



Word frequency effects and intelligence testing

Adrian Furnham*, Alastair McClelland

University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

ARTICLE INFO

Article history:

Received 19 September 2009

Received in revised form 30 November 2009

Accepted 1 December 2009

Available online 4 January 2010

Keywords:

Word frequency

Intelligence

ABSTRACT

Participants completed two near identical reasoning tests in different orders. One completed the **Baddeley (1968)** verbal reasoning test which had the words 'precede' or 'does not precede' on half the items while the other substituted this for 'comes before' or 'does not come before', a version used by **Hartley and Holt (1971)**. Word frequency tables suggests the second version would be easier than the first. Those who did the first followed by the second version increased their score by over 10 points, while those who did the reverse order showed no difference between scores. Implications for testing non-native speakers is discussed.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

This study is concerned with how word usage in verbal intelligence tests has a significant effect on test scores. It will examine two versions of the Baddeley Reasoning Test published over 40 years ago and which is still used in research in many fields (Furnham & Monsen, 2009; Ling, Burton, Salt, & Muncer, 2009; Schlittmeier, Hellbruck, Thaden, & Vorlander, 2008; Venetjoki, Kaarlela-Tuomaala, Keskinen, & Hongisto, 2006). What is particularly attractive about the test is that it has been shown to be a reliable test of cognitive ability that only takes 3 min. By the end of 2009 the test had over 400 citations in the *Web or Science* as it is used extensively in ergonomics and health psychology as a simple, reliable measure, "sensitive to a number of stresses" (Baddeley, 1968, p. 341).

The Baddeley Reasoning Test (BRT) was published in 1968 and described as "three-minute reasoning test based on grammatical transformation" and as "a simple reasoning test involving the understanding of sentences of various levels of complexity" (Baddeley, 1968, p. 341). The test involves noting whether a statement i.e. "A is followed by B – AB" is true or false. There are only six types of sentences: A follows/precedes B; A is followed by/is not followed by B; A is preceded by/is not preceded by B, though the A and B are interchanged frequently as the "answer" which is either AB or BA.

The original paper had evidence of test–retest reliability ($r = .80$, $N = 18$, on successive days) and validity ($r = .59$, $N = 29$, British Army Reasoning Test). The paper noted reasonably small practice effects when testing people five times in 1 h.

* Corresponding author. Tel.: +44 207 679 5395; fax: +44 207 436 4276.
E-mail address: a.furnham@ucl.ac.uk (A. Furnham).

Numerous studies over the years have provided evidence of the test's predictive and concurrent validity (Chamorro-Premuzic, Furnham, & Moutafi, 2004; Furnham, Zhang, & Chamorro-Premuzic, 2006b). Thus, for instance, Batey, Chamorro-Premuzic, and Furnham (2009) found the BRT was positively correlated with General Knowledge ($r = .40$, $N = 72$), while Furnham, Rawles, and Iqbal (2006a) found it correlated significantly with the Wonderlic measure ($r = .56$, $N = 70$). Earlier Furnham and Chamorro-Premuzic (2006) found that BRT correlated with the Wonderlic ($r = .40$, $N = 112$) and General Knowledge ($r = .29$, $N = 110$).

Soon after Baddeley (1968) published his test, Hartley and Holt (1971) used a slightly adapted text to examine the validity and reliability of the test (Simplified Baddeley's Three-Minute Reasoning Test – SBTMRT). They found a correlation of $r = .70$ ($N = 68$) with the AH4 and a seven day interval test–retest reliability $r = .66$. Hartley, Hogarth, and Mills (1972) in three experiments found however the test less valid and reliable than in their previous study. They concluded the quest for a short, reliable, easy to administer and score test needed to continue.

However, there was an important difference between the Baddeley and Hartley and Holt versions of the test. The test has two concepts: follows and precedes. Thus, one item is: 'A follows B – BA' or another is: 'B is preceded by A – BA'. Some questions use the word 'not'. So an item may be: 'B is not followed by A – BA' or 'A is not preceded by B – BA'. Hartley used 'comes before' as opposed to 'precedes'. Thus in the Baddeley the item is 'B precedes A – AB' whereas in Hartley it is 'B comes before A – AB'. There is no difference in the test for about half the items using the phrase 'is/is not followed by' but for the remaining items the Baddeley test uses 'is/is not preceded by' while the Hartley test uses 'comes before' or 'comes after'.

Silver, Phelps, and Dunlap (1989) also varied the test arguing that the passive voice takes longer to process than the active voice.

They noted that the Hartley and Holt (1971) modification from *precedes* to *comes before* adds to the sentence length and therefore processing time. They therefore substituted *leads* for *precedes*. They found in a timed computer administered version of the test that *leads* produced more correct responses per unit time than *precedes* but argue for keeping the original wording. They also pondered the word frequency issue and argued that the word *follows* has a higher frequency occurrence than *leads* which should lead to faster recognition time but that their results did not support this hypothesis.

However neither Hartley and Holt (1971) nor Silver et al. (1989) required participants to do both versions (original plus changed) of the test in different orders. Thus they could not look for differences or carry-over effects which were concerned with having completed one version first, then observing higher scores on the second version.

This paper looks at differences in candidates' performance in both near identical tasks. There is an extensive literature which suggests that information processing is substantially affected by both word frequency and word length (e.g., Balota et al., 2007). If one consults the British National Corpus and examines differences in the two word groups 'precede' vs 'come before/after', it seems that the latter is around four times more frequently used than the former.

Given that the BRT is a speeded, fluid, word processing test based on grammatical transformation it was predicted that participants would do better at the simpler SBTMRT version of the test which differs only in the use of the precede/come before/after phrase on approximately half the items. This has implications for the norms of tests but also testing people who are less fluent in English and whose results maybe much more affected by the word frequency problem inherent in this test.

2. Method

2.1. Participants

Eighty-eight economics students (43 male) aged 18–24 ($M = 21.14$ years, $SD = 4.43$) took part. They were all first year economic students who if not native speakers passed tests to study at university level.

2.2. Tests

1. BRT (Baddeley, 1968). A 64 item scale based on the original paper.
2. SBTMRT (Hartley & Holt, 1971). A 64 item scale identical to the above in directions and layout with the exception as described in the Section 1.

2.3. Procedure

A class of first year students were randomly divided into two groups: one of 40 and the other of 48. One completed the BRT first while the other group did the SBTMRT, under strict test conditions. After this they were given a mini lecture on an unrelated topic. Thirty minutes later the groups did the other version of the test. Three days later they were given their scores on both tests and the results discussed.

3. Results

The summary statistics are presented in Table 1. The results were subjected to a 2 (Test Version) \times 2 (Presentation Order) mixed ANOVA. As predicted, there was a significant main effect

Table 1

Mean performance on the two versions of the reasoning test under the two orders of presentation (standard deviations are in parentheses).

Presentation Order	Test Version	
	BRT	SBTMRT
BRT–SBTMRT	27.25 (11.33)	36.60 (11.39)
SBTMRT–BRT	38.31 (14.12)	35.50 (12.64)

of Test Version, $F(1, 86) = 10.55$, $p < .01$ ($\eta_p^2 = 10.9\%$) with overall performance on the SBTMRT ($M = 36.05$) being higher than the BRT ($M = 32.78$). There was also a significant main effect of Presentation Order, $F(1, 86) = 4.02$, $p < .05$ ($\eta_p^2 = 4.5\%$), and a significant interaction between the two independent variables, $F(1, 86) = 36.53$, $p < .001$ ($\eta_p^2 = 29.8\%$). In a simple cross-over design such as the one employed here, a main effect of Presentation Order indicates the presence of a *differential carry-over effect* from the first to the second testing session – and this is evident from an inspection of Table 1. When the SBTMRT followed the BRT there was a 9.35 point increase in performance, $t(39) = 5.97$, $p < .001$ ($d = 0.94$), whereas when the BRT followed the SBTMRT there a much smaller (albeit significant) increase of 2.81 points, $t(47) = 2.18$, $p < .05$ ($d = 0.31$). Another way of viewing the results is to note that performance on the SBTMRT did not depend upon order of presentation of the tests ($M = 35.50$ when first, and $M = 36.60$ when second), $|t| < 1$, ($d = 0.09$) but performance on the BRT was critically dependent on order of presentation ($M = 27.25$ when first, and $M = 38.31$ when second) $t(86) = 4.00$, $p < .001$ ($d = 0.86$). The significant interaction between Test Version and Presentation Order indicates a *non-specific carry-over or practice effect*; overall performance on the second testing session ($M = 37.46$) was higher than on the first ($M = 31.38$).

4. Discussion

A basic requirement of all tests, particularly intelligence tests are that they are reliable. The test–retest reliability of both versions of these two versions of the same test were satisfactory rather than particularly good. It is assumed by researchers who use this test that the scores are both reliable and valid and there is sufficient concurrent validity of the original Baddeley version. However no test norms have been published though individual researchers who have collected data over the years report evidence of a normal distribution.

What this study has demonstrated is the subtle power of high and low frequency words in short verbal reasoning tests. There was almost one standard deviation difference in the scores of students who did the “easier” high frequency word version first. They scored 38 out of 64 (59% correct) compared to 27 out of 64 (42% correct). Thus these tests cannot be considered as alternative forms.

This study highlights a number of important issues. The first is the effect of practice on cognitive ability testing. This study showed a dramatic practice or carry-over effect from an “harder test” to an “easier test” but not the other way around. Indeed those who did the “easier” version first, did less well in the subsequent test raising issues about how much and why ability test scores rise with practice. Second, whilst word frequency effects have demonstrable effects on performance it may be hypothesized that these are magnified in non-native speakers: that is, the difference between different versions for native speakers would be smaller than for non-native speakers. While there are always problems associated with giving verbal or grammatical reasoning tests to groups who are non-homogeneously native speakers it is possible that this test

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات