Working Memory Assessment: Construct Validity of the Brown-Peterson Test

Marie Geurten[1], Eric Vincent[2], Martial Van der Linden[3], Françoise Coyette[4], & Thierry Meulemans[1]

[1]University of Liège, Psychology and Neuroscience of Cognition Unit, Belgium

[2]Neuropsychological Rehabilitation Unit, University Hospital of Liège, Belgium

[3]University of Geneva, Switzerland

[4]Neuropsychological Rehabilitation Unit, Saint-Luc Hospital, Belgium

Running head: Validity of the Brown-Peterson Test

Correspondence concerning this article should be addressed to Marie Geurten, University of Liège B33 Trifacultaire – Quartier Agora, Place des Orateurs 1, 4000 Liège – Belgium; E-mail: mgeurten@ulg.ac.be; Phone: +32 4 366 59 43

Eric Vincent, University of Liège; Quai Godefroid Kurth 45, 4020 Liège – Belgium; Email: eric.vincent@chu.ulg.ac.be; Phone: +32 4 270 32 73

Martial Van der Linden, University of Geneva; Phone: +41 22 379 93 41; Email: Martial.vanderlinden@unige.ch

Françoise Coyette, University of Louvain-la-Neuve ; Avenue Hippocrate 10, 1200 Bruxelles ; Email: francoise.coyette@uclouvain.be

Thierry Meulemans, University of Liège B33 Trifacultaire - Quartier Agora, Place des Orateurs, 1, 4000 Liège - Belgium; Phone: +32 4 366 23 74; E-mail: thierry.meulemans@ulg.ac.be

Abstract

The classical Brown-Peterson task is generally supposed to assess working memory capacities. To date, however, the construct validity of the task remains mostly unexamined. In this context, the aim of the present study was to demonstrate the convergent and the divergent validity as well as the clinical and the developmental sensitivity of a computerized version of the Brown-Peterson test. A group of 726 French-speaking participants aged from 18 to 86 years and 47 patients who had sustained a traumatic brain injury (TBI) were administered the Brown-Peterson task and various other cognitive tasks assessing executive functioning, verbal and visual memory, or processing speed. The correlation analyses revealed the good convergent of the task, which was shown to be able to distinguish between participants with TBI and control participants. We found an effect of age and education level on the different scores recorded for the Brown-Peterson test. Normative data taking into account the influence of the latter variables were thus provided. On the whole, these findings seem to confirm the validity of the Brown-Peterson task as a tool to assess working memory abilities.

Keywords: Brown-Peterson task; Working memory; Construct validity; Normative data

Working Memory Assessment: Construct Validity of the Brown-Peterson Test

The Brown-Peterson task was originally created by Brown (1958) and elaborated by Peterson and Peterson (1959) to examine the hypothesis that memory trace decay is a cause of forgetting. Specifically, the Brown-Peterson technique was designed to measure the effects of both interference and delay on short-term memory performance. Over time, various versions of the task have been developed (e.g., Belleville, Chatelois, Fontaine, & Peretz, 2002; Mertens, Gagnon, Coulombe, & Messier, 2006; Ryan & Butters, 1980; for a recent adaptation in children, see Rai & Harris, 2013). However, the basic principle of the Brown-Peterson paradigm always remains the same. Typically, participants are instructed to recall a few items (usually, three consonants) after variable delays (usually, from 0 to approximately 30 seconds) during which an interference task must be performed (e.g., mental addition, repetition of numbers, counting backward).

In addition to the memory component of the task, which requires participants to retain information in memory for a short period, several neuropsychological studies have demonstrated that executive functions are also involved in the Brown-Peterson task. For instance, Korsakoff patients' deficits in the interference condition of the Brown-Peterson task were found to be more related to their executive dysfunctions than to their memory problems (Leng & Parkin, 1989). Similarly, patients with frontal lesions have been shown to perform poorly on the Brown-Peterson test (Kapur, 1988). The fact that both memory and executive skills seem to be required to complete the task generally leads researchers and practitioners to consider the Brown-Peterson procedure as a measure of working memory abilities (Bherer, Belleville, & Peretz, 2001; Fleming, Goldberg, Gold, & Weinberger, 1995; Floden, Stuss, & Craik, 2000; Mertens et al., 2006). Consistent with this view, Mertens et al. (2006) revealed significant correlations between the Brown-Peterson task and other classical measures of working memory, such as the backward digit span and the letter-number sequencing of the Wechsler Memory Scale (WMS-III; Wechsler, 1997).

To date, however, even though the Brown-Peterson procedure is widely used to assess working memory skills in both clinical (Anil et al., 2003; Belleville et al., 2002; Callahan et al., 2014) and experimental (Fleming et al., 1995; Kane & Engle, 2000; Neath & Brown, 2012; Stuss, Stethem, & Poirier, 1987; Wang, Ren, Li, & Schweizer, 2015) settings, the task's construct validity has not been examined in detail. Consequently, it is still unclear whether the Brown-Peterson test is truly a good measure of working memory abilities. In the same vein, no research has been conducted to examine the task's ability to discriminate between patients with working memory deficits and healthy participants, which would indicate that the task is a useful and sensitive instrument to help neuropsychologist to describe the specific cognitive situation of their patients. Moreover, authors who have investigated the effects of age and education level on the task have reported conflicting results. Specifically, several studies found no effect of age on Brown-Peterson measures (Bherer et al., 2001; Stuss et al., 1987), while others showed a negative association between chronological age and performance (Anil et al., 2003; Callahan et al., 2014; Floden et al., 2000; Mertens et al., 2006). Similarly, some authors found that participants' level of education had a positive influence on their Brown-Peterson scores (Anil et al., 2003; Bherer et al., 2001; Callahan et al., 2014; Morrow & Ryan, 2002) while others found no such effect (Boone, 1999). However, the lack of age and educational effects probably results from the small sample sizes used in previous studies. The above-mentioned experiments never included more than 90 subjects (for an exception, see Callahan et al., 2014, who established normative data for the Brown-Peterson task with a sample of 595 elderly participants).

For these reasons, the first aim of the present study is to provide evidence for the convergent and divergent validity, as well as the clinical and age sensitivity, of a computerized version of the Brown-Peterson test. To test the convergent and divergent validity, a battery of neuropsychological tests commonly used by clinicians in their day-to-day practice was administered to a large sample of healthy participants. Correlation analyses were then carried out to support the Brown-Peterson test's validity as a measure of working memory abilities. Specifically, we expect to observe higher correlations with tasks assessing verbal working memory and executive abilities than with tasks assessing other cognitive functions (such as long term memory tests).

Secondly, we also examined whether the Brown-Peterson technique was able to discriminate between patients with impaired working memory abilities and healthy participants. Working memory deficits are frequently reported following a traumatic brain injury (TBI) (e.g., McAllister, Flashman, McDonald, & Saykin, 2006; McDowell, Whyte, & D'Esposito, 1997; Perlstein et al., 2004). Accordingly, a group of patients who had sustained mild to severe TBI was included in the study so that we could test the Brown-Peterson test's discriminant validity.

Finally, the influence of participants' age and education level was also investigated. For this purpose, we recruited a larger number of participants from a wider range of ages than those enrolled in previous studies. If the Brown-Peterson has a good age and educational sensitivity, we expect to demonstrate a decrease in participants' performance with age and an increase in their performance with the number of years of education. Depending on the results of the statistical analyses, normative data taking into account the influence of participants' age and education level will be extracted for the Brown-Peterson task.

Methods

Participants

Healthy group. A total of 726 French-speaking individuals whose ages ranged from 18 to 86 years (Mean = 49.78 years, SD = 19.94) participated in the study. The number of years of education of participants ranged from 3 to 26 years (Mean = 12.64, SD = 3.51). Sixteen additional participants were tested but not included in the analyses because they were not able to complete all the tasks. All subjects had a normal or corrected-to-normal vision and hearing. None had an established diagnosis of learning disability or a history of neurological or psychiatric disorders. Fifty-four percent of the subjects were females. The sample was recruited from the French community of Liège, Namur, and Luxembourg in Belgium over a ten-year period from 2005 to 2015. They all volunteered to participate. No remuneration was provided. Demographic variables are displayed in Table 1.

Clinical group. A group of 47 French-speaking individuals (18 females) who had sustained mild (n = 23), moderate (n = 9), or severe TBI (n = 15) from closed head trauma participated in the study. They were recruited from the Neuropsychological Rehabilitation Unit of the University Hospital in Liège, Belgium. None of them was involved in litigation regarding their disability claims. Injury severity was determined by the Glasgow Coma Scale score (GCS; Teasdale & Jennett, 1974) on admission (severe ≤ 8; moderate > 8 and < 13; mild ≥ 13) or by the duration of unconsciousness (severe > 6 hours; moderate < 6 hours and > 1 hour; mild < ½ hour). Exclusion criteria included a history of psychiatric disorder (including substance abuse disorder), an established diagnosis of developmental disability or mental deficiency, and a pretrauma history of neurological disorder. Each patient was matched as closely as possible with a healthy participant for sex, age and education level, all *ps*>.86. To this end, the Mahalanobis distance matching (MDM) method was used. Demographic and clinical data on the patients and their matching controls are displayed in Table 1.

< Table 1 >

Material

Brown-Peterson test. A new computerized version of the classical Brown-Peterson test was administered to all subjects (Brown, 1958; Peterson & Peterson, 1959). Figure 1 illustrates the details of the procedure. Three consonants were visually presented in the center of the screen at a rate of one per second. The three consonants were chosen so that they did not form any common acronyms. Participants were instructed to read each letter aloud and to keep them in mind in order to be able to recall them after an unannounced delay of 0, 5, 10, or 20 seconds. These intervals were chosen because they are used in a common clinical version of the Brown-Peterson task (Coyette et al., 2003; Morris, 1986). During the delay period, participants were asked to perform an interference task. Specifically, several pairs of numbers were verbally presented to subjects (e.g., ‘7-9’) who were required to repeat them in the reverse order (e.g., ‘9-7’). Once the delay period had elapsed, a response box appeared in the center of the computer screen and the participants were instructed to recall the three consonants in the correct order. For the 0-second interval, recall occurred immediately after the presentation of the third consonant. The computerized version of the task allowed for a more rigorous control of each delay period. A total of 24 trials (6 trials per time delay) were administered to each participant in a pseudo-random order – that is, the different trigrams and the different delay periods were presented arbitrarily but in the same order for each participant. A fixed-order procedure was preferred because it allowed the data to be used when comparing the performance of participants in a clinical context. Before the test, a practice trial was administered to familiarize the subjects with the procedure. The proportion of consonants recalled in the correct order for each time delay (i.e., 0, 5, 10, and 20 seconds) was used as dependent variable.

<Figure 1>

Other cognitive tasks. A neuropsychological battery of tests was administered to examine the Brown-Peterson task's convergent and divergent validity. All of the selected tasks are frequently used by clinical neuropsychologists in their day-to-day practice (Strauss, Sherman, & Spreen, 2006). For convergent validity, we chose tasks that have been demonstrated to appraise short-term memory, working memory, and executive capacities (i.e., cognitive functions that are commonly supposed to be involved in the Brown-Peterson task). More specifically, the *digit span subtest* of the WMS-III (Wechsler, 1997) was used to assess short-term memory capacities. Forward and backward digit spans were used as outcome measures in our analyses. A computerized version of the *Paced Auditory Serial-Addition Task* (PASAT; Gronwall, 1977) was used as a measure of participants' updating abilities (i.e., working memory abilities). In this task, 60 numbers (from 1 to 9) are divided into four trials that differ in terms of the speed with which the numbers are presented (one number every 2.4, 2.0, 1.6, or 1.2 seconds). Subjects are asked to add each number to the one that immediately preceded it. The outcome measures were the number of correct responses for each of the four experimental trials. Executive functions were assessed using an interference score computed for the *Stroop test* (Regard, 1981). This score was obtained by subtracting the median reaction time in the naming part of the task from the median reaction time in the interference part of the task (interference index; see Meulemans, 2008).

For divergent validity, we used tasks assessing long-term memory, processing speed, and general vocabulary abilities. Verbal long-term memory abilities were appraised using the French version of the *Buschke Selective Reminding Test* (SRT; Buschke & Fuld, 1974; Van der Linden et al., 2004). The dependent variable was the total number of items recalled across the ten trials of the task. Visual long-term memory abilities were tested using a computerized version of the *Visual Pattern* task (Della Sala, Gray, Baddeley, & Wilson, 1997) during which participants had to learn a matrix pattern that exceeded their working memory span by three items. The number of trials to recall the matrix pattern was used as the dependent measure. The total times to complete the color and word segments of the *Stroop Test* were used as measures of processing speed. Finally, the French version of the *National Adult Reading Test* (fNART; Blair & Spreen, 1989) was administered to assess general vocabulary abilities. In this task, participants were required to read 34 irregular French nouns. The dependent variable was the number of words correctly pronounced.

Procedure

Both patients and healthy participants were enrolled and tested following written informed consent and with the agreement of the participating institution's ethics committee. Healthy subjects were tested individually at home. Patients were tested in the institution where they were recruited. All the computerized tasks were administered using version 9.5 of Toolbook software (SumTotal Systems Inc., Gainesville, Florida). Healthy subjects participated in a session lasting approximately 60 minutes during which half of them were given the tasks in the following random order: (1) the digit span task, (2) the Brown-Peterson task, (3) the Buschke SRT, (4) the visual pattern test, (5) the PASAT, (6) the Stroop test, and (7) the fNART. The other half of the participants completed the tasks in the opposite order. The patients were given the Brown-Peterson task as a part of a more general neuropsychological evaluation that was carried out after their admission to the Neuropsychological Rehabilitation Unit in Liège. The latter evaluation was individualized for each patient depending on their cognitive complaints.

Results

Data Analyses

Statistical analyses were done using *Statistica* software version 10 (Hill & Lewicki, 2007). The first goal of our study was to examine the construct validity and the age and clinical sensitivity of the Brown-Peterson task. Construct validity was tested using the known-group technique, which involves administering the measurement instrument to groups expected to differ due to known characteristics (Portney & Watkins, 2000). To this end, we carried out analyses of covariance (ANCOVAs) to examine whether the task was sensitive to delay period, age, and educational differences. The delay (0, 5, 10, or 20-second) was included as a within-subject factor. Participants' chronological age (in years) and their number of years of education were treated as continuous variables and were thus included as covariates in the following analyses. Furthermore, correlation analyses between the four measures recorded from the Brown-Peterson test (i.e., the rate of correct responses for the 0, 5, 10 and 20-second time delays) and the scores recorded for the other neuropsychological tests were carried out to check the task's convergent and divergent validity. Finally, Receiver Operating Characteristic (ROC) curve analyses were used to assess the ability of the different levels of difficulty of the Brown-Peterson test to discriminate between patients with TBI and healthy participants. Significant level was set at .05, unless otherwise noted. Preliminary analyses no order or gender effect on any of the dependent variables, all $ps > .13$.

Delay, Age, and Education Effects

ANCOVAs were conducted to examine the effect of the delay period (0, 5, 10, or 20 seconds) on performance. The influence of age and education level was also investigated. Our results revealed a medium main effect of the delay period, $F(3, 2166) = 32.86$, $p<.001$, $\eta^2= .11$[1]. Specifically, post hoc analyses (Tukey (1949); Honestly Significant Differences' [HSD] test) indicated that performance decreased significantly each time the interval increased, all *ps*<.016 (means ranged from .98 to .77). We also found a medium main effect of age, $F(3, 722) = 139.04$, $p<.001$, $\eta^2 = .11$, and a medium effect of education level, $F(1, 722) = 155.54$, $p<.001$, $\eta^2 = .07$. A significant delay x age, $F(3, 2166) = 58.85$, $p<.001$, $\eta^2 = .05$, and delay x education, $F(3, 2166) = 45.10$, $p<.001$, $\eta^2 = .09$, interactions were also found. For this reason, we chose to conduct multiple regression analyses to examine the effect of age and education on participants' performance for the four different delay periods of the Brown-Peterson task. The results of these analyses are presented in the following section. No other main or interaction effect reached significance, *Fs*<1.10.

Age and Education Sensitivity

Multiple regression analyses were conducted to examine the effects of age and education level on each of the four scores for the Brown-Peterson task. Education level was found to predict participants' score for the 0-second, $\beta = 0.199$, 95% CI [.13, .27], $p<.001$, $R^2 = .03$; for the 5-second, $\beta = 0.353$, 95% CI [.29, .42], $p<.001$, $R^2 = .12$; for the 10-second, $\beta = 0.314$, 95% CI [.25, .38], $p<.001$, $R^2 = .11$; and for the 20-second time delays, $\beta = 0.345$, 95% CI [.28, .41], $p<.001$, $R^2 = .12$. These results indicated that participants with a higher education level performed better at the Brown-Peterson task, but this effect was larger for 5-second, 10-second,

[1] Cohen (1988) defined effect size of partial eta squared as "small, $\eta^2 = .01$," "medium, $\eta^2 = .06$," and "large, $\eta^2 = .13$."

and 20-second intervals than for 0-second interval. This pattern suggests that educational level makes little difference for the 0-second delay, but have a larger influence on the three longer delays.

Chronological age was not found to predict participants' performance for the 0-second interval. However, age was found to predict participants' score for the 5-second, $\beta = -0.273$, 95% CI [-.34, -.21], $p<.001$, $R^2 = .09$; for the 10-second, $\beta = -0.354$, 95% CI [-.42, -.29], $p<.001$, $R^2 = .14$; and for the 20-second intervals, $\beta = -0.344$, 95% CI [-.41, -.28], $p<.001$, $R^2 = .13$. The negative Beta coefficient consistently showed that performance decreased with age. On the whole, these results revealed the scale to have good age sensitivity, except for the 0-second delay period. Given the number of statistical analyses and the need the balance the amount of type 1 and type 2 errors, we employed a false discovery rate method for multiple testing. The false discovery rate control the expected proportion of falsely rejected null hypothesis (Benjamini & Hochberg, 1995). In the present case, we ordered our 15 p-values from smallest to largest, the p-value associated with the first finding that could be considered as significant was .046.

Normative data for age and level of education were constructed on the basis of these results. Specifically, as statistical analyses revealed a significant effect of age and education level for all the Brown-Peterson's scores, we regrouped participants into two education categories: ED = 1: 12 years of education or less and ED = 2: more than 12 years of education, respectively. Our sample was divided into four age groups: 18–39 (n = 241; early adulthood), 40–59 (n = 185; middle adulthood), 60–69 (n = 159; late adulthood), and 70–86 years (n = 141; aging). These normative data are presented in the Appendix.

Construct Validity

Convergent validity. To examine the Brown-Peterson task's validity as a test of working memory, partial correlation analyses were carried out between the four scores on the Brown-Peterson task and the scores for the digit span test, the PASAT, and the interference portion of the Stroop test. As previous analyses demonstrated the effect of age and education level on the Brown-Peterson task, the influence of chronological age (in years) and the number of years of education was controlled for. Chronological age and education level were thus used as continuous variables in all the correlation analyses. Partial correlations with the different Brown-Peterson scores are displayed in Table 2. Correlation matrix between other cognitive scores is provided as supplemental material. The results of the statistical analyses indicated no correlation between the score for the 0-second delay of the Brown-Peterson task and the other cognitive measures, all $r_p<.18$. However, significant medium correlations were found between the score for the 5-second interval of the Brown-Peterson task and the score for the 2-second trial of the PASAT, $r_p = .25$, $p = .01$, the score for the 1.6-second trial of the PASAT, $r_p = .25$, $p = .01$, and the backward digit span score, $r_p = .28$, $p<.001$[2]. Similarly, significant partial correlations were found between the score for the 10-second interval of the Brown-Peterson task and the score for the 1.6-second trial of the PASAT, $r_p = .32$, $p = .001$, the score for the interference part of the Stroop, $r_p = –.26$, $p = .006$, and the scores for the forward, $r_p = .26$, $p = .007$, and backward, $r_p = .41$, $p<.001$, digit spans. Finally, significant correlations were found between the score for the 20-second delay of the Brown-Peterson task and the 2-second trial of the PASAT, $r_p = .37$, $p<.001$, the 1.6-second trial of the PASAT, $r_p = .35$, $p<.001$, the score for the interference part of the Stroop, $r_p = –.26$, $p = .006$, and the backward digit span, $r_p = .27$, $p = .006$.

[2] Cohen (1988) defined effect size correlations as "small, r = .10," "medium, r = .24," and "large, r = .37."

< Table 2 >

Divergent validity. According to our hypothesis, the Brown-Peterson task is primarily a test of working memory. To investigate the divergent validity of the task (i.e., to determine that the Brown-Peterson task is not too similar to another test), partial correlation analyses (controlling for age and education level) were carried out between scores on the Brown-Peterson task and the scores for the SRT, the visual pattern task, the color and word parts of the Stroop, and the fNART. These correlations were expected to be smaller than those obtained for the tasks selected to examine the convergent validity of the Brown-Peterson task. Once again, no correlation was found between the 0-second interval of the Brown-Peterson test and the other cognitive measures, $r_p<.13$. As Table 2 shows, significant correlations were found between the score for the 5-second interval and scores for the color part of the Stroop, $r_p = -.28$, $p = .003$, and the fNART, $r_p = .33$, $p<.001$. Similarly, significant correlations were found between the score for the 10-second interval and scores for the color portion of the Stroop, $r_p = -.31$, $p = .001$, and the fNART, $r_p = .46$, $p<.001$. Finally, the same correlations were revealed between the score for the 20-second interval and the color part of the Stroop, $r_p = -.31$, $p = .001$, and the fNART, $r_p = .45$, $p<.001$. No other correlation reached significance, all $r_p<.19$. Once again, a false discovery rate method (Benjamini-Hochberg procedure) was used to balance the amount of type 1 and type 2 errors in these analyses. Specifically, we ordered our 48 p-values from smallest to largest, the p-value associated with the first finding that could be considered as significant was .015.

Discriminant Validity

First, the ability of the different scores recorded for the Brown-Peterson task to discriminate between patients who had sustained mild to severe TBI and healthy participants matched for age and education level was explored. The results of the 2 (Group) x 4 (Delay period) ANOVA showed a large main effect of group, $F(1,92) = 35.54$, $p<.001$, $\eta^2 = .28$. A large effect of delay was also found, $F(3,276) = 75.79$, $p<.001$, $\eta^2 = .45$. Post hoc analyses (Tukey's test) indicated that performance decreased significantly each time the interval increased, all $ps<.001$, except between the two longer intervals, $p = .37$. Finally, a group x delay period interaction of large size was revealed, $F(3,276) = 16.13$, $p<.001$, $\eta^2 = .15$. Specifically, post hoc analyses (Tukey's test) indicated significant differences between groups for the 5-second interval (mean = .91 and .73 for healthy participants and patients, respectively); the 10-second interval (mean = .85 and .62); and the 20-second interval (mean = .84 and .57), all $ps<001$, but not for the 0-second interval, $p = .99$.

The ROC curve method was used to further investigate the discriminant validity of the four scores of the Brown-Peterson task. Conventionally, the area under the curve would be 1.0 for a measure that discriminates perfectly between patients and healthy participants, and .50 for a measure that discriminates with an accuracy of no better than chance. The results revealed that the area under the curve was .52 for the 0-second interval, .75 for the 5-second interval, .77 for the 10-second interval, and .80 for the 20-second interval. The positive predictive values (PPV; i.e., the proportion of people with a positive test result who actually have the condition) and the negative predictive values (NPV; i.e., the proportion of people with a negative test result who do not have the condition) for the best cutoff scores are presented in Table 3. These values were determined so that the optimum balance could be found between sensitivity and specificity.

< Table 3 >

However, it is possible that patients who have sustained mild head injury may have only very slight difficulties with complex cognitive tasks when compared to healthy people (e.g., Binder, Rohling, & Larrabee, 1997). The lack of major cognitive impairments in patients with mild TBI might have artificially reduced the task's clinical sensitivity. For this reason, we chose to investigate the ability of the Brown-Peterson scores to discriminate between patients who had sustained moderate to severe TBI (n = 24) and healthy participants. The results of the 2 (Group) x 4 (Delay period) ANOVA were the same than those obtained when all the patients where included in the analysis.

The results of the ROC curve method indicated that the area under the curve was .48 for the 0-second interval of the Brown-Peterson test, .79 for the 5-second interval, .77 for the 10-second interval, and .87 for the 20-second interval. The PPV and the NPV for the best cutoff scores are presented in Table 3.

Discussion

Although the Brown-Peterson task is frequently used in both clinical and experimental settings (e.g., Belleville et al., 2002; Stuss et al., 1987), no study had previously been carried out to fully examine the test's psychometric properties or establish its sensitivity to working memory impairments. In this context, the primary aim of our study was to provide evidence for the construct validity of the Brown-Peterson procedure. To this end, a computerized version of the task was employed. Our study produced several interesting findings.

Convergent Validity

Firstly, the significant partial correlations found between the different scores on the Brown-Peterson test and the interference part of the Stroop indicate that the Brown-Peterson task involves executive abilities. The significant correlation between the 10-second delay of the Brown-Peterson task and the forward digit span suggests that the Brown-Peterson test may partially involve short-term memory. Combined with the significant correlations found with the backward digit span and the scores for the different trials of the PASAT, this seems to confirm that the Brown-Peterson procedure is an appropriate measure of working memory (e.g., Bherer et al., 2001; Fleming et al., 1995; Mertens et al., 2006). However, although significant, the correlations between the different scores recorded for the Brown-Peterson task and the other working memory measures are only moderate. This pattern could be explained by the fact that the Brown-Peterson test delivers information about participants' cognitive efficiency that differs the information provided by the other neuropsychological tests. Indeed, while other short-term memory tests used in this study allow one to appraise either the processing or the storage function of working memory, the Brown-Peterson task seems to assess both components (for a theoretical model of how these two functions interact during a complex span task similar to the Brown-Peterson, see Barrouillet, Bernardin, & Camos, 2004). According to Barrouillet et al. (2004), the concurrent examination of both components provides practitioners with additional information about the nature of their patients' cognitive problems, which suggests that the task should be used concurrently with other classical tests of working memory during neuropsychological evaluations.

Divergent Validity

The absence of correlation between the Brown-Peterson task and either the Buschke SRT or the visual pattern test suggests that the Brown-Peterson task specifically assesses short-term rather than long-term memory abilities. However, the significant relationships found with the color part of the Stroop and the fNART seem to indicate that the Brown-Peterson also requires good processing speed and vocabulary skills. Similar results were already found by Stuss et al. (1989; see also Floden et al., 2000). This pattern might be explained if one assumes that participants with good language abilities and/or processing speed are more likely to implement strategies that will help them to retain the three consonants during the delay interval (for studies demonstrating the influence of strategies on memory performance, see Dunlosky & Kane, 2007; Geurten, Catale, & Meulemans, 2015; Geurten, Lejeune, & Meulemans, 2015; McNamara & Scott, 2001). For example, a participant who has a good vocabulary and is able to quickly process information might be more likely to find a word that comprises the three to-be-remembered-letters in the correct order (e.g., SPN = Spin), increasing the probability of recalling it on the upcoming test. Nonetheless, other studies should be carried out to confirm this hypothesis and investigate (a) whether the implementation of strategies truly accounts for performance on the Brown-Peterson task, and (b) which cognitive variables are involved in the implementation of these strategies.

Age and Educational Sensitivity

From a developmental point of view, it appears that performance on the different intervals of the Brown-Peterson task (except for the 0-second interval) decreases with age. These findings are consistent with the results of recent studies carried out using the Brown-Peterson technique that found a negative association between participants' age and their scores for the Brown-Peterson task (e.g., Callahan et al., 2014; Mertens et al., 2006). More generally, these results are also coherent with the lifespan trajectory of working memory. According to various theoretical models, working memory abilities are generally expected to decrease with age (e.g., Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Salthouse, 1994). Moreover, the influence of participants' education level has also been established, suggesting that the lack of results observed in some previous studies probably could be due to small samples sizes rather than from an absence of effect. Thus, normative data adjusted for age and education level are provided in the Appendix. Interestingly, although age and education level were shown to influence participants' performance to the Brown-Peterson task, no age x education level interaction was found. Even if an absence of significant result is not sufficient in itself to conclude to an absence of effect, this lack of finding seems to indicate that a high level of education does not protect older participants from a reduction of performance as compared to young participants.

Discriminant Validity

Finally, we examined the ability of the different Brown-Peterson task scores to distinguish between patients with TBI and their matched controls. Comparisons revealed significant differences between the two groups on the 5-, 10-, and 20-second intervals of the task. Furthermore, ROC analyses revealed a good level of specificity and sensitivity, particularly for the 20-second interval. Importantly, we should note that, although impaired working memory is very frequently reported following a TBI (e.g., McAllister et al., 2006; McDowell et al., 1997; Perlstein et al., 2004), TBIs are not always followed by such impairments. In this context, we may have artificially underestimated the clinical sensitivity of the task (which is already good) by including participants who did not have significant working memory problems. On the whole, however, our findings confirm that the Brown-Peterson task can be considered as a useful tool to examine working memory problems in participants with TBI.

In conclusion, although other investigations should be conducted to further examine some aspects of its psychometric properties (e.g., the convergent validity and the divergent validity were not totally satisfactory), the analysis of the construct validity of the computerized version of the Brown-Peterson task suggests that the task could be a proper measure of working memory. Furthermore, we also established that this task has good clinical and age sensitivity of the three longer intervals of the task. Moreover, new normative data taking the influence of age and education level into account were obtained.

References

Anil, A. E., Kivircik, B. B., Batur, S., Kabakçi, E., Kitis, A., Güven, E., . . . Arkar, H. (2003). The Turkish version of the auditory consonant trigram test as a measure of working memory: A normative study. *The Clinical Neuropsychologist, 17*, 159–169. doi:10.1076/clin.17.2.159.16510

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General, 133*, 83–100. doi:10.1037/0096-3445.133.1.83

Belleville, S., Chatelois, J., Fontaine, F., & Peretz, I. (2002). *Mémoria: Batterie informatisée d'évaluation de la mémoire pour Mac et PC.* Montreal, QC: Institut Universitaire de Gériatrie de Montréal.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289–300. doi:10.2307/2346101

Bherer, L., Belleville, S., & Peretz, I. (2001). Education, age, and the Brown-Peterson technique. *Developmental Neuropsychology, 19*, 237–251. doi:10.1207/S15326942DN1903_1

Binder, L. M., Rohling, M. L., & Larrabee, G. J. (1997). A review of mild head trauma. part I: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology, 19,* 421–431. doi:10.1080/01688639708403870

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the national adult reading test. *Clinical Neuropsychologist, 3,* 129–136. doi:10.1080/13854048908403285

Boone, K. B. (1999). Neuropsychological assessment of executive functions: Impact of age, education, gender, intellectual level, and vascular status on executive test scores. In B. L.

Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (pp. 247–260). New York: Guilford Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12–21. doi:10.1080/17470215808416249

Buschke, H., & Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology, 24*, 1019–1025.

Callahan, B. L., Belleville, S., Ferland, G., Potvin, O., Tremblay, M.-P., Hudon, C., & Macoir, J. (2014). Normative data for a computer-assisted version of the auditory three-consonant Brown-Peterson paradigm in the elderly French-Quebec population. *The Clinical Neuropsychologist, 28*, 317–332. doi:10.1080/13854046.2013.873082

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Coyette, F., Verreckt, E., Seron, X., Desgranges, B., Adam, S., & Eustache, F. (2003). Rééducation des troubles de l'administrateur central de la mémoire de travail. In T. Meulemans, B. Desgranges, S. Adam, & F. Eustache (Eds.), *Evaluation et prise en charge des troubles mnésiques* (pp. 293–314). Marseille: Solal.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). *Visual pattern test.* Bury St Edmunds: Thames Valley Test Company.

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology, 60*, 1227–1245. doi:10.1080/17470210600926075

Fleming, K., Goldberg, T. E., Gold, J. M., & Weinberger, D. R. (1995). Verbal working memory dysfunction in schizophrenia: Use of a Brown-Peterson paradigm. *Psychiatry Research, 56*, 155–161. doi:10.1016/0165-1781(95)02589-3

Floden, D., Stuss, D. T., & Craik, F. I. M. (2000). Age differences in performance on two versions of the Brown-Peterson task. *Aging, Neuropsychology, and Cognition*, *7,* 245–259. doi:10.1076/anec.7.4.245.795

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience, 8,* 1298–1300.

Geurten, M., Catale, C., & Meulemans, T. (2015). When children's knowledge of memory improves children's performance in memory. *Applied Cognitive Psychology, 29,* 244–252. doi:10.1002/acp.3102

Geurten, M., Lejeune, C., & Meulemans, T. (2015). Time's up! Involvement of metamemory knowledge, executive functions, and time monitoring in children's prospective memory performance. *Child Neuropsychology*. Advance online publication. doi:10.1080/09297049.2014.998642

Gronwall, D. M. A. (1977). Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual and Motor Skills, 44*, 367–373. doi:10.2466/pms.1977.44.2.367

Hill, T., & Lewicki, P. (2007). *Statistics methods and applications.* Tulsa, OK: StatSoft, 2007.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 336–358. doi:10.1037/0278-7393.26.2.336

Kapur, N. (1988). *Memory disorders in clinical practice.* London: Butterworth-Heinemann.

Leng, N. R. C., & Parkin, A. J. (1989). Aetiological variation in the amnesic syndrome: Comparisons using the Brown-Peterson task. *Cortex, 25*, 251–259. doi:10.1016/S0010-9452(89)80041-3

McAllister, T. W., Flashman, L. A., McDonald, B. C., & Saykin, A. J. (2006). Mechanisms of working memory dysfunction after mild and moderate TBI: Evidence from functional MRI and neurogenetics. *Journal of Neurotrauma, 23*, 1450–1467. doi:10.1089/neu.2006.23.1450

McDowell, S., Whyte, J., & D’Esposito, M. (1997). Working memory impairments in traumatic brain injury: evidence from a dual-task paradigm. *Neuropsychologia*, *35*, 1341–1353. doi:10.1016/S0028-3932(97)00082-1

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition*, *29*, 10–17.

Mertens, V. B., Gagnon, M., Coulombe, D., & Messier, C. (2006). Exploratory factor analysis of neuropsychological tests and their relationship to the Brown-Peterson task. *Archives of Clinical Neuropsychology, 21*, 733–739. doi:10.1016/j.acn.2006.08.005

Meulemans, T. (2008). La batterie GREFEX. In O. Godefroy & GREFEX (Eds.), *Fonctions exécutives et pathologies neurologiques* (pp. 217–229). Marseille: Solal.

Morris, R. G. (1986). Short-term forgetting in senile dementia of the Alzheimer’s type. *Cognitive Neuropsychology, 3*, 77–97. doi:10.1080/02643298608252670

Morrow, L. A., & Ryan, C. (2002). Normative data for a working memory test: The four word short-term memory test. *The Clinical Neuropsychologist*, *16*, 373–380. doi:10.1076/clin.16.3.373.13850

Neath, I., & Brown, G. D. A. (2012). Arguments against memory trace decay: A SIMPLE account of Baddeley & Scott. *Frontiers in Psychology, 3,* 35. doi:10.3389/fpsyg.2012.00035

Perlstein, W. M., Cole, M. A., Demery, J. A., Seignourel, P. J., Dixit, N. K., Larson, M. J., & Briggs, R. W. (2004). Parametric manipulation of working memory load in traumatic brain injury: Behavioral and neural correlates. *Journal of the International Neuropsychological Society, 10,* 724–741. doi:10.1017/S1355617704105110

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198. doi:10.1037/h0049234

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: applications to practice* (Vol. 2): Prentice Hall Upper Saddle River, NJ.

Rai, M. K., & Harris, R. J. (2013). The Modified Brown-Peterson Task: A Tool to Directly Compare Children and Adult's Working Memory. *The Journal of Genetic Psychology, 174,* 153–169. doi:10.1080/00221325.2011.653839

Regard, M. (1981). *Stroop test – Victoria version.* Victoria, BC: Neuropsychological Laboratory, University of Victoria.

Ryan, C., & Butters, N. (1980). Learning and memory impairments in young and old alcoholics: Evidence for the premature-aging hypothesis. *Alcoholism: Clinical and Experimental Research, 4,* 288–293. doi:10.1111/j.1530-0277.1980.tb04816.x

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology, 8,* 535–543. doi:10.1037/0894-4105.8.4.535

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary.* New York: Oxford University Press.

Stuss, D. T., Stethem, L. L., & Poirier, C. A. (1987). Comparison of three tests of attention and rapid information processing across six age groups. *Clinical Neuropsychologist, 1*, 139–152. doi:10.1080/13854048708520046

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet, 2,* 81–84.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99–114.

Van der Linden, M., Adam, S., Agniel, A., Baisset-Mouly, C., Bardet, F., Coyette, F., . . . Thomas-Antérion, C. (2004). *L'évaluation des troubles de la mémoire: Présentation de quatre tests de mémoire épisodique (avec leur étalonnage).* Marseille: Solal.

Wang, T., Ren, X., Li, X., & Schweizer, K. (2015). The modeling of temporary storage and its effect on fluid intelligence: Evidence from both Brown-Peterson and complex span tasks. *Intelligence, 49,* 84–93. doi:10.1016/j.intell.2015.01.002

Wechsler, D. (1997). *WMS-III: Wechsler memory scale administration and scoring manual.* New York: Psychological Corporation.

Appendix

Practitioners who would like to obtain this computerized version of the Brown-Peterson task are invited to contact the last author. Normative data are presented below.

Means, Standard Deviations, and Percentile Scores for Brown-Peterson Raw Data Depending on Age and Education Level

	Mean	SD	P1	P5	P10	P25	P50	P75	P90	P95
18–39 years – ED = 1 (n = 65)										
0-second	0.97	0.06	0.76	0.83	0.89	1.00	1.00	1.00	1.00	1.00
5-second	0.86	0.17	0.27	0.56	0.67	0.82	0.89	1.00	1.00	1.00
10-second	0.81	0.18	0.40	0.45	0.56	0.67	0.83	0.94	1.00	1.00
20-second	0.82	0.18	0.31	0.50	0.61	0.72	0.83	1.00	1.00	1.00
18–39 years –ED = 2 (n = 176)										
0-second	0.99	0.03	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5-second	0.93	0.10	0.67	0.72	0.78	0.89	1.00	1.00	1.00	1.00
10-second	0.90	0.13	0.44	0.61	0.78	0.83	0.94	1.00	1.00	1.00
20-second	0.88	0.14	0.50	0.56	0.67	0.83	0.94	1.00	1.00	1.00
40-59 years – ED = 1 (n = 72)										
0-second	0.97	0.08	0.69	0.83	0.91	1.00	1.00	1.00	1.00	1.00
5-second	0.81	0.20	0.33	0.41	0.50	0.67	0.89	1.00	1.00	1.00
10-second	0.76	0.18	0.35	0.47	0.50	0.67	0.78	0.89	0.94	1.00
20-second	0.70	0.24	0.20	0.28	0.33	0.56	0.72	0.89	1.00	1.00
40-59 years – ED = 2 (n = 113)										
0-second	0.98	0.04	0.81	0.89	0.94	1.00	1.00	1.00	1.00	1.00
5-second	0.94	0.08	0.76	0.78	0.83	0.89	1.00	1.00	1.00	1.00
10-second	0.86	0.16	0.43	0.55	0.61	0.78	0.94	1.00	1.00	1.00
20-second	0.82	0.16	0.42	0.50	0.61	0.71	0.86	0.94	1.00	1.00
60–69 years – SES = 1 (n = 51)										
0-second	0.98	0.05	0.81	0.86	0.89	1.00	1.00	1.00	1.00	1.00
5-second	0.77	0.21	0.25	0.42	0.50	0.61	0.83	0.94	1.00	1.00

10-second	0.67	0.23	0.11	0.25	0.33	0.56	0.72	0.83	0.89	1.00
20-second	0.64	0.27	0.08	0.17	0.28	0.44	0.67	0.83	1.00	1.00
60–69 years – ED = 2 (n = 108)										
0-second	0.99	0.02	0.89	0.94	1.00	1.00	1.00	1.00	1.00	1.00
5-second	0.88	0.13	0.45	0.61	0.67	0.83	0.89	1.00	1.00	1.00
10-second	0.81	0.19	0.28	0.41	0.50	0.69	0.89	0.94	1.00	1.00
20-second	0.79	0.19	0.28	0.44	0.53	0.67	0.83	0.94	1.00	1.00
70–86 years – ED = 1 (n = 63)										
0-second	0.97	0.07	0.74	0.83	0.89	0.97	1.00	1.00	1.00	1.00
5-second	0.67	0.21	0.17	0.33	0.34	0.53	0.67	0.89	0.94	0.94
10-second	0.54	0.26	0.07	0.17	0.18	0.31	0.56	0.78	0.89	0.89
20-second	0.51	0.26	0.00	0.12	0.22	0.33	0.50	0.72	0.89	0.89
70–86 years – ED = 2 (n = 78)										
0-second	0.98	0.05	0.75	0.88	1.00	1.00	1.00	1.00	1.00	1.00
5-second	0.82	0.20	0.29	0.44	0.50	0.72	0.89	1.00	1.00	1.00
10-second	0.69	0.22	0.25	0.32	0.39	0.50	0.72	0.89	0.97	1.00
20-second	0.70	0.24	0.10	0.27	0.39	0.56	0.78	0.89	1.00	1.00

Note. SD = Standard Deviation; ED = Education level; ED = 1: 12 years of education or less; ED = 2: more than 12 years of education.

Table 1

Distribution of Clinical and Demographic Data (Means and Standard Deviations) for the TBI Group, the Healthy Group, and the Whole Control Sample

	TBI (n=47)	Matched Healthy (n=47)	Whole Healthy (n=726)
Demographic			
Sex (no. of females)	18	18	397
Age (years)	37.70 (12.89)	37.77 (12.86)	49.78 (19.94)
Education level (years)	13.00 (2.26)	13.11 (2.98)	12.64 (3.51)
Clinical			
GCS	12.33 (4.48)		
Length of coma (hours)	53.80 (181.85)		
Months since injury	27.49 (45.09)		
Type of injury (no. of participants)			
RTA	31		
Fall	8		
Violence	5		
Other	3		

Note. TBI = Traumatic brain injury; GCS = Glasgow coma scale; RTA = Road traffic accident

Table 2

Partial Correlation Matrix for Each of the Scores of the Brown-Peterson Task and the Scores of the Other Cognitive Measures

	0-second	5-second	10-second	20-second
PASAT 2.4	.18	.16	.22	.21
PASAT 2.0	.15	.25*	.20	.37**
PASAT 1.6	.14	.25*	.32*	.35**
PASAT 1.2	.07	.16	.15	.22
Forward digit span	.15	.18	.26*	.21
Backward digit span	.05	.28*	.41**	.27*
Stroop – Color	–.01	–.28*	–.31*	–.31*
Stroop – Word	–.07	–.09	–.20	–.19
Stroop – Interference	–.13	–.22	–.26*	–.26*
Visual pattern	–.06	–.06	–.19	–.19
SRT	.11	.03	.10	.04
fNART	.13	.33**	.46**	.45**

*p<.015 **p<.001

Note. PASAT = Paced Auditory Serial-Addition Task; SRT = Buschke Selective Reminding Test; fNART = French version of the National Adult Reading Test

Table 3

Positive and Negative Predictive Values for Four Scores of the Brown-Peterson Task

	All TBI			Mo/S TBI		
	Cutoff	PPV	NPV	Cutoff	PPV	NPV
5-second delay	.86	.69	.77	.86	.75	.75
10-second delay	.83	.79	.70	.83	.76	.74
20-second delay	.83	.70	.68	.70	.80	.88

Note. TBI = Traumatic Brain Injury; Mo/S = Moderate to Severe; PPV = Positive predictive value; NPV = Negative predictive value

Figure captions

Figure 1. Description of the Brown-Peterson procedure.