Published in final edited form as:

Neuropsychology. 2015 January ; 29(1): 82–91. doi:10.1037/neu0000079.

Neuroanatomical correlates of behavioral rating versus performance measures of working memory in typically developing children and adolescents

Nazlie Faridi^a, Sherif Karama^{a,b,*}, Miguel Burgaleta^b, Matthew White^c, Alan C. Evans^b, Vladimir Fonov^b, D. Louis Collins^b, and Deborah P. Waber^{c,*}

^aDepartment of Psychiatry, McGill University, Montreal, Canada

^bMcConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

^cDepartment of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, USA

Neuropsychologists use two primary classes of assessment tools: performance measures and structured questionnaires. Questionnaires have become an increasingly popular strategy to assess executive functioning, which has been characterized as "a set of general-purpose control mechanisms, often linked to the pre-frontal cortex of the brain, that regulate the dynamics of human cognition and action (Miyake & Friedman, 2012, p. 8). Whereas laboratory measures provide a sampling of cognitive aspects of executive functioning in the context of well-structured tasks, questionnaire measures are thought to provide an indication of the individual's executive functioning as manifest in the context of day to day life.

The Behavioral Rating Inventory of Executive Functioning (BRIEF) is a widely used questionnaire instrument, with versions capturing behavior from preschool to adulthood (Gioia, Isquith, Guy, & Kenworthy, 2000). It includes parent, teacher and self-report options. For school-age children and adolescents, the BRIEF is an observational questionnaire completed by parents or teachers that yields two main indices, Behavioral Regulation and Metacognition, each with component subscales. A number of research studies, however, have documented a concerning lack of correspondence between performance measures and specific BRIEF scales that purport to measure the same component of executive functioning. McAuley, Chen, Goos, Schachar, and Crosbie (2010) provide a comprehensive review of 11 studies in which laboratory measures of executive function (EF) and the BRIEF were administered. Collectively, these studies find only inconsistent associations between parent ratings on the BRIEF and a variety of performance measures of EF. Nevertheless, in some settings, the observational scores can be more sensitive to differences between clinical and control groups than the performance measures (Mrakotsky et al., 2012; Piper et al., 2011; Toplak, Bucciarelli, Jain, & Tannock, 2009), although the reasons for this are unclear.

*Correspondence to: Sherif Karama, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, 3801 University, Montreal, QC, H3A 2B4, sherif.karama@mcgill.ca; or Deborah P. Waber, Department of Psychiatry, Boston Children's Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, deborah.waber@childrens.harvard.edu.
Miguel Burgaleta is now based at Facultad de Psicología, Universidad Autónoma de Madrid, Campus de Cantoblanco, Madrid, Spain.

Respondents might, for example, be influenced by knowing that the child has a clinical condition or exposure.

In their own study, McAuley et al. (2010) report higher correlations of the BRIEF with other (questionnaire) measures of behavior disruption than with laboratory measures of executive functioning. They offer a number of explanations for this finding, ranging from a hypothesis that the performance measures assess underlying skills whereas the BRIEF assesses application of these skills, to a hypothesis that the BRIEF does not assess executive functioning to the extent that is commonly believed. They raise the concern, moreover, that the BRIEF is really a measure of behavioral problems and not executive functioning.

Another strategy that can be used to elucidate the discrepancy between performance and questionnaire measures of EF is to reference neural correlates. Whereas many laboratory tasks can be directly translated to analog functional neuroimaging paradigms, questionnaire measures cannot. Ratings on these scales can, however, be correlated with morphometric neuroimaging. Mahone, Martin, Kates, Hay, and Horská (2009) correlated a performance-based measure of working memory (WJ-III Auditory Working Memory) and the BRIEF Parent Report Working Memory Scale with measures of lobar gray and white matter volumes (standardized for total brain volume) in a sample of 35 typically developing children. They contrasted these correlations with a set of nonexecutive outcomes (WJ-III Spatial Relations, CBCL Anxious/Depressed Scale). There was a significant correlation between the BRIEF Working Memory Scale and frontal gray matter volume and a somewhat smaller but still statistically significant correlation between the WJ-III Auditory Working Memory test and frontal gray matter volume. Despite the association of both outcomes with the volume of frontal grey matter, the correlation between the two working memory measures was small and not statistically significant. The association of these working memory measures with frontal gray matter volume was consistent with predictions based on the established association between working memory and frontal cortex demonstrated on functional neuroimaging paradigms (Casey, Cohen, Jezzard, & Turner, 1995; Kwon, Reiss, & Menon, 2002). The lack of correspondence between the laboratory and observational measures, although not surprising given the literature, remains perplexing. The different types of measures, however, might be associated with different regions of the frontal cortex.

Working memory is generally defined as the ability to hold information in memory for short time periods for use in complex tasks (Baddeley, 1998). Performance measures typically require the person to maintain information in a short-term store and manipulate it mentally in the context of a more complex task. Thus, for example, the digit span task, a standard measure of working memory, requires that the respondent listen to a string of random digits and then recite them backwards. The BRIEF includes items that are believed to capture working memory capacity in everyday life. Representative items include “Has trouble finishing tasks (chores, homework),” “When given three things to do, remembers only the first or last,” and “Has a short attention span.”¹

¹Reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, Florida 33549, from the Behavior Rating Inventory of Executive Function by Gerard A. Gioia, Peter K. Isquith, Steven C. Guy and Lauren Kenworthy, Copyright 1996, 1998, 2000 by PAR, Inc. Further reproduction is prohibited without permission from PAR, Inc.

In the present study, which was modeled after the Mahone et al. (2009) study, we correlated volumetric and cortical thickness measures obtained from neuroimaging with performance and questionnaire measures of working memory. The data were derived from the NIH MRI Study of Normal Brain Development, a large population-based sample that was studied longitudinally, providing ample power to detect relevant associations and dissociations. Brain measurements included lobar volumes, as well as volumes of subcortical structures and local cortical thickness measurements. We evaluated associations of these morphometric measurements with laboratory measures of verbal and spatial working memory, as well as with the Working Memory scale from the BRIEF. Associations with particular, ideally different, morphometric features could inform our appreciation of the basis for the association or lack of association between the laboratory and questionnaire modalities, potentially providing greater insight into what the BRIEF, and in particular the Working Memory scale, measures. As a contrast, we carried out comparable analyses of the BRIEF Inhibition and Emotional Control scales. These latter scales were chosen as contrast variables because their neural substrates would be assumed to differ from those associated with working memory in that they do not entail regulation of cognitive processes as much as action and emotion.

Based on the Mahone et al. (2009) findings, we predicted that the working memory measures, both performance and observational, would be specifically associated with the volume of frontal grey matter, but not with the other lobar volumes, and that larger volumes would correlate with better functioning. We further predicted that the analyses of cortical thickness would highlight specific regions of frontal cortex as highly correlated with performance or observational measures of working memory. Based on results from prior studies, we predicted that cortical thickness in left lateral prefrontal regions would be associated with the working memory measures (Casey et al., 1995; Klingberg, 2006).

Method

Study Design and Participants

Data were obtained from the Pediatric MRI Data Repository (Release 4.0) of the NIH MRI Study of Normal Brain Development, a longitudinal project developed to characterize healthy brain maturation in relation to behavior in a large, multi-site study (Evans & Brain Development Cooperative Group [BDCG], 2006). Children and adolescents living near six Pediatric Study Centers across the United States were recruited via a population-based sampling method designed to limit biases. The overall study sample consisted of 433 healthy children aged 4 to 21 years, and was demographically representative of the US population in terms of variables including gender, race, and socioeconomic status (Waber et al., 2007). Exclusion criteria included but were not limited to: IQ < 70, history of medical illness with CNS implications, and any Axis I psychiatric disorder (other than simple or social phobia, adjustment disorder, oppositional defiant disorder, enuresis, encopresis, or nicotine dependency; see Waber et al., 2007 for a complete list of inclusion and exclusion criteria).

Participants underwent brain MR imaging and extensive neuropsychological testing on up to three occasions at two year intervals. For the purposes of this report, participants under 6 and over 16 years of age were excluded so that the versions of the neuropsychological measures

would be consistent across participants. Within this age range, 57.2% completed all three visits, 29% completed only visits 1 and 2, and 12.2% completed the first visit only (Waber, Forbes, Almlil, Blood, & BDCG, 2012). The number of visits for individuals whose scans were used in this analysis is displayed in Table 1.

Neuropsychological Measures

The neuropsychological measures included the Behavior Rating Inventory of Executive Function (BRIEF) as well as performance measures of verbal and spatial working memory. These were administered at each visit, typically on the same day as the scan. Raw scores were used for all analyses so as to avoid distortions that could potentially have been introduced by the norming process.

BRIEF.—The parent version of the BRIEF (Gioia et al., 2000) was administered at each visit. The primary outcomes were the Working Memory (WM), Emotional Control (EC) and Inhibition (INH) scales; the latter two were included as contrast conditions. Raw scores were used in all analyses. On all scales, higher scores reflect more behavioral problems.

Wechsler Intelligence Scale for Children-III (WISC-III) Digit Span (DS).—This task asks the child to repeat random digit strings of increasing length, both in the order the numbers are presented (a measure of short-term memory) and backwards (a measure of working memory) (Wechsler, 1991). The primary outcome was the Backward score, with higher scores reflecting better performance.

Cambridge Neuropsychological Test Battery (CANTAB) Spatial Working Memory (SWM).—(CeNeS, 1998). This is a computer-administered serial order pointing task (Petrides & Milner, 1982) in which participants search through an increasing number of boxes (from two to eight) to locate hidden blue squares, or tokens. Selecting a box already found to contain a token is considered a “within search error,” whereas returning to search a box that has already been found to be empty is a “between search error.” Six and seven year olds were administered a maximum of six boxes to avoid frustration and fatigue (Luciana & Nelson, 1998; Waber et al., 2007). The primary outcome was Total Errors, with higher scores indicating more errors.

MRI Protocol and Processing

Imaging data were obtained for each participant on the day of or within a maximum of 28 days of psychometric testing at each visit. Given the age range of participants, ~25 min of data were acquired on 1.5T scanners at the six study sites. Multiple contrasts were obtained (T1-weighted, T2-weighted, and proton-density weighted). A 3D T1-weighted spoiled gradient recalled echo sequence was selected, providing 1mm sagittally-acquired isotropic data for the entire head. Because the maximum number of slices on the GE scanners was 124, the slice thickness was increased to ~1.5mm. Inter-site reliability was evaluated with American College of Radiology phantoms, as well as with living phantoms scanned at each site. (For a complete description, please see Evans & BDCG, 2006.)

All MRI images were processed at the Montreal Neurological Institute (MNI) with the CIVET pipeline (version 1.1.9), an automated structural image analysis tool (Ad-Dab'bagh et al., 2006; Collins, Neelin, Peters, & Evans, 1994; Mazziotta, Toga, Evans, Fox, & Lancaster, 1995; Sled, Zijdenbos, & Evans, 1998; Talairach & Tournoux, 1988; Zijdenbos, Forghani, & Evans, 2002). The CLASP algorithm, contained in this pipeline, was used to generate cortical thickness at 40,962 vertices per hemisphere, calculated as the distance between the cerebrospinal fluid-grey matter and grey matter-white matter interfaces (Chung et al., 2001; Kabani, Le Goualher, MacDonald, & Evans, 2001; Kim et al., 2005; Lerch & Evans, 2005; Lyttelton, Boucher, Robbins, & Evans, 2007; MacDonald, Kabani, Avis, & Evans, 2000).

Volumetric measures were obtained with the ANIMAL algorithm, a validated, fully-automated segmentation method (Collins, Holmes, Peters, & Evans, 1995). In addition, an augmented multi-segmentation and label fusion technique was used to derive hippocampus and amygdala volumes (Collins & Pruessner, 2010).

Data Set

Because of the strict quality control process for the imaging measures (Evans & BDCG, 2006), a number of participants with high-quality lobar and basal ganglia volume data lacked adequate cortical thickness (90 cases) or hippocampus/amygdala data (41 cases). In order to preserve comparability of the volumetric results, cases (i.e., a given individual on a given test day) lacking hippocampus and amygdala data were excluded from all volumetric analyses. In order to maximize sample size, however, cases without cortical thickness data were retained for the volumetric analysis, and vice versa. As a result, volumetric analyses were conducted on data derived from 347 participants, with 663 scans in total, whereas the cortical thickness analysis sample included 350 participants, with a total of 615 observations. There were no significant differences between the two highly-overlapping samples in terms of age, gender, handedness or neuropsychological scores. Demographic characteristics of the longitudinal sample used in volumetric analyses are presented in Table 1.

Statistical Analysis

Mixed-effects linear regression models were used to assess associations between neuroimaging measures and behavioral outcomes in order to account for missing data and repeated measurements over time. These models were adjusted for the effects of age, gender, and a measure of total brain volume. Analyses of associations between behavioral measures included only age and gender as covariates. Handedness and scanner were each considered as covariates but were found to be non-contributory and dropped from the models.

Volumetric analyses.—Statistical analyses involving brain volumes were carried out with SPSS 19.0 (SPSS Inc., Chicago, Illinois). The neuropsychological variables were linearly regressed against lobar (frontal, temporal, parietal and occipital) grey and white matter volumes, as well as caudate, putamen, globus pallidus, amygdala and hippocampus volumes in a mixed-effects model controlling for age, gender, and (total brain volume – volume of

interest). Each model was fitted using restricted maximum likelihood, and a first-order autoregressive covariance structure was assumed for the error terms.

$$Y_{ij} = \beta_1 \text{Age}_i + \beta_2 \text{Gender}_i + \beta_3 (\text{Total Brain Volume}_i - \text{Volume of Interest}_i) + \beta_4 (\text{Volume of Interest}_i) + \gamma_i + \varepsilon_{ij},$$

where Y_{ij} represents the behavioural variable of interest for subject i at time j , β_1 , β_2 , β_3 and β_4 are regression coefficients, γ_i is the random intercept for subject i , and ε_{ij} is the error for subject i at time j . The behavior variables, age, total brain volume, and volume of interest measures were standardized for analysis, which yield standardized parameter estimates for comparison across models. The “Age by Volume of Interest” interaction was also analyzed. Collinearity diagnostics, including the variance inflation factor (VIF) and condition index (CI), were used to assess the degree of collinearity among variables in each model. If high collinearity ($\text{VIF} > 3$ or $\text{CI} > 20$ for any variable in the model) was found, the model was rerun excluding the total brain volume variable and reassessed for collinearity issues. Unless otherwise stated in the reporting of results, collinearity was not an issue in the models for volumetric analyses, and removing the total brain volume measure did not significantly affect results.

For all volumetric analyses, bilateral structures were combined since left and right components were very highly correlated. When correlation coefficients for left and right sides were < 0.9 , hemispheres were analyzed separately (right-handed individuals only). This was the case for three structures: hippocampus, $r(345) = .89$, $p < .001$; amygdala $r(345) = .80$, $p < .001$; and globus pallidus, $r(345) = .70$, $p < .001$.

A Bonferroni correction was applied, accounting for the non-independence of the measures of interest, leading to an adjusted significance threshold of $\alpha = .0047$ (see <http://www.quantitativeskills.com/sisa/calculations/bonfer.htm>).

Cortical thickness analyses.—Relationships between cortical thickness and the neuropsychological measures were evaluated using Surfstat (Worsley, Taylor, Tomaiuolo, & Lerch, 2004; <http://www.math.mcgill.ca/keith/surfstat/>), a statistical toolbox created for MATLAB (The MathWorks, Inc). Each individual’s native-space cortical thickness was linearly regressed in a first-order mixed effects model against each of the psychometric scores, while accounting for the effects of age, gender, and a proxy measure of brain volume (pBV, calculated as the sum of total white matter, intracerebral CSF and subcortical grey matter). As previously demonstrated by our group, cortical grey matter accounts for 40% of total brain volume, and is very highly correlated with mean cortical thickness (> 0.75 ; Karama et al., 2011). Thus, correcting for total brain volume would, in effect, partially correct for cortical thickness itself. For this reason, pBV was included in the regression models in place of total brain volume (Karama et al., 2011).

Additionally, age was included as a first-order term in the model, rather than a cubic or quadratic term, since the age effect on cortical thickness in the total study sample has been shown to be best described by a first order linear function (Brain Development Cooperative

Group, 2011; Nguyen et al., 2012). The “Age by Behavioral Variable” interaction was also analyzed.

A whole-brain correction, using random field theory (RFT) with $p < .05$, was used to account for multiple comparisons (Worsley, Taylor, Tomaiuolo, & Lerch, 2004).

Results

Associations between Performance and Questionnaire Measures

Table 2 shows the results of the mixed model analyses demonstrating associations among the BRIEF and the performance measures. The BRIEF measures were all highly intercorrelated, and the CANTAB SWM Total Errors was highly correlated with Digit Span Backward. The BRIEF WM scale was significantly correlated with CANTAB SWM Total Errors but not the DS Backward.

Intraclass correlations for each of the neuropsychological variables ranged from 0.71 for DS Backward to 0.85 for BRIEF WM, indicating relatively consistent scores across visits.

Associations between Brain Volumes and Neuropsychological Measures

BRIEF.—There were no meaningful associations between BRIEF WM and frontal grey matter ($p = 0.6314$), nor were there associations with other lobar volumes. In contrast, there were significant negative correlations between temporal grey matter volume and the BRIEF INH ($\beta = -0.15$, S.E. = 0.05, $p = 0.0011$) and EC ($\beta = -0.13$, S.E. = 0.04, $p = 0.0032$) subscales. The temporal grey matter and total brain matter volumes were found to be collinear in these models, and therefore the above results were obtained from a model excluding the total brain volume measure. Since the coefficients from the reduced model were also more conservative (closer to 0) than the full model, these estimates do not overstate the association of temporal grey matter with BRIEF INH or EC.

Performance Measures.—There were similarly no statistically significant associations between frontal gray and white matter volumes and the performance measures of working memory. However, there were negative correlations between SWM Total Errors and DS Backward with right hippocampal volume ($\beta = -0.16$, S.E. = 0.05, $p = 0.0018$ and $\beta = 0.17$, S.E. = 0.06, $p = 0.0034$, respectively). DS Backward was also associated with the volume of the amygdala ($\beta = 0.18$, S.E. = 0.05, $p = 0.001$, right amygdala; $\beta = 0.17$, S.E. = 0.05, $p = .0015$ left amygdala).

The “Age by Volume of Interest” interaction was not significant in all models for both BRIEF and performance measures.

Associations between Cortical Thickness and Neuropsychological Measures

BRIEF.—WM was negatively correlated with the cortical thickness of both the left and right parahippocampal gyri (PHG), to a greater extent on the left (Figure 1a). EC was negatively correlated with right PHG thickness only (Figure 1b). INH was not significantly correlated with any cortical thickness measures. The negative correlations signify that for each of these scales, greater cortical thickness was associated with lower levels of problems.

Performance Measures.—DS Backward and CANTAB SWM were not significantly associated with any region of cortical thickness.

There were also no significant interaction effects of age for either BRIEF or performance measures.

Discussion

The primary goal of this study was to leverage structural neuroimaging to gain insight into the frequently observed lack of correspondence between performance and questionnaire measures of executive functioning, in particular working memory. An unexpected but reliable association was found between the BRIEF WM scale and cortical thickness in the PHG, whereas performance measures of working memory were associated with the volumes of deeper medial temporal lobe structures, the hippocampus and amygdala. Contrary to expectation, we did not confirm associations between working memory and frontal measurements, either volumes or cortical thickness, for either the performance measures or the BRIEF WM scale, despite the power of this very large data set to detect associations. Nor were frontal measurements correlated with any other BRIEF measure.

The BRIEF EC and INH scales were included as contrast variables, and indeed their associations differed from those of the WM scale. EC and INH were negatively correlated with temporal lobe volume, whereas WM was not. The BRIEF WM scale was associated with cortical thickness in PHG bilaterally, but more prominently on the left side. EC, in contrast, was significantly associated only with the right PHG, and INH was not associated with any cortical thickness measure. Since a number of items from the WM scale implicitly involve language and sequential processing, the laterality differences would be consistent with the known functional association of the left hemisphere with language and sequencing, and the right with social and emotional cognition. Unfortunately, the magnitude of the effects was insufficient to prove such a distinction statistically and it must therefore be viewed as speculative. Nevertheless, there does appear to be meaningful specificity in the association of the BRIEF WM scale with particular neuroanatomical measurements.

We also examined intercorrelations between the performance and questionnaire measures. Although a number of studies have failed to document associations between the BRIEF scales and performance measures of executive functions, we did find a statistically significant association between the BRIEF WM and CANTAB SWM, albeit of small magnitude. These findings are partially consistent with a previous report demonstrating an association between the BRIEF WM scale and a composite measure combining WISC Digit Span and Spatial Span scores (Toplak et al., 2009).

Our failure to replicate the findings of Mahone et al. (2009) may reflect the instability of their relatively small sample size and the fact that they did not correct for multiple comparisons. We were also unable to replicate the results of Pangelinan et al. (2011), who demonstrated an association between putamen volume and CANTAB SWM in a smaller (217 observations) and younger subset (<13 years old) of the NIH database. Again, this discrepancy may reflect differences in statistical methods, as they neither corrected for

multiple comparisons nor controlled for total brain volume, and their twotier analysis (first generating residuals from linear mixed effects modeling, then analyzing these residuals with Pearson's correlations) may have artificially inflated degrees of freedom.

The present findings raise two important issues. The first is the lack of correspondence between measures purported to assess the same function, in this case working memory, and what this implies for neuropsychological research and practice. The second is the significance of the unanticipated but apparently reliable association of the BRIEF WM scale with measures of cortical thickness in the PHG.

With respect to the first issue, the discrepancy between performance and observational measures of "working memory," in terms of their association with one another as well as with the neuroanatomical measures, highlights the risk of overinterpreting diagnostic labels associated with a particular test, leading to assumptions of greater equivalence than is the case. The term "working memory," as originally framed (Baddeley, 1986), referred to a discrete cognitive operation that could be demonstrated in well structured experimental paradigms and was relatively independent of context. The developers of the BRIEF, in contrast, conducted a factor analysis of a large number of behavioral descriptors thought to reflect executive functioning, as endorsed by parents and teachers, and identified a cluster that cohered statistically. Since the items in this factor resonated with the general understanding of working memory, they assigned it that label. Although the scale may well reflect processes that are loosely related to the original WM construct and, importantly, assess "real world" behaviors, it is not surprising that equivalence between the questionnaire and performance measures cannot be demonstrated.

Indeed, it was this observed discrepancy that motivated the present study and thus leads to the second issue, that is, the relevance of neuroanatomical correlates for appreciating the significance of the behavioral measures. In this regard, the association between the BRIEF WM scale and the PHG is of greatest interest. Medial temporal lobe (MTL) structures have long been viewed as critical to episodic (Eichenbaum, Otto, & Cohen, 1992; Scoville & Milner, 1957), or autobiographical (Tulving, 2002), memory. In episodic memory, encoding and retrieval are integrally related to the context within which the memory was formed. The classic differentiation between episodic and semantic memory is the difference between knowing that "Paris is where I first tasted snails" versus knowing that "Paris is the capital of France."

Recently, a "binding of item and context" model has been proposed to parse the specific contributions to episodic memory of three MTL structures: the hippocampus, parahippocampal cortex (posterior PHG) and perirhinal cortex (anterior PHG) (Diana, Yonelinas, & Ranganath, 2007; Eichenbaum, Yonelinas, & Ranganath, 2007). According to this model, parahippocampal cortex encodes information about the context in which items to be remembered are presented, whereas the perirhinal cortex encodes information about the items themselves. The data then converge in the hippocampus, where 'item' and 'context' are merged into an episodic memory.

Bergmann and colleagues (Bergmann, Rijpkema, Fernández, & Kessels, 2012), using fMRI, assessed both working memory and long-term memory with a relational memory task, using pairs of faces and houses. Activation in left PHG was uniquely related to working memory but not long-term memory. The authors inferred that parahippocampal regions “may be involved in transient memory representations, supporting performance over short delays.” In another study (Howard, Kumaran, Ólafsdóttir, & Spiers, 2011) PHG was specifically sensitive to scene novelty, consistent with the view that it supports neural representation of the global context of episodes or the local layout of an environment.

There is, further, an emerging literature on the functional development of PHG in typically developing children and young adults. Gheiti and colleagues (Gheiti, DeMaster, Yonelinas, & Bunge, 2010) used an incidental learning paradigm to investigate developmental differences in medial temporal lobe (MTL) function. They documented a progressive differentiation between item and context over developmental time. In younger children (8-year olds), activation in MTL regions was associated with recollection of both the item and contextual cues, whereas in 14-year olds, activation in these regions was associated with recollection of contextual cues but not items. We did evaluate potential interactions with age in the present study; however, none were detected.

Since the BRIEF is explicitly contextual in nature, that is, items are referenced to a child’s behavioral patterns in the context of everyday life, its specific association with the PHG is interpretable. In particular, it would make sense that the BRIEF would bear a relationship to the medial temporal lobe system, which is central to the binding of items to context. Although we know of no other studies of this type in children, memory in daily life, as assessed by a self-report questionnaire, has previously been linked to medial temporal lobe thickness in a healthy elderly population (Bjørnebekk, Westlye, Walhovd, & Fjell, 2010).

Whereas the BRIEF scales were associated with cortical thickness in the PHG, the performance measures of working memory were associated with right hippocampal volume. Such associations have been previously demonstrated in fMRI paradigms of working memory (Abrahams et al., 1999; Ranganath & Blumenfeld, 2005). In addition, we found an association between Digit Span Backward and amygdala volume, consistent with findings from an fMRI study in which event-related amygdala activation was correlated with performance during an emotionally neutral working memory task (Schaefer et al., 2006). It is important to note, however, that the present findings pertain only to anatomy, whereas the above referenced studies assessed function.

The BRIEF EC scale, included as a contrast condition, did in fact align differently than the WM scale. It was negatively correlated not only with PHG thickness (right side only) but also with temporal lobe volume bilaterally. This latter finding is consistent with several neuroimaging studies implicating these structures in emotional appraisal, with alterations in volume and function being linked to emotional dysregulation in adults with Bipolar Disorder (Almeida, Akkal, et al., 2009; Almeida, Mechelli, et al., 2009), Social Anxiety Disorder (Liao et al., 2011) and Schizophrenia (Surguladze et al., 2006). Perhaps more relevant to this developmental study, temporal lobe morphology has been implicated in children with schizophrenia prodrome (Cullen et al., 2012) and in adolescents with major depressive

disorder (Shad, Muddasani & Rao, 2012). The negative correlation of INH with temporal lobe volume is also supported by the functional neuroimaging literature. Enhanced temporal lobe activity during response inhibition tasks has been observed in children with ADHD when contrasted with an unaffected control group (Schulz et al., 2004).

The overlap between items included in the BRIEF WM scale and inattentive symptoms of ADHD has been previously noted, not only by critics (McAuley et al., 2010), but by the authors of the BRIEF themselves (Gioia et al., 2000), who suggest that working memory and the ability to sustain performance and attention may be difficult to distinguish. A previous study from our group, based on the same database, reported a parallel set of analyses using the Attention Problems scale from the Child Behavior Checklist (Ducharme et al., 2012). That study identified age-dependent associations with cortical thickness in multiple regions of frontal cortex typically associated with the attention network. One area in left medial temporal cortex was identified, suggesting potential overlap. Overall, however, the findings for this attention scale were quite different from those for the BRIEF WM, suggesting that they can be distinguished.

The present study also had limitations. First, there were only two performance measures of working memory and the task demands were somewhat different, potentially attenuating our ability to detect associations between performance measures of working memory and brain measures. Second, because the sample had been screened for disorders (e.g., ADHD), the strengths of associations between both the performance and behavioral measures and the brain measures could have been constrained.

In sum, this study provides new insights into the typically observed dissociation between performance and behavioral rating measures that purport to assess the same function, in particular, working memory. The findings suggest that the BRIEF Working Memory scale is sensitive to the momentary binding of item and context that occurs on an ongoing basis in a real world environment. It therefore may reflect functions akin to episodic or autobiographic memory to a greater extent than working memory *per se*. If so, it would measure an important capacity to which standard laboratory performance measures are not sensitive because they are context independent. Whether this is the case and how the observed anatomical correlates may be related to function cannot be determined on the basis of these data; however, the neuroanatomical correlations demonstrated here provide clues to appreciating similarities and differences between behavioral rating and performance measures of working memory. Moreover, they situate the BRIEF in particular within the purview of contemporary cognitive neuroscientific models.

Acknowledgments

This project was conducted by the Brain Development Cooperative Group and supported by the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute of Neurological Disorders and Stroke (Contract #s N01-HD02-3343, N01-MH9-0002, and N01-NS-9-2314, -2315, -2316, -2317, -2319 and -2320). Special thanks to the NIH contracting officers for their support. S.K. was supported by a Fellowship from the Fonds de Recherche En Santé du Québec. We also acknowledge the important contribution and remarkable spirit of John Haselgrove, Ph.D. (deceased).

References

- Abrahams S, Morris RG, Polkey CE, Jarosz JM, Cox TCS, Graves M, & Pickering A (1999). Hippocampal involvement in spatial and working memory: A structural MRI analysis of patients with unilateral mesial temporal lobe sclerosis. *Brain and Cognition*, 41, 39–65. [PubMed: 10536085]
- Ad-Dab'bagh Y, Lyttelton O, Muehlboeck J, Lepage C, Einarson D, & Mok K (2006). The CIVET image-processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research. In Corbetta M (Ed.), *Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping Florence, Italy*.
- Almeida JRC, Akkal D, Hassel S, Travis MJ, Banihashemi L, Kerr N ...Phillips ML. (2009). Reduced gray matter volume in ventral prefrontal cortex but not amygdala in bipolar disorder: Significant effects of gender and trait anxiety. *Psychiatry Research*, 171(1), 54–68. doi:10.1016/j.psychres.2008.02.001 [PubMed: 19101126]
- Almeida JRC, Mechelli A, Hassel S, Versace A, Kupfer DJ, & Phillips ML (2009). Abnormally increased effective connectivity between parahippocampal gyrus and ventromedial prefrontal regions during emotion labeling in bipolar disorder. *Psychiatry Research*, 174(3), 195–201. doi: 10.1016/j.psychres.2009.04.015 [PubMed: 19910166]
- Baddeley A (1986). *Working memory*. Applied Cognitive Psychology (Vol. 2). Oxford: Oxford University Press. doi:10.1002/acp.2350020209
- Baddeley A (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, 8(2), 234–8. [PubMed: 9635207]
- Bergmann HC, Rijpkema M, Fernández G, & Kessels RPC (2012). Distinct neural correlates of associative working memory and long-term memory encoding in the medial temporal lobe. *NeuroImage*. doi:10.1016/j.neuroimage.2012.03.047
- Bjørnebekk A, Westlye LT, Walhovd KB, & Fjell AM (2010). Everyday memory: Self-perception and structural brain correlates in a healthy elderly population. *Journal of the International Neuropsychological Society*, 16(6), 1115–26. doi:10.1017/S1355617710001025 [PubMed: 20946708]
- Brain Development Cooperative Group. (2011). Total and regional brain volumes in a population-based normative sample from 4 to 18 years: The NIH MRI study of normal brain development. *Cerebral Cortex*, 22, 1–12. doi:10.1093/cercor/bhr018 [PubMed: 21613470]
- Casey B, Cohen J, Jezzard P, & Turner R (1995). Activation of prefrontal cortex in children during a nonspatial working memory task with functional MRI. *Neuroimage*, 2(3), 221 Retrieved from <http://ukpmc.ac.uk/abstract/MED/9343606> [PubMed: 9343606]
- Chung MK, Worsley KJ, Taylor J, Ramsay J, Robbins S, & Evans AC (2001). Diffusion smoothing on the cortical surface. *NeuroImage*, 13(6), 95. doi:10.1016/S1053-8119(01)91438-7
- Collins DL, Holmes C, Peters T, & Evans AC (1995). Automatic 3D model-based neuroanatomical segmentation. *Human Brain Mapping*, 208(1995), 190–208. doi:10.1002/hbm.460030304
- Collins DL, Neelin P, Peters T, & Evans AC (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, (18), 192–205. [PubMed: 8126267]
- Collins DL, & Pruessner JC (2010). Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *NeuroImage*, 52(4), 1355–66. doi:10.1016/j.neuroimage.2010.04.193 [PubMed: 20441794]
- Cullen AE, De Brito SA, Gregory SL, Murray RM, Williams SC, Hodgins S, Laurens KR (2012). Temporal lobe volume abnormalities precede the prodrome: A study of children presenting antecedents of schizophrenia. *Schizophrenia Bulletin*. Advance online publication. doi: 10.1089/cap.2011.0005.
- Diana RA, Yonelinas AP, & Ranganath C (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*, 11(9), 379–86. doi: 10.1016/j.tics.2007.08.001 [PubMed: 17707683]
- Ducharme S, Hudziak JJ, Botteron KN, Albaugh MD, Nguyen T-V, Karama S, & Evans AC (2012). Decreased regional cortical thickness and thinning rate are associated with inattention symptoms

- in healthy children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(1), 18–27. doi:10.1016/j.jaac.2011.09.022 [PubMed: 22176936]
- Eichenbaum H, Yonelinas AP, & Ranganath C (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30, 123–52. doi:10.1146/annurev.neuro.30.051606.094328
- Eichenbaum H, Otto T, & Cohen NJ (1992). The hippocampus—what does it do? *Behavioral and Neural Biology*, 57(1), 2–36. doi:10.1016/0163-1047(92)90724 [PubMed: 1567331]
- Evans AC, & Brain Development Cooperative Group (2006). The NIH MRI study of normal brain development. *NeuroImage*, 30(1), 184–202. doi:10.1016/j.neuroimage.2005.09.068 [PubMed: 16376577]
- Ghetti S, DeMaster DM, Yonelinas AP, & Bunge SA (2010). Developmental differences in medial temporal lobe function during memory encoding. *Journal of Neuroscience*, 30 (28), 9548–9556. [PubMed: 20631183]
- Gioia GA, Isquith PK, Guy SC, & Kenworthy L (2000). Behavior rating inventory of executive function. Odessa, FL: Psychological Assessment Resources.
- Howard L, Kumaran D, Ólafsdóttir H, & Spiers H (2011). Double dissociation between hippocampal and parahippocampal responses to object-background context and scene novelty. *Journal of Neuroscience*, 31(14), 5253–5261. Retrieved from <http://www.jneurosci.org/content/31/14/5253.short> [PubMed: 21471360]
- Kabani N, Le Goualher G, MacDonald D, & Evans AC (2001). Measurement of cortical thickness using an automated 3-D algorithm: a validation study. *NeuroImage*, 13(2), 375–80. doi:10.1006/nimg.2000.0652 [PubMed: 11162277]
- Karama S, Colom R, Johnson W, Deary IJ, Haier R, Waber DP, ... Evans AC (2011). Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. *NeuroImage*, 55(4), 1443–53. doi:10.1016/j.neuroimage.2011.01.016 [PubMed: 21241809]
- Kim JS, Singh V, Lee JK, Lerch J, Ad-Dab'bagh Y, MacDonald D, ... Evans, A. C. (2005). Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage*, 27(1), 210–21. doi:10.1016/j.neuroimage.2005.03.036 [PubMed: 15896981]
- Klingberg T (2006). Development of a superior frontal-intraparietal network for visuospatial working memory. *Neuropsychologia*, 44(11), 2171–7. doi:10.1016/j.neuropsychologia.2005.11.019 [PubMed: 16405923]
- Kwon H, Reiss AL, & Menon V (2002). Neural basis of protracted developmental changes in visuospatial working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), 13336–41. doi:10.1073/pnas.162486399 [PubMed: 12244209]
- Lerch JP, & Evans AC (2005). Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, 24(1), 163–73. doi:10.1016/j.neuroimage.2004.07.045 [PubMed: 15588607]
- Liao W, Xu Q, Mantini D, Ding J, Machado-de-Sousa JP, Hallak JEC, ... Chen, H. (2011). Altered gray matter morphometry and resting-state functional and structural connectivity in social anxiety disorder. *Brain Research*, 1388, 167–77. doi:10.1016/j.brainres.2011.03.018 [PubMed: 21402057]
- Luciana M, & Nelson CA (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, 36(3), 273–93. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9622192> [PubMed: 9622192]
- Lytelton O, Boucher M, Robbins S, & Evans A (2007). An unbiased iterative group registration template for cortical surface analysis. *NeuroImage*, 34(4), 1535–44. doi:10.1016/j.neuroimage.2006.10.041 [PubMed: 17188895]
- MacDonald D, Kabani N, Avis D, & Evans AC (2000). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, 12(3), 340–56. doi:10.1006/nimg.1999.0534 [PubMed: 10944416]
- Mahone EM, Martin R, Kates WR, Hay T, & Horská A (2009). Neuroimaging correlates of parent ratings of working memory in typically developing children. *Journal of the International*

- Neuropsychological Society, 15(1), 31–41. doi:10.1017/S1355617708090164 [PubMed: 19128526]
- Mazziotta JC, Toga AW, Evans A, Fox P, & Lancaster J (1995). A probabilistic atlas of the human brain: Theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *NeuroImage*, 2(2), 89–101. doi:10.1006/nimg.1995.1012 [PubMed: 9343592]
- McAuley T, Chen S, Goos L, Schachar R, & Crosbie J (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function? *Journal of the International Neuropsychological Society*, 16(3), 495–505. doi:10.1017/S1355617710000093 [PubMed: 20188014]
- Miyake A & Friedman NP (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. [PubMed: 22773897]
- Mrakotsky C, Forbes PW, Bernstein JH, Grand RJ, Bousvaros A, Szigethy E, & Waber DP (2012). Acute cognitive and behavioral effects of systemic corticosteroids in children treated for inflammatory bowel disease. *Journal of the International Neuropsychological Society*, 1–14. doi: 10.1017/S1355617712001014
- Nguyen T-V, McCracken J, Ducharme S, Botteron KN, Mahabir M, Johnson W, ... Karama S (2012). Testosterone-related cortical maturation across childhood and adolescence. *Cerebral Cortex*, (3), 1–9. doi:10.1093/cercor/bhs125
- Pangelinan MM, Zhang G, VanMeter JW, Clark JE, Hatfield BD, & Haufler AJ (2011). Beyond age and gender: Relationships between cortical and subcortical brain volume and cognitive-motor abilities in school-age children. *NeuroImage*, 54, 3093–3100. doi: 10.1016/j.neuroimage.2010.11.021 [PubMed: 21078402]
- Petrides M, & Milner B (1982). Deficits on subject-ordered tasks after frontal and temporal lobe-lesions in man. *Neuropsychologia*, 20(3), 249–262. [PubMed: 7121793]
- Piper BJ, Acevedo SF, Kolchugina GK, Butler RW, Corbett SM, Honeycutt EB, ... Raber J (2011). Abnormalities in parentally rated executive function in methamphetamine/polysubstance exposed children. *Pharmacology, Biochemistry, and Behavior*, 98(3), 432–9. doi:10.1016/j.pbb.2011.02.013
- Ranganath C, & Blumenfeld RS (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*. 9(8), 374–308. doi:10.1016/j.tics.2005.06.009 [PubMed: 16002324]
- Schaefer A, Braver TS, Reynolds JR, Burgess GC, Yarkoni T, & Gray JR (2006). Individual differences in amygdala activity predict response speed during working memory. *The Journal of Neuroscience*. 26(40), 10120–8. doi:10.1523/JNEUROSCI.2567-06.2006 [PubMed: 17021168]
- Schulz K, Fan J, Tang CY, Newcorn JH, Buchsbaum MS, Cheung AM, & Halperin JM (2004). Response inhibition in adolescents diagnosed with attention deficit hyperactivity disorder during childhood: An event-related fMRI study. *American Journal of Psychiatry*, 161(9), 1650–1657. [PubMed: 15337656]
- Scoville WB, & Milner B (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, 20(1), 11–21. doi:10.1136/jnnp.20.1.11
- Shad MU, Muddasani S, Rao U (2012) Gray matter differences between healthy and depressed adolescents: A voxel-based morphometry study. *Journal of Child And Adolescent Psychopharmacology*. 22(3), 190–197. doi: 10.1089/cap.2011.0005. [PubMed: 22537357]
- Sled JG, Zijdenbos AP, & Evans AC (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1), 87–97. doi: 10.1109/42.668698 [PubMed: 9617910]
- Surguladze S, Russell T, Kucharska-Pietura K, Travis MJ, Giampietro V, David AS, & Phillips ML (2006). A reversal of the normal pattern of parahippocampal response to neutral and fearful faces is associated with reality distortion in schizophrenia. *Biological Psychiatry*, 60(5), 423–31. doi: 10.1016/j.biopsych.2005.11.021 [PubMed: 16487943]
- Talairach J, & Tournoux P (1988). Co-planar stereotactic atlas of the human brain 3dimensional proportional system: An approach to cerebral imaging. New York: Theime Medical Publishers.
- Toplak ME, Bucciarelli SM, Jain U, & Tannock R (2009). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with

- attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology*, 15(1), 53–72. doi: 10.1080/09297040802070929 [PubMed: 18608232]
- Tulving E (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, (53), 1–25. doi:10.1146/annurev.psych.53.100901.135114
- Waber DP, De Moor C, Forbes PW, Almli CR, Botteron KN, Leonard G, ... & Brain Development Cooperative Group (2007). The NIH MRI study of normal brain development: Performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *Journal of the International Neuropsychological Society*, 13(5), 729–46. doi:10.1017/S1355617707070841 [PubMed: 17511896]
- Waber DP, Forbes PW, Almli CR, Blood EA, & Brain Development Cooperative Group (2012). Four-year longitudinal performance of a populationbased sample of healthy children on a neuropsychological battery: The NIH MRI study of normal brain development. *Journal of the International Neuropsychological Society*, 18, 179–190. doi:10.1017/S1355617711001536 [PubMed: 22364826]
- Wechsler D (1991). *Wechsler Intelligence Scale for Children (Third Edition)*. New York: Psychological Corporation.
- Worsley KJ, Taylor JE, Tomaiuolo F, & Lerch J (2004). Unified univariate and multivariate random field theory. *NeuroImage*, 23 Suppl 1, S189–95. doi:10.1016/j.neuroimage.2004.07.026 [PubMed: 15501088]
- Zijdenbos AP, Forghani R, & Evans AC (2002). Automatic “pipeline” analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis. *IEEE transactions on medical imaging*, 21(10), 1280–91. doi:10.1109/TMI.2002.806283 [PubMed: 12585710]

Figure 1.

a. Associations between cortical thickness (CT) and BRIEF Working Memory (WM). *Figure 1b.* Associations between CT and BRIEF Emotional Control. Both figures show negative contrasts in a first-order model controlling for age, gender, and the proxy measure of brain volume (pBV). RFT-corrected, $n=615$. Blue areas are significant at the cluster level, while red and yellow indicate significance at the vertices. *Figure 1c.* Scatter plot of BRIEF WM against CT (mm) in a left parahippocampal gyrus cluster. Controlled for age, gender, and pBV. $n=615$.

Table 1

Descriptive statistics. Age, gender, handedness, and neuropsychological variables are reported for the longitudinal sample (N=663). Household income, maternal education and race are reported for the first visit (N=347). Statistics were comparable for the subsample with cortical thickness measures (Observations, N=615; Number of Participants, N=350).

Number of Observations (Number of Participants)	663 (347)
1 scan, <i>n</i> (%)	115 (33.1%)
2 scans, <i>n</i> (%)	148 (42.7%)
3 scans, <i>n</i> (%)	84 (24.2%)
Mean age in years (range)	11.5 (6.0 – 16.9)
Female, <i>n</i> (%)	360 (54.3%)
Non Right-handed, <i>n</i> (%)	72 (10.9%)
Household Income	
Less than \$50,000	97 (27.9%)
\$50,000 to \$100,000	173 (49.9%)
Over \$100,000	77 (22.2%)
Maternal Education	
Less than high school	4 (1.1%)
High school	46 (13.3%)
Some college	104 (30.1%)
College degree	111 (32.1%)
Some graduate school	19 (5.5%)
Graduate school degree	62 (17.9%)
Race	Total ^a (both parents/ one parent)
White, Non-Hispanic	298 (251/47)
Black	37 (28/9)
Hispanic	46 (13/33)
Asian	9 (3/6)
Native Hawaiian/ Pacific Islander	
American Indian or Alaskan Native	6 (0/6)
BRIEF ^b	
Working Memory, Mean (range)	47.6(35–80)
Inhibition, Mean (range)	46.9(36–83)
Emotional Control, Mean (range)	44.7(35–75)
CANTAB Spatial Working Memory	
Total Errors, Mean (range)	32.6 (0–19)
WISC-III Digit Span Raw Score	
Backward, Mean (range)	5.6 (1–14)

^aTotal number of children with one or both parents in the racial/ethnic category

^bBRIEF T-scores are presented here, but raw scores are used in all other analyses.

Table 2

Mixed models with standardized betas showing associations among behavioral and performance measures.

		BRIEF Working Memory	BRIEF Emotional Control	BRIEF Inhibition	WISC-III Digit Span Backward
BRIEF					
Emotional	β	0.39			
Control	SE	0.03			
	<i>p</i>	0.001			
BRIEF					
Inhibition	β	0.51	0.57		
	SE	0.03	0.03		
	<i>p</i>	<0.001	<0.001		
WISC-III	β	-0.07	-0.06	-0.04	
Digit Span	SE	0.04	0.04	0.04	
Backward	<i>p</i>	0.06	0.12	0.26	
CANTAB Spatial	β	0.15	0.11	0.09	-0.16
Working	SE	0.04	0.04	0.04	0.04
Memory	<i>p</i>	<0.001	0.01	0.04	<0.001