Does the scoring of late responses affect the outcome of the paced auditory serial addition task (PASAT)?

Julie Balzano a, Nancy Chiaravalloti a,b, Jeannie Lengenfelder a,b, Nancy Moore a, John DeLuca a,b,c,*

a Kessler Medical Rehabilitation Research and Education Corporation, Neuropsychology and Neuroscience Laboratory, West Orange, NJ, United States
b UMDNJ—New Jersey Medical School, Department of Physical Medicine and Rehabilitation, United States
c UMDNJ—New Jersey Medical School, Departments of Physical Medicine and Rehabilitation and Neurosciences, United States

Accepted 13 September 2006

Abstract

The Paced Auditory Serial Addition Task (PASAT) is a cognitive task purported to measure working memory, speed of information processing, and sustained and divided attention (Spreen, O. & Strauss, E. (1998). A compendium of neuropsychological tests: administration, norms, and commentary (2nd ed.). New York, NY: Oxford University Press.) The current study examined whether treating late responses on the PASAT as correct or as incorrect can significantly affect the outcome of the test in individuals with Multiple Sclerosis (MS).

Subjects consisted of 59 individuals with MS and 37 healthy controls (HC). Scoring method was found to affect the number of individuals who were considered impaired on the PASAT. When participants were penalized for late responses significantly more participants, in both groups, were found to be impaired on the 2 s trial of the task (MS: p < 0.01; HC: p < 0.05).

Results indicate that that the specific scoring of the PASAT should be reported in studies that use the test.

© 2006 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Multiple sclerosis; Processing speed; Working memory; PASAT; Scoring

1. Introduction

The Paced Auditory Serial Addition Task (PASAT) is a cognitive task purported to measure working memory, speed of information processing, and sustained and divided attention (Spreen & Strauss, 1998), commonly used in both clinical and research settings. The PASAT was originally developed to measure changes in performance during recovery from closed-head injuries (i.e. concussion). It is now, however, administered to a variety of clinical populations including those with traumatic brain injury (Brooks, Fos, Greve, & Hammond, 1999; Gronwall, 1977), multiple sclerosis (MS) (DeLuca, Chelune, Tulsby, Lengenfelder, & Chiaravalloti, 2004; Miller, Rudick, Cutter, Bailer, & Fischer, 2000), lupus erythematous (Shucard et al., 2004), pain disorder (Sjogren, Thomsen, & Olsen, 2000), cancer (Sjogren, Olsen, Thomsen, & Dalberg, 2000), hypoglycemia (Schachinger, Cox, Linder, Brody, & Keller, 2003), and asthma (Weersink,
The PASAT has been shown to be particularly sensitive to the cognitive deficits commonly seen in MS (e.g. Diamond, DeLuca, Kim, & Kelley, 1997; Fischer, Rudick, Cutter, & Reingold, 1999).

In its most common forms, the PASAT is administered via audiotape, by which a series of single digit numbers between one and nine are randomly presented. Participants are instructed to add each of the numbers presented on the tape to each preceding number and to say out loud the totals. The dependent variable is the total number of correct responses given on each trial. Several versions of the PASAT exist, and can differ on factors such as the number of trials administered (i.e. two or four trials), the number of items within each trial (i.e. 61 or 50 items), the length of interstimulus intervals (ISI: 3 s/2 s or 2.4 s/2.0 s/1.6 s/1.2 s), and the medium through which the task is presented (i.e. audiotape, computer). While the auditory version of the PASAT is most commonly used, a visual version does exist (PVSAT), and was found to be comparable to the auditory test (Nagels et al., 2005). Participants with MS have been shown to be significantly impaired on both the PASAT and the PVSAT (Diamond et al., 1997).

As the speed at which each PASAT item is presented increases (e.g. the ISI), it is common for participants to give “late responses.” A late response is defined as a response given after the next number is presented on the tape. However, the issue of whether or not such responses should be included as correct in the total PASAT score is not consistently addressed by all administration manuals. Gronwall’s (1977) test instructions for the commonly used four trial version of the PASAT do not address the issue of late responses, nor does the administration and scoring manual for the two trial version of the PASAT, part of The Multiple Sclerosis Functional Composite scale (MSFC; Fischer et al., 1999). Spreen and Strauss (1998), however, in commenting on the four trial version, indicate that responses made after the next number has begun to be presented are to be scored as incorrect and not counted in the total score: “to be correct, a response must be made before presentation of the next stimulus” (p. 245). Tombaugh (2006), in a review of the literature on the PASAT, refers to late responses as errors, but it is unclear how such responses have been treated in past research. Additionally, some clinicians, in spite of contrary test instructions, may choose to score late responses on the PASAT as correct to enable a more accurate diagnosis of deficits. On a more theoretical level, counting late responses as incorrect may be helpful in diagnosing speed of information processing deficits, but may penalize an otherwise intact working memory system.

A large number of normative studies have been published on the different versions of the PASAT. Again, it is not clear from many of these studies how late responses were treated. Most normative studies examining the four trial version, the 50- and 100-item short forms, and the two trial version from the MSFC do not address the issues of whether or not late responses are included in the scoring (Roman, Edwall, Buchanan, & Patton, 1991; Brittain, La Marche, Reeder, Roth, & Boll, 1991; Diehr, Cherner, Wolfson, Miller, & Grant, 2003; Wiens, Fuller, & Crossen, 1997; Solari, Radice, Manneschi, Motti, & Montanari, 2005). Diehr, Heaton, Miller, and Grant (1998), in contrast to the studies mentioned above, clearly states that participants were instructed to make each response before the next number was presented on the tape, however, it is still not indicated how late responses, if they happened to occur, were treated. As the PASAT is a common neuropsychological tool used by both researchers and clinicians, it is important that the test’s reliability be maintained, and that the test is administered and scored in a consistent manner.

The primary purpose of the present study was to examine whether treating late responses on the PASAT as correct or as incorrect can significantly alter the outcome of the test in persons with MS, a subject group known to show impairments in processing speed and working memory (e.g. Archibald & Fisk, 2000; Bagert, Camplair, & Bourdette, 2002; DeLuca et al., 2004; Rao, 2004).

2. Method

2.1. Participants

Ninety-six participants, 59 individuals with MS and 37 healthy individuals served as subjects. Individuals were excluded from participation if they were not between the ages of 18 and 65, if they had a history of any neurological disease (other than MS for the MS group), if they had any major psychiatric disturbance, if English was not their first language, or if they had a history of drug and/or alcohol abuse. Demographic data are presented in Table 1.

All participants in the MS group were diagnosed with clinically definite MS according to the criteria of Poser et al. (1983), and were between 22 and 63 years of age ($M = 44.9$, S.D. = 9.5), completing between 11 and 20 years of formal
Table 1
Demographic information for MS and HC groups (n = 96)

<table>
<thead>
<tr>
<th></th>
<th>Total MS (n = 60), mean (S.D.)</th>
<th>Total HC (n = 38), mean (S.D.)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>44.9 (9.5)</td>
<td>43.6 (11.2)</td>
<td>ns</td>
</tr>
<tr>
<td>Education (years)</td>
<td>14.5 (2.2)</td>
<td>14.9 (2.0)</td>
<td>ns</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>81.4</td>
<td>70.3</td>
<td>ns</td>
</tr>
<tr>
<td>Years since diagnosis (MS)</td>
<td>9.6 (7.4)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

education ($M = 14.5, \text{S.D.} = 2.2$). Of the 59 participants in the MS group, 40 (67.8%) were diagnosed with relapsing remitting MS, five (8.5%) with primary progressive MS, nine (15.3%) with secondary progressive MS, three (5.1%) with progressive relapsing MS, and two (3.4%) with an undetermined type.

Participants in the healthy control group were between 22 and 64 years of age ($M = 43.6, \text{S.D.} = 11.2$), completing between 12 and 19 years of formal education ($M = 14.9, \text{S.D.} = 2.0$). There were no significant differences between the groups in terms of age, education, or gender composition.

2.2. Procedure

All data was collected as part of a larger research study on information processing speed in persons with MS. Prior to recruitment, the study protocol was reviewed and approved by the Kessler Medical Rehabilitation Research and Education Corporation (KMRRREC) Institutional Review Board. Participants with MS were recruited primarily through referrals from the outpatient Gimbel MS Center in Teaneck, NJ. Healthy Control participants were primarily recruited through flyers posted at the Kessler Institute for Rehabilitation in West Orange, NJ and through advertisements in local newspapers. Additionally, both healthy controls and participants with MS were also recruited through participation in previous studies in the Neuroscience and Neuropsychology Laboratory of KMRRREC. All participants were initially screened over the phone and then scheduled to complete testing if they met the study inclusion criteria mentioned above. All participants signed an informed consent form prior to study participation and were paid for their participation.

2.3. Materials

The Paced Auditory Serial Addition Task (PASAT), the cognitive dimension of the MSFC, was administered to all participants in accordance with standardized procedures (Fischer et al., 1999). Sixty-one single digit numbers between one and nine were randomly presented via audiotape. The participants were instructed to add each of the numbers presented on the tape to each preceding number and to say out loud the totals. Two trials of this task were administered: one in which the numbers were presented at a rate of one every 3 seconds (3 s), and another in which the numbers were presented at a rate of one every 2 seconds (2 s). All participants were given a visual presentation of the task following the initial instructions, and an 11-digit long practice string on audiotape before completing each trial.

For the present study, each participant’s performance on both the 3 and 2 s trials of the PASAT was scored using two different methods. In the first method, late responses were counted as correct (Late-C), and correct responses were defined as all appropriate responses, even if they were provided after the next number had begun to be presented by the tape. In the second method, late responses were not counted as correct (Late-I), and correct responses were defined as only those appropriate responses that were spoken before the next number was presented on the tape. For both methods, the dependent measure was the total number of correct responses, out of a possible 60, on a given trial. All responses were recorded by the test administrator.

For both methods, standard scores were calculated for each trial using normative data on the PASAT (Rao, Leo, Bernardin, & Unverzagt, 1991; Anon, 2006 http://www.firc.mcw.edu/pasat/PDF/PASAT_Manual.pdf). Participants were classified as “impaired” on the PASAT if they had scores greater than or equal to 1.5 standard deviations below the mean (seventh percentile), as calculated from normative data (Rao et al., 1991; Anon, 2006 http://www.firc.mcw.edu/pasat/PDF/PASAT_Manual.pdf). As is the case with most presentations of instructions and norms for the PASAT, the correct method for handling late responses was not addressed in the manual.
3. Results

The effect of scoring methods on PASAT outcome was analyzed by a 2 (group: MS, HC) × 2 (time: 3 s, 2 s) × 2 (condition: Late-C, Late-I) repeated measures analysis of variance (ANOVA). This data is presented in Fig. 1. By definition, a significant main effect of condition was found, indicating that mean PASAT scores were higher when late responses were counted as correct (Late-C) \( F(1,93) = 163.55, p < 0.001 \). As expected, a significant main effect of time was found where the mean number of correct responses was significantly higher in the 3 s versus 2 s PASAT trials \( F(1,93) = 176.84, p < 0.001 \). Finally, a significant main effect of group was observed, indicating that mean PASAT scores were significantly higher in the healthy control group compared to the MS group \( F(1,93) = 27.86, p < 0.001 \).

A significant time × condition interaction was observed, indicating that the difference between mean Late-C versus Late-I scores was larger on the 2 s trial versus the 3 s trial, collapsed across groups \( F(1,93) = 50.90, p < 0.001 \) (Fig. 2). In other words, penalizing participants for late responses had a greater effect on 2 s PASAT scores than it did on 3 s PASAT scores for both groups. However, the interaction of condition × group was not significant, with both groups demonstrating the same pattern of higher levels of performance when late responses were counted as correct that incorrect.

The mean number of late responses given on the PASAT was analyzed by a 2 (group: MS, HC) × 2 (time: 3 s, 2 s) repeated measures ANOVA. Correct and incorrect late responses were combined because the mean number of incorrect late responses alone was very small [MS = 0.3898 (3 s), 0.7119 (2 s); HC = 0.2973 (3 s), 0.4324 (2 s)]. No significant main effect of group was found, indicating that there was no significant difference between groups on the number of late responses across trials of the PASAT. The time × group interaction was also not significant for the number of late responses given by MS and HC groups across PASAT trials. As expected, a significant main effect of time \( F(1,94) = 50.46, p < 0.001 \) was found, indicating that both the MS and HC groups produced a greater number of late responses on the 2 s trial of the PASAT as compared with the 3 s trial.

Fig. 1. Mean PASAT scores for Late-C and Late-I scoring for the MS and HC groups.

Fig. 2. Mean PASAT scores for Late-C and Late-I scoring.
Table 2
Percent impairment on the 3 and 2 s trials of the PASAT for Late-C and Late-I scoring methods for the MS group

<table>
<thead>
<tr>
<th></th>
<th>3 s Late-C (n = 59)</th>
<th>3 s Late-I (n = 59)</th>
<th>2 s Late-C (n = 58)</th>
<th>2 s Late-I (n = 58)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impaired, % (≥−1.50S.D.)</td>
<td>41 (n = 24)</td>
<td>46 (n = 27)</td>
<td>24 (n = 14)</td>
<td>52 (n = 30)</td>
</tr>
<tr>
<td>Not impaired, % (≤−1.50S.D.)</td>
<td>59 (n = 35)</td>
<td>54 (n = 32)</td>
<td>76 (n = 44)</td>
<td>48 (n = 28)</td>
</tr>
<tr>
<td>F</td>
<td>$\chi^2 = 0.3$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p<0.01$.

Table 3
Percent impairment on the 3 and 2 s trials of the PASAT for Late-C and Late-I scoring methods for the HC group

<table>
<thead>
<tr>
<th></th>
<th>3 s Late-C (n = 37)</th>
<th>3 s Late-I (n = 37)</th>
<th>2 s Late-C (n = 37)</th>
<th>2 s Late-I (n = 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impaired, % (≥−1.50S.D.)</td>
<td>5 (n = 2)</td>
<td>8 (n = 3)</td>
<td>0 (n = 0)</td>
<td>11 (n = 4)</td>
</tr>
<tr>
<td>Not impaired, % (≤−1.50S.D.)</td>
<td>95 (n = 35)</td>
<td>92 (n = 34)</td>
<td>100 (n = 37)</td>
<td>89 (n = 33)</td>
</tr>
<tr>
<td>F</td>
<td>$\chi^2 = 0.28$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p<0.05$.

Table 4
Percent impairment for MS and HC groups on the 2 s trial of the PASAT for Late-C and Late-I scoring methods

<table>
<thead>
<tr>
<th></th>
<th>MS (n = 58)</th>
<th>HC (n = 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late-C, % impaired (≥−1.50S.D.)</td>
<td>24 (n = 14)</td>
<td>0 (n = 0)</td>
</tr>
<tr>
<td>Late-I, % impaired (≥−1.50S.D.)</td>
<td>52 (n = 30)</td>
<td>11 (n = 4)</td>
</tr>
<tr>
<td>F</td>
<td>$\chi^2 = 4.74$</td>
<td></td>
</tr>
</tbody>
</table>

* $p<0.05$.

To determine whether the proportion of individuals impaired on each trial of the PASAT was altered according to which scoring method was used, a $\chi^2$-test of association (Glass & Hopkins, 1984) was performed on both the 3 and 2 s trials of the test for MS and HC groups. These data are presented in Table 2 (MS) and Table 3 (HC). For the MS group, significantly more subjects were found to be impaired on the 2 s trial of the PASAT when Late-I versus Late-C scoring was used ($\chi^2(1) = 9.64, p < 0.01$). No such significant difference was found on the 3 s trial within the MS group ($p > 0.100$). Similarly, for the HC group, significantly more subjects were found to be impaired on the 2 s trial of the PASAT when Late-I versus Late-C scoring was used ($\chi^2(1) = 4.86, p < 0.05$). No such significant difference was found on the 3 s trial within the HC group ($p > 0.100$).

To examine whether scoring method had a disproportionate effect on the 2 s trial of the PASAT between the MS and the HC groups, a 2 × 2 (group × scoring method) $\chi^2$-test was conducted. Significantly more participants with MS were found to be impaired when Late-I versus Late-C scoring was used on the 2 s trial of the test ($\chi^2(1) = 4.74, p < 0.05$) (Table 4). In other words, scoring method had a significantly greater effect within the MS group than the healthy control group on the proportion of individuals classified as impaired on the 2 s trial of the PASAT.

4. Discussion

The primary objective of the present study was to determine whether treating late responses on the PASAT as correct or as incorrect would significantly affect test performance and interpretation. Scoring method was found to affect the number of individuals who were found to be impaired on the PASAT. When participants were penalized for late responses (i.e. Late-I scoring) significantly more participants were found to be impaired on the task.

As expected, subjects across groups scored significantly better on the 3 s versus 2 s trial of the PASAT, regardless of how late responses were scored. Also as expected, the HC group produced higher PASAT scores than MS participants on both trials, regardless of how late responses were scored (e.g. Archibald & Fisk, 2000; Bagert et al., 2002; DeLuca et al., 2004; Rao, 2004). Although Late-C scoring, by definition, resulted in higher PASAT scores on the 3 and 2 s trials of the test, it is most important to note that this was true for both the MS and HC groups, as demonstrated by the lack...
A significant interaction of condition (Late-C, Late-I) by group (MS, HC). Therefore, the scoring method utilized for the PASAT yields the same result in both the MS and healthy control groups.

Scoring method did, however, have a significant effect on the proportion of individuals, in both groups, who were classified as impaired on the 2 s trial of the PASAT. Specifically, when participants were penalized for late responses (i.e. when late responses were not counted as correct (Late-I)) on the 2 s trial, a significantly greater proportion of individuals in both the MS and HC groups were classified as impaired, although the effect was significantly higher in the MS group. Because virtually all late responses at the 2 s level were correct (very rarely was a late response incorrect), it is suggested that the increase in impairment was due to processing speed, rather than working memory accuracy. In other words, it seems that participants were correctly manipulating the numbers (i.e. working memory), but were not doing so quickly enough at the 2 s level (i.e. processing speed).

These findings indicate that studies which have not penalized participants for giving late responses (Late-C scoring) are being overly conservative in detecting impairment. Not counting late responses (Late-I scoring) penalizes subjects for being slow; hence the PASAT becomes more sensitive to processing speed impairments. In contrast, counting late responses (Late-C scoring) as correct adjusts for potential processing speed deficits, making the PASAT more a test of working memory accuracy. It is critical to match the method used for scoring to the method used for collecting the normative data. Thus, any study that uses the PASAT should report the specific scoring method. Unfortunately, normative studies generally do not provide this information, leaving the scoring method up to the individual user.

While the increase in the proportion of impaired individuals on the 2 s trial when Late-I scoring was used was present in both groups, it was significantly greater within the MS group. Processing speed deficits have been shown to be common in individuals with MS (Bagert et al., 2002; Chiaravalloti, Christodoulou, Demaree, & DeLuca, 2003; DeLuca et al., 2004; Rao, 2004). Furthermore, processing speed, rather than working memory, has been named a primary deficit in individuals with this disorder (Demaree, DeLuca, Gaudino, & Diamond, 1999). Thus, it could be expected that individuals with MS would show increased levels of impairment when they are penalized for late responses. As the PASAT is a common tool used for assessing cognitive functions in persons with MS, it is important to consider the scoring of late responses when interpreting results in both clinical and research settings, particularly when evaluating processing speed impairments.

Despite the significance of findings noted above, the current study did suffer from some methodological limitations that should be addressed in future research. First, the MS sample included in the current study was largely composed of individuals with a relapsing–remitting disease course. Only 15% (n = 9) of the MS participants in this study were diagnosed with a secondary progressive disease course, a course known to be more severe and tap different areas of cognition (e.g., both processing speed and working memory) than relapsing–remitting (e.g., only processing speed) (DeLuca et al., 2004). Thus, examination of only secondary progressive MS subjects may lead to different results than those observed in the present study. Future studies on the use of the PASAT in MS should seek a participant sample that is more representative of the different subtypes of MS.

Additionally, the restriction of the study sample to individuals with MS limits the conclusions we can draw about the utilization of the PASAT in general. Future studies on the PASAT should also examine the effect of differing methods for scoring late responses in other clinical populations, especially in those where the PASAT is commonly used.

Despite the methodological limitations noted, study findings were able to shed some light on the effects of differences in scoring methods on the classification of impaired versus not impaired in the PASAT. Overall, it is important to note that scoring method does significantly impact the interpretation of PASAT performance for both clinical and research purposes. Thus, the specific scoring criteria utilized for a given PASAT administration should always be noted and the subsequent interpretation of the PASAT should take scoring method into account.

Acknowledgments

This project was supported by grant #RO1 HD38249 and the Henry H. Kessler Foundation. The authors thank June Halper and the Gimbel MS center for their help in subject recruitment.

References
