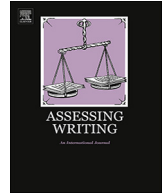




ELSEVIER

Contents lists available at ScienceDirect

Assessing Writing

journal homepage: www.elsevier.com/locate/asw

Researching the comparability of paper-based and computer-based delivery in a high-stakes writing test

Sathena Chan*, Stephen Bax, Cyril Weir

Centre for Research in English Language Learning and Assessment, University of Bedfordshire, Hitchin Road, Luton, LU2 8LE, United Kingdom

ARTICLE INFO

Keywords:

Test equivalence
Score equivalence
Cognitive validity
Computer-based testing of writing
Delivery mode
Second language writing assessment

ABSTRACT

International language testing bodies are now moving rapidly towards using computers for many areas of English language assessment, despite the fact that research on comparability with paper-based assessment is still relatively limited in key areas. This study contributes to the debate by researching the comparability of a high-stakes EAP writing test (IELTS) in two delivery modes, paper-based (PB) and computer-based (CB). The study investigated 153 test takers' performances and their cognitive processes on IELTS Academic Writing Task 2 in the two modes, and the possible effect of computer familiarity on their test scores. Many-Facet Rasch Measurement (MFRM) was used to examine the difference in test takers' scores between the two modes, in relation to their overall and analytic scores. By means of questionnaires and interviews, we investigated the cognitive processes students employed under the two conditions of the test. A major contribution of our study is its use – for the first time in the computer-based writing assessment literature – of data from research into undergraduates' cognitive processes within real-world academic settings as a comparison with test takers' cognitive processes during academic writing under test conditions. In summary, this study offers important new insights into academic writing assessment in the computer mode.

1. Introduction

In line with the increasingly key role of technology in all areas of higher education, computer-based (CB) assessment is becoming more and more common in most university disciplines (Newman, Couturier, & Scurry, 2010). In a similar fashion, many international language testing bodies now routinely use computers for various areas of Academic English writing assessment. In a study to compare ESL writers' performances on pen-and-paper and computer-delivered tests, Lee (2002) noted that test takers now believed a computer test to be more authentic and valid in relation to the target ESL contexts.

The International English Language Test System (IELTS) test, which is one of the most widely used tests of English language proficiency for educational, professional, and migration purposes, does not currently offer computer-based options. However, it seems more than likely, given the increased authenticity and other perceived benefits of CB testing, that in the near future IELTS will need to move towards offering computer-based options alongside traditional paper-and-pencil (PB) modes. In preparation for a possible move towards the CB assessment of IELTS, research was conducted some years ago to investigate differences between the CB and PB testing of IELTS writing (Wise & Plake, 1989). Although that research is still of relevance, in the intervening years students' increased familiarity with computers in both learning and assessment, as well as developments in test delivery technology, necessitate a fresh look at the questions of equivalence the study raised.

* Corresponding author.

E-mail address: sathena.chan@beds.ac.uk (S. Chan).

<https://doi.org/10.1016/j.asw.2018.03.008>

Received 8 September 2017; Received in revised form 24 March 2018; Accepted 25 March 2018

1075-2935/© 2018 Elsevier Inc. All rights reserved.

McDonald (2002) identified two fundamental types of equivalence which need to be examined when a pencil-and-paper writing test is offered alongside a computer delivered version and the two versions continue to co-exist side by side. The first, score equivalence, relates to the results of the test takers' performance and the concern is whether the scores obtained between the two modes are statistically equivalent and interchangeable. While score equivalence is often considered the most important issue in the delivery mode equivalence research, Mead and Drasgow (1993), who conducted a widely referenced meta-analysis of 159 correlations between paper-based and computer-based scores on writing tests, note that one should not assume that test takers use the same writing processes under different delivery conditions, especially when time-constraints are imposed. A second type of equivalence that needs to be examined, therefore, relates to the underlying construct that is being measured. Given that writing is a cognitively complex and socially situated activity, it is clearly impossible to achieve complete equivalence between the two conditions. However, in the context of direct writing assessment, it is essential to establish that the constructs operationalised by the tests are equally comparable between the two modes and in addition match as far as possible what students are expected to do in the target language use (TLU) domain (Bachman and Palmer, 1996).

Some research has been conducted to examine the cognitive processes of writers completing IELTS writing tasks (Yu, Rea-Dickins, & Kiely, 2011 on AWT1), but evidence of the cognitive validity of IELTS writing between the two modes is lacking, as is any comparison of either with the constructs underlying real life writing activities. Our aim is to examine the extent to which the results of computer-based IELTS, as a direct writing assessment, are statistically equivalent and construct valid as compared to the results from the paper-and-pencil IELTS. We also compare writing in both modes to real life writing in a university setting. The findings will contribute to establishing an evidence base of comparability that is a necessary pre-requisite to the introduction of a CB version of the IELTS writing test.

2. Literature review

2.1. Computer-based assessment of (academic) writing

In line with the shift towards computer-based academic writing in real life, many international high-stakes language testing organisations are moving towards the CB testing of writing, in some cases, abandoning the PB mode altogether. Cambridge English Language Assessment (<http://www.cambridgeenglish.org/>) offers CB versions of KET, PET, FCE, CAE and CPE in more than 350 test centres in 64 countries. The TOEFL iBT (<https://www.ets.org/toefl/ibt/about>) has already been taken in 1355 test centres in 149 countries, with the PB format now being phased out completely. Pearson (<https://pearsonpte.com/>) offers the PTE Academic test in CB mode only, and states that "more than 27 million test questions making use of this technology have been delivered, responded to, and automatically scored for individuals from over 100 countries around the world" (Pearson, 2012, p.7). The British Council has launched a CB test – Aptis (<https://www.britishcouncil.org/exam/aptis>). As almost all major academic writing assessments offer some forms of CB essay tasks, the momentum towards the need of CB writing test is compelling and those who do not follow in this direction risk being left behind and losing market share. Drawing on McDonald's (2002) work on the impact of individual variables on test equivalence, and on Mead and Drasgow's (1993) meta-analysis, we will now consider two types of equivalence, scoring and cognitive in turn.

2.2. Score equivalence

The literature of score equivalence in writing test between paper-based and computer-based modes presents a varied picture as regards outcomes. Early research by Mazzeo and Harvey (1988) suggested that CB tests at the time tended to be more difficult than PB versions, perhaps partly owing to test takers' lack of familiarity with the technology involved. Some found inconsistent results of the effect of delivery mode on performance. For example, Burke and Cizek (2006), in their study examining eighty 6th grade students, found that score equivalence was dependent on the prompt variable.

However, more recent research show that a CB test may elicit better performance from writers. Russell and Plati (2000) found that grades 8 and 10 students performed significantly better when they composed extended composition items under CB conditions. Wolfe and Manolo (2005) found that scores given to essays written in CB mode are in fact "slightly more reliable than scores assigned to handwritten essays and exhibit higher correlations with TOEFL multiple-choice sub-scores". Goldberg, Russell, and Cook (2003) performed a meta-analysis of 26 writing studies that were conducted from 1992 to 2002 concerning grades K-12 students. The results showed that students produced significantly better texts in terms of quality (effect size = 0.41, n = 15) and quantity of writing (effect size = 0.50, n = 14) under CB conditions. However, it is worth noting that only six of the studies examined the effects of writing mode on revisions, (which will be discussed more fully later), and they yielded inconclusive results.

On the other hand, a large body of more recent research concerning large-scale language tests showed that, depending on appropriate design, the scores across the CB and PB modes can be considered comparable (Russell, 1999; Taylor, Jamieson, Eignor, & Kirsch, 1998; Wright & Linacre, 1994). Taylor et al. (1998) studied the comparability of PB and CB versions for the 1996 administration of the TOEFL exam and found no significant differences in score for test takers taking the two different versions. Likewise, Wright and Linacre (1994) contended that PB and CB versions of writing tests yield very similar scores. Russell (1999), who examined over 1000 participants in a test of basic proficiency in reading, writing, and mathematics in CB and PB modes, found no significant difference in scores between the two modes. Based on performance of 262 participants, Wise and Plake (1989) reported that the difference between the PB and experimental CB versions of IELTS was not significant.

It appears that, provided that the test design is carefully constructed, score equivalence is achievable across the two modes in

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