Assessing Dual-Task Performance Using a Paper-and-Pencil Test: Normative Data

Sergio Della Sala 1,2,*, Jennifer A. Foley 1,2, Nicoletta Beschin 3, Mike Allerhand 2, Robert H. Logie 1,2

1 Human Cognitive Neuroscience, University of Edinburgh, Edinburgh, UK
2 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
3 Neuropsychological Service, Rehabilitation Department, Azienda Ospedaliera S.A. Abate Gallarate, Gallarate, VA, Italy

*Corresponding author at: Tel: +44 131 650 3440; fax: +44 131 651 3230.
E-mail address: sergio@ed.ac.uk (S. Della Sala).

Accepted 7 May 2010

Abstract

Although several studies have described dual-tasking ability in normal aging, Mild Cognitive Impairment and Alzheimer’s disease, no normative data for dual-task performance exist. Dual-tasking ability of 436 healthy individuals, aged 16–88 years, was assessed using a new paper-and-pencil dual-task paradigm. In this study, no age effect was detected, providing strong evidence that age does not affect dual-tasking abilities. Psychometric data for this new assessment are presented, which may enable clinicians and researchers to use this paradigm as a means of examining attentional control in dual-tasking.

Keywords: Assessment; Attention; Elderly; Geriatrics; Aging

Introduction

Dual-tasking is the ability to perform two tasks concurrently. Dual-task paradigms have been instrumental in forging cognitive models of attentional control (Fernandes & Moscovitch, 2000; Naveh-Benjamin, Craik, Guez & Kreuger, 2005), and in investigations of deficits in attentional control following focal and diffuse brain damage, including neurodegenerative disease (Baddeley, Baddeley, Bucks & Wilcock, 2001; Balota & Faust, 2001; Logie, Cocchini, Della Sala & Baddeley, 2004). However, its use in clinical practice has been limited by the absence of a portable and easy-to-use version of the task, and by the lack of normative data.

Experimental paradigms assess dual-tasking ability by examining the ability to perform two tasks separately and then together. The difference between the performance on each of the single tasks and in the dual-task condition provides an index of dual-tasking ability. Therefore, it is important to minimize any difference in performance on each of the single tasks across individuals (Salthouse, Ragan & Prill, 1984). Thus, the demand of individual tasks must be calibrated according to the ability of each individual to equate single task performance across participants.

The nature of the two tasks is a further important factor in the assessment of dual-tasking ability (MacPherson, Della Sala, Logie & Wilcock, 2007). To enable examination of the ability to perform two tasks simultaneously, dual-task assessments should assess concurrent performance on two tasks that do not compete for the same processing mechanism (Allport & Wylie, 2000; Monsell, 2003), such as aural presentation of digits for spoken recall coupled with visual and motor tracking (Logie et al., 2004).

In addition, unless performance on both of these tasks is considered, this trade-off between tasks may go undetected (Della Sala, Baddeley, Papagno & Spinnler, 1995). Thus, an accurate assessment of dual-tasking ability should include an overall measure of the impact of dual task on the performance of both tasks.

Several studies reported dual tasking to be poorer in older adults than in younger adults, with this age difference being exacerbated by increasing task demands (Anderson, Craik & Naveh-Benjamin, 1998; Craik, Govoni, Naveh-Benjamin &
Anderson, 1996; Craik & McDowd, 1987; Fernandes and Moscovitch, 2000; Hartley & Little, 1999; Lindenberger, Marsiske & Baltes, 2000; McDowd & Craik, 1988; Naveh-Benjamin et al., 2005). However, in these studies, the demands of each of the two individual tasks were not calibrated to the ability of each participant to equate single-task performance across age groups. Therefore, the age effect could have arisen because of baseline differences between age groups in their single-task abilities (Salthouse, Fristoe, Lineweaver & Coon, 1995). Dual tasking is affected by aging when one of the two tasks involves reaction times (Logie, Della Sala, MacPherson & Cooper, 2007; Macht & Buschke, 1983). Once care is taken to calibrate the performance on individual tasks so as to equate single-task performance across groups, and to use tasks that are not time based, this age effect disappears (Baddeley, Logie, Bressi, Della Sala & Spinnler, 1986; Belleville, Rouleau & Caza, 1998; Logie et al., 2004; Salthouse et al., 1995). It has been suggested, therefore, that dual-tasking ability is not affected by normal aging when single-task differences with age are taken into account (Logie et al., 2004).

In contrast, several studies that have calibrated the single-task performance between groups have found a striking impairment in dual-tasking ability in people with AD when compared with healthy age-matched controls (Baddeley, Bressi, Della Sala, Logie & Spinnler, 1991; Baddeley et al., 1986; Della Sala et al., 1995; Holtzer, Burright & Donovick, 2004; Logie et al., 2004; MacPherson, Della Sala & Logie, 2004; MacPherson et al., 2007; Morris, 1986; Morris & Baddeley, 1988; Sebastian, Menor & Elosua, 2006). This dual-task deficit becomes more pronounced with the development of the disease and increasing severity (Baddeley et al., 1991, 2001; Della Sala et al., 1995). The specificity of the dual-task decrement in AD is further underlined by the lack of dual-task deficits in older people with chronic depression, even when they have been equated with AD patients for episodic memory performance (Kaschel, Logie, Kazén & Della Sala, 2009). So, episodic memory appears to be sensitive to AD, but is not specific to the disease, whereas dual-task impairments appear to be specific to AD compared with chronic depression and with healthy aging. The cognitive function thought to be required for performing two concurrent tasks has been conceived as one of the executive functions within the working memory model (Baddeley, 1996; Baddeley et al., 1996, 1997) with the AD impairments in dual tasks reflecting a specific deficit of this function (Baddeley et al., 1986, 1991; Della Sala et al., 1995; Della Sala & Logie, 2001; MacPherson et al., 2004, 2007; Sebastian et al., 2006).

This dual-task impairment appears to be robust over practice (Baddeley, Cocchini, Della Sala, Logie & Spinnler, 1999) and remains with many different task combinations, including memory and motor tasks (Baddeley et al., 1986), two memory tasks (Cocchini, Logie, Della Sala, MacPherson & Baddeley, 2002; MacPherson et al., 2007), and two everyday tasks, like walking and talking (Cocchini et al., 2004).

Although the results were clear and have been replicated in a wide range of studies, the literature summarized above made use of computerized methods of assessing dual-tasking ability. These methods have a limited clinical utility, because they require bulky and expensive equipment, which necessitate training before usage.

Earlier attempts to develop simpler versions for clinical use (Della Sala et al., 1995; Baddeley, Della Sala, Gray, Papagno & Spinnler, 1997) discusses the assessment of dual-tasking ability using digit sequence memory (with the experimenter reading out the sequences for oral recall) with either a paper-and-pencil maze test (Della Sala et al., 1995; Baddeley et al., 1997) or a paradigm based on Fitts’ law (Baddeley et al., 1997). Neither of these was successful as many participants found the Fitts’ law task too difficult, even in the single-task condition, and the mazes test was insufficiently reliable, rendering these tasks inadequate for the confident and accurate assessment of dual-tasking ability in clinical practice.

This paper describes a new version of dual task that uses paper and pencil for the tracking element. The aims of this study were to collect normative data on healthy individuals and identify the psychometric properties of a simplified version of the dual-task procedure.

Materials and Methods

Participants

A total of 486 people took part in this study. The older participants were recruited from the Somma Lombardo Hospital orthopedic department and from different social groups for older people. The younger participants were recruited among the relatives of the older people. The inclusion criteria were: No known neurological or psychiatric disease, no uncorrected visual or auditory impairment, and Mini-Mental Status Examination score above the age-specific cut-off. Participants were either tested in the Somma Lombardo Hospital or in their social groups, in Lombardy, Italy. Participants were assessed individually by an experienced Clinical Psychologist in a quiet room, free from distraction, either in the morning or in the afternoon.
Procedure

Each participant completed the dual-task assessment. This consisted of performing digit recall and tracking tasks separately and then simultaneously. This dual-task assessment, relevant testing material and the full instructions for its administration and scoring, can be found at www.psy.ed.ac.uk/people/sdsala/tests/sdsdualtask/.

Before commencing the digit recall task, digit span of each individual was established. Participants heard a list of digits at a rate of one per second. They were then asked to repeat these digits back in the same order as they heard them. The initial sequence length was two digits long and participants were presented with six sequences at each sequence length. If five out of the six sequences were recalled correctly, the digit sequence was lengthened by one digit. Once participants could no longer recall five out of the six digit sequences, digit span was taken as the maximum length at which the participant was able to recall five out of six digit sequences correctly. Next, participants heard a series of sequences, at an individual span length, for immediate oral serial-ordered recall over a period of 90 s. The number of lists that each participant heard and recalled during the 90-s period varied depending on their digit span, and thus the performance measure was the proportion of digits recalled correctly in the correct serial position.

The tracking task consisted of using a pencil to draw a line through circles arranged in a path around a sheet of A3 paper. Participants were given a shortened version for a practice trial, with only 17 circles, to ensure that they understood the task demands. After this, the participant was presented with the full version comprising 319 circles and was asked to start at one end of the path and draw a line through each successive circle as quickly as they could over a 90-s period. The performance measure was the number of circles crossed within the allotted time.

In the dual-task condition, participants were asked to perform the tracking (circle crossing) task at the same time as listening to and repeating back the digit sequences they heard, for another 90 s. The performance measures were the proportion of digits recalled accurately and the number of circles crossed by the pencil. Proportional performance in digit recall ($p_m$) was calculated by measuring the change in digit recall between single- ($m_{\text{single}}$) and dual-task ($m_{\text{dual}}$) conditions, where $m$ is the proportion of digits recalled accurately, and using:

$$p_m = 100 - \frac{(m_{\text{single}} - m_{\text{dual}}) \times 100}{m_{\text{single}}}$$

Proportional performance in tracking ($p_t$) was calculated by measuring the change in tracking between single- ($t_{\text{single}}$) and dual-task ($t_{\text{dual}}$) conditions, where $t$ is the number of circles drawn through, and using:

$$p_t = 100 - \frac{(t_{\text{single}} - t_{\text{dual}}) \times 100}{t_{\text{single}}}$$

Proportional performance in both tasks combined ($\mu$) was calculated by using:

$$\mu = \frac{p_m + p_t}{2}$$

Reliability

To examine the test–retest reliability of the new dual-task assessment, a smaller sample of participants, representative of the larger group, completed the assessment a second time. All participants who agreed to be retested within the proposed interval were included in this reliability analysis. The time between the first and second assessments ranged from 1 hr to 1 day, based on the availability of the participant. This consisted of 176 healthy participants, aged 18–88 years (mean $= 51.86$, $SD = 21.25$) and an education range of 2–21 years (mean $= 10.48$, $SD = 4.88$).

Results

Normative Data for the New Dual-Task Assessment

The data were not normally distributed, as they were significantly positively skewed ($z_{\text{skewness}} = 8.07$) and leptokurtic ($z_{\text{kurtosis}} = 21.07$). The data skew appeared to be caused by a few particularly high scores (as seen in Fig. 1). Therefore, the
data were trimmed to exclude outliers more than two standard deviations (SDs) above or below the mean decrement in digit recall, tracking, or combined performance.

Fifty outliers were removed in total: 25 more than 2 SDs above and 25 more than two SDs below the mean decrement in digit recall, tracking, or combined dual-task performance. These 50 outliers (30 females, 57.7%) ranged in age between 23 and 87 years (mean = 62.12, SD = 20.24) and in education from 2 to 20 years (mean = 9.52, SD = 5.04). The remaining 436 individuals (254 females, 58.3%) ranged in age between 16 and 88 years (mean = 48.46, SD = 18.98) and in education from 2 to 22 years (mean = 11.62, SD = 4.45). Further participant demographic information, as a function of participant age, is found in Figs. 2 and 3. The outliers were significantly older than the remaining group ($U = 6,879.50$, $p < .001$), but there was no significant difference in the combined dual-tasking ability ($U = 11,274.00$, $p = .95$).

The outliers were relatively equally distributed about the mean, as seen in Fig. 4, indicating that their removal should not introduce bias in the remaining data.

---

**Fig. 1.** Histogram of the distribution of the index of dual task performance ($\mu$) before data trimming.

**Fig. 2.** Gender of participants, as a function of age.
Group means and SDs for the digit span, digit recall, and tracking tasks are presented in Table 1. Group means and SDs for dual-task performance of digit recall, tracking, and combined are presented in Table 2.

**Relationship between Dual-Task Performance and Gender, Age, and Education**

Point-biserial correlation analyses showed that digit recall, tracking, and combined dual-task performance were unaffected by gender.

The dual-task performance of the tracking task was significantly correlated with years of education ($\rho = .21, p < .001$) and age of participants ($\rho = -.25, p < .001$). Similarly, the combined dual-task performance was significantly correlated with years of education ($\rho = .21, p < .001$) and age of participants ($\rho = -.52, p < .001$). However, the dual-task performance of the digit recall task was not affected by education ($\rho = -.01, p = .79$) or age ($\rho = .04, p = .20$). When the effect of education
was partialled out, only tracking remained significantly correlated with age ($r = -0.16, p < .01$), but, crucially, not the combined dual-task performance ($r = -0.06, p = .11$).

**Reliability**

There was a significant test–retest correlation for dual-task performance in digit recall ($r = .59, p < .001$), tracking ($r = .73, p < .001$), and combined ($r = .69, p < .001$). There was a significant increase in digit recall [$t(175) = -2.27, p < .05$], tracking [$t(175) = -3.84, p < .001$], and combined dual-task performance [$t(175) = -4.34, p < .001$], but these increases were small in magnitude (mean increase = 1.65%, 4.89%, 3.27%, respectively).

**Dual-Task Performance: Normative Data**

Table 1. Performance on digit span, digit recall (in the single- and dual-task conditions), and tracking (in the single- and dual-task conditions)

<table>
<thead>
<tr>
<th></th>
<th>Healthy participants [mean (SD)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit span</td>
<td>4.95 (1.06)</td>
</tr>
<tr>
<td>Digit recall (single)</td>
<td>89.53 (10.81)</td>
</tr>
<tr>
<td>Digit recall (dual)</td>
<td>87.82 (10.02)</td>
</tr>
<tr>
<td>Tracking (single)</td>
<td>161.84 (96.54)</td>
</tr>
<tr>
<td>Tracking (dual)</td>
<td>155.17 (91.37)</td>
</tr>
</tbody>
</table>

Table 2. Dual-task performance of digit recall, tracking, and combined

<table>
<thead>
<tr>
<th></th>
<th>Healthy participants [mean (SD)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit recall</td>
<td>96.75 (6.94)</td>
</tr>
<tr>
<td>Tracking</td>
<td>97.14 (15.46)</td>
</tr>
<tr>
<td>Combined</td>
<td>97.48 (8.78)</td>
</tr>
</tbody>
</table>

Fig. 5 shows the distribution of combined dual-task performance ($\mu$), with a normal curve drawn in for comparison purposes. The distribution appears to be relatively normal, and passes the Shapiro–Wilk test of normality.

The uncertainty inherent in normality judgments based on hard cut-off scores is quantified using confidence levels. If cut-off scores are derived from sample estimates using parametric methods, the confidence levels are subject to assumptions about the distributions. However, the data obtained from task performance are typically skewed. The current data are skewed toward the
higher scores ($z_{skewness} = 0.51$). It is argued elsewhere (Capitani, 1997) that, for such data, thresholds for normality judgments are better established using nonparametric methods. The nonparametric tolerance interval (Wilks, 1941) is distribution-free in the sense that the resulting tolerance limits depend solely upon the sample size and the coverage and confidence level parameters; they are entirely independent of the unknown distribution of data in the population. Nevertheless, the cut-off scores based on these limits can be regarded as inferential and population-based, since Wilks provided the exact sampling distribution of the tolerance interval that covers a given proportion of population values with a given confidence level.

Following Capitani (1997), the normality judgments are based on the two cut-off scores, both derived from one-sided tolerance intervals: The outer and inner tolerance limits. The outer limit is calculated such that with 95% confidence the unbounded interval above the limit will cover at least 95% of normal population values. The corresponding inner limit is calculated such that with 95% confidence the interval above the limit will cover at most 95% of normal population values (see Supplementary Material for the method of calculation). The outer tolerance limit, which is the more conservative of the two, is used as the cut-off to classify a score as “abnormal”. The inner tolerance limit is used as the cut-off to judge a score “normal”. Scores between the outer and inner limits are considered “borderline”. Fig. 6 shows the scores in ranked order, and shows the cut-off scores corresponding to the outer and inner tolerance limits. Scores < 81.0% are classified as “abnormal” on the basis that, with 95% confidence, the interval above this threshold covers at least 95% of the normal population, so the risk of classifying a score as abnormal when it is in fact normal is at most 5%. Similarly, scores > 84.4% are judged to be “normal” as the interval below that threshold covers at most 95% of the abnormal population, so the risk of judging a score normal when it is in fact abnormal is at most 5%.

The nonparametric one-tailed lower 5% tolerance limits (with 95% confidence) for combined dual-task performance is 81.0%. This cut-off is then adjusted for education (see Table 3), to provide the corrected combined dual-task cut-off scores for the various levels of education.

Table 3. Corrections to be added to the raw scores according to the level of schooling and the adjusted combined dual-task performance cut-off scores

<table>
<thead>
<tr>
<th>Level of schooling</th>
<th>Education correction</th>
<th>Dual-task performance cut-off score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>-1.29</td>
<td>79.71</td>
</tr>
<tr>
<td>Secondary</td>
<td>1.29</td>
<td>82.29</td>
</tr>
<tr>
<td>College/university</td>
<td>3.87</td>
<td>84.87</td>
</tr>
<tr>
<td>Post-graduate</td>
<td>5.16</td>
<td>86.16</td>
</tr>
</tbody>
</table>

Fig. 6. Cut-off scores using outer and inner one-sided nonparametric tolerance limits, $N = 436$, coverage $\beta = 0.95$, confidence level $\gamma = 0.95$
Discussion

Our aims were to establish a set of norms for a simplified version of a dual-task procedure that could be used for research and clinical purposes. The data gathered with the simpler dual-task procedure were not normally distributed. There were a few outliers considerably above the rest of the distribution, which created a positive skew and therefore the data were trimmed by two SDs. This technique removes outlier bias, but it is possible that this technique in itself gives a false impression of a ‘normal’ dual-task performance. However, the outlier analysis revealed that the excluded data points were relatively well balanced either side of the mean, suggesting that the exclusion of these would not, in itself, introduce bias.

The remaining data are normally distributed. There is an indication that the scores might be skewed toward the higher end, but this trend was not significant. Conservative non-parametric methods (Capitani, 1997; Wilks, 1941) were used to identify cut-offs for ‘normal’ performance, which allow for some skewness in the data. Such conservative methods will limit type I errors when interpreting dual-task performance, and may prove useful in clinical practice. This paper-and-pencil assessment is simpler, easier to use, and cheaper than the existing computerized methodologies, and the test–retest reliability is sufficiently high, unlike previous assessments, making it ideally suitable for clinical use. Moreover, as the assessment is individually calibrated to each individual participant, thereby eliminating any single-task differences and only focusing on the intra-individual differences between single- and dual-task performance, the assessment also eliminates the possible measurement error by avoiding involvement of other cognitive processes, which are not the focus of interest here. Indeed, the bespoke nature of the dual-task assessment also ensures that it minimizes the effect of any cultural or educational variation. Therefore, the normative data should be widely generalizable. Although it is recognized that as it is not possible to calibrate the tracking task to individual level of ability, the dual-task assessment may be vulnerable to some single-task differences, although this was not found in this study.

The participants showed minimal reductions in performance under dual-task conditions, in either the digit recall, tracking, or the combined measure of dual-task performance. This suggests that, in general, healthy people demonstrate no difficulties in dual tasking under these testing conditions. No effect of age was detected, supporting the suggestion that the cognitive function involved in successfully coordinating the concurrent performance of two distinct tasks is unaffected by healthy aging (Baddeley et al., 1986; Belleville et al., 1998; Salthouse et al., 1995).

Previous work using laboratory versions of the dual-task procedure had indicated that dual-task impairments could be signatures of AD that are not shown in other patient groups, such as chronic depression (Kaschel et al., 2009) or MCI (Foley, Kaschel, Logie, & Della Sala, submitted), or in healthy older people (Baddeley et al., 1986, 1991; Della Sala et al., 1995; Holtzer et al., 2004; Logie et al., 2004; MacPherson et al., 2004, 2007; Morris, 1986; Morris & Baddeley, 1988; Sebastian et al., 2006). This accompanies the well-established episodic memory deficit that is characteristic of AD but is not necessarily specific to the disease. This dual-task paradigm could therefore ideally complement measures of episodic memory in assessing cognitive competence in people with AD. Episodic memory tests are currently considered to be the most useful assessments to detect AD, but they are not specific to AD even if they are highly sensitive to the disease. Episodic memory impairment can be associated with a wide range of disorders and, to a certain extent, normal aging, which can lead to diagnostic uncertainty. Therefore, elements more specific to AD need to be added to improve the diagnostic utility of these tests. Moreover, as individuals with AD quickly reach ‘floor’ on episodic memory assessments, these tests are unsuitable for following up patients over time. As the dual-task assessment is adjusted to individual level of ability, this assessment can be used with patients for longer. This study has shown that the dual-task assessment is unaffected by normal aging and offers a set of norms as well as a simplified test protocol that could be used in a clinical setting. Thus, combining the sensitivity of episodic memory tests with the specificity of this dual-task assessment may prove to be particularly useful for singling out patients with AD and following up their performance over time.

Supplementary Material

Supplementary material is available at Archives of Clinical Neuropsychology online.

Funding

This work was supported by the grant AS-90-2007 awarded by the Alzheimer’s Society to SDS and RHL. JAF is the appointed Research Fellow on this grant.
Conflict of Interest

None declared.

Acknowledgement

We are grateful to Dr Anna Cantagallo who assisted us in testing some of the older adult control participants.

References


