Assessing Writing xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Assessing Writing

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

Going online: The effect of mode of delivery on performances and perceptions on an English L2 writing test suite

Tineke Brunfaut[a,*], Luke Harding[a,*], Aaron Olaf Batty[b]

[a] Lancaster University, Department of Linguistics and English Language, County South, Lancaster, LA1 4YL, United Kingdom
[b] Keio University, Shonan Fujisawa Campus, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan

ARTICLE INFO

Keywords:
Paper-based testing of writing
Computer-based testing of writing
Online testing of writing
Mode of delivery
Perceptions
Second language writing assessment

ABSTRACT

In response to changing stakeholder needs, large-scale language test providers have increasingly considered the feasibility of delivering paper-based examinations online. Evidence is required, however, to determine whether online delivery of writing tests results in changes to writing performance reflected in differential test scores across delivery modes, and whether test-takers hold favourable perceptions of online delivery. The current study aimed to determine the effect of delivery mode on the two writing tasks (reading-into-writing and extended writing) within the Trinity College London *Integrated Skills in English* (ISE) test suite across three proficiency levels (CEFR B1-C1). 283 test-takers (107 at ISE I/B1, 109 at ISE II/B2, and 67 at ISE III/C1) completed both writing tasks in paper-based and online mode. Test-takers also completed a questionnaire to gauge perceptions of the impact, usability and fairness of the delivery modes. Many-facet Rasch measurement (MFRM) analysis of scores revealed that delivery mode had no discernible effect, apart from the reading-into-writing task at ISE I, where the paper-based mode was slightly easier. Test-takers generally held more positive perceptions of the online delivery mode, although technical problems were reported. Findings are discussed with reference to the need for further research into interactions between delivery mode, task and level.

1. Introduction

Following significant technological developments and vast increases in computer accessibility, the past three decades have seen the introduction of several computer-based language testing systems. In some cases, tests have been conceptualised as computer-based from inception (e.g. Dialang, PTE Academic); in other cases, test developers aimed to replace a paper-based test with a computer-based system, or envisioned a parallel offer across the two modes of delivery (e.g. IELTS' concurrent paper-based and computer-based delivery of their reading, listening and writing components). Based on a comprehensive review of the computer-based testing literature, Davey (2011) concluded that the key motivations for adopting computer-based testing approaches are: (a) to target new constructs; (b) to achieve more accurate and efficient scoring; and (c) to make test administration more accessible, efficient and cost-effective. Factors such as market demand or policy requirements are also likely to play a role in the decision to move a paper-based test online. This article presents a comparative study within the context of motivation (c): the decision to add an online alternative to a primarily paper-based test (the Trinity College London ISE suite writing test) in order to make test delivery more efficient and accessible.[1]

* Corresponding author.
E-mail addresses: t.brunfaut@lancaster.ac.uk (T. Brunfaut), l.harding@lancaster.ac.uk (L. Harding), abatty@sfc.keio.ac.jp (A.O. Batty).

[1] Note that in the remainder of this article we mostly use the term 'computer-based' instead of 'online' to clarify the nature of the device and because this is the more frequently used term in prior research. We appreciate the additional internet connectivity aspect of the online mode, but this did not constitute a specific focus of the present study (but see the trial project referred to in section 2).

https://doi.org/10.1016/j.asw.2018.02.003
Received 6 September 2017; Received in revised form 7 December 2017; Accepted 9 February 2018
1075-2935/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

At face value, writing, of all the language skills, may be the most suitable to test in a computer-based environment. Since the spread of word processing software, and more latterly mobile technologies, a considerable proportion of day-to-day writing tasks are now completed on computers and other electronic devices. Scholars such as Jin and Yan (2017) have consequently argued for computer-based writing assessment. However, research has also shown that the writing medium may have an effect on the writing process (e.g., Van Waes & Schellens, 2003), and that writing processes may in turn influence text quality (e.g., Breetvelt, van den Bergh & Rijlaarsdam, 2009). A key question, therefore, is raised in those situations where a writing test offered in one delivery mode is replaced by another mode, or where two modes exist simultaneously and are intended to be parallel: does the mode of delivery affect test performance? This is especially important since task-related factors such as task layout, response mode or editing functions, and writer-related factors such as handwriting or computer skills may exert an influence. Depending on their effect and on their relevance, such factors may cause undesirable or construct-irrelevant variance. Therefore, researchers such as Choi, Kim and Boo (2003) have called for comparative studies on the impact of delivery mode as a prerequisite for the validation of computer-based tests (in contexts of paper-based replacements or parallel use of both modes).

1.1. Paper- versus computer-based testing of second language writing: performance results

Previous research on the impact of delivery mode on writing test scores, which has primarily concentrated on independent writing tasks, has not led to uniform conclusions. For example, in an early study with a counterbalanced repeated measures design, Owston, Murphy and Wideman (1992) found that a group of computer-experienced eighth-graders obtained significantly higher scores on all writing criteria for their paper-based writing performances compared with their computer-based scores. Yu, Livingston, Larkin, and Bonett (2004) also found that pre-service teachers gained systematically higher scores on essays in a paper-based writing test than those they produced in a computer-based test, after controlling for essay topic, essay version, and test-taker characteristics. Similarly, in a relatively recent study, Chen, White, McCloskey, Soroui and Chun (2011) established that (young) adults had better overall results on the paper-based version of an adult written literacy assessment than on its computer-based counterpart (regardless of their gender or level of education). Analysing their data in more detail, Chen et al. observed different delivery mode effects on the writing tasks in their study (a complaint letter, opinion letter and request letter) and that these differences were related to test-takers' characteristics (employment status in the complaint letter task, and age and race/ethnicity in the opinion letter task). They concluded that the computer mode might negatively affect the performance level of specific groups of test-takers (e.g., unemployed or 65+). Other studies which also identified a delivery mode effect on writing test scores, however, observed the effect to go in the opposite direction. Li (2006), for instance, found that a group of advanced English second language (ESL), adult learners gained higher scores on their argumentative writing for performances composed on a computer as opposed to those written on paper. Jin and Yan (2017) found that Chinese ESL college students produced longer texts with fewer language errors when completing expository essay tasks from the College English Test (CET) in computer-based mode as opposed to paper-based mode.

By contrast, a number of studies have found no impact of delivery mode on test-takers' mean writing scores, but have revealed effects at a sub-group or individual level. For example, Endres (2012) found no impact of delivery mode on average writing scores. However, when looking at individual test-takers' results, several obtained higher scores in the computer-based mode, and Endres warned that "writing in the two modes is a different experience for candidates and should not simply be considered comparable" (p.31). Breland, Lee, and Muraki (2005) also found similar observed mean scores between handwritten versus computer-typed performances of TOEFL writing tasks overall and did not detect any real task-effect differences. However, at a sub-group level, a delivery mode effect was discovered, but in contrast to Endres (2012), with systematically lower scores on the computer-based mode for those test-takers with lower ESL ability. Wolfe and Manalo (2005), who also focussed on the TOEFL, found a similar association between ESL proficiency and score effects of writing exam delivery mode. More specifically, they observed that less proficient test-takers (as measured by their scores on TOEFL multiple-choice items) benefitted from the handwritten mode, while no score differences between the delivery modes were found for more proficient test-takers.

Finally, when exploring the potential role of rating approach on the scoring of performances produced in different modes, Lee (2004) did not observe systematic writing score differences between the writing section of a paper- versus computer-based ESL placement test when using the test's conventional holistic rating approach. However, when looking at analytic ratings of the written performances along the criteria of organization, content, linguistic expression, and use of sources (using a scale developed for research purposes), Lee found that raters' analytic scores on all criteria were significantly higher on the computer-based version than the paper-based version. Powers, Fowles, Farnum, and Ramsey (1994), on the other hand, who looked into the effect on scoring of the medium in which written performances were presented to raters, found that performances presented in handwritten mode received higher scores on average than those presented in word-processed format (regardless of whether the texts had originally been written in handwritten or word-processed mode and transformed into the other mode later). It should be kept in mind, however, that Powers et al.'s research was conducted at a time when word-processed writing was less widespread, and raters might have been more used to handwritten texts. Given rapid developments in accessibility, use of, and familiarity with computers and word-processing in the new millennium, it is thus uncertain how transferrable these findings are to the present time.

The studies surveyed above suggest that, while results of comparative studies remain mixed, there is value in considering the effect of delivery mode at different levels of proficiency. Also, since Lee (2004) is one of the only papers which has considered delivery mode effects across different analytic scoring criteria, further evidence is required to understand interactions between delivery mode and different aspects of written performance as reflected in scale criteria. Furthermore, there is value in considering the nature of the writing task; a gap which has not been fully addressed in the literature to date. Comparative studies have typically focused on investigating mode of delivery for 'independent' writing tasks: those tasks which consist simply of a prompt or instructions

Download English Version:

https://daneshyari.com/en/article/6831498

Download Persian Version:

https://daneshyari.com/article/6831498

Daneshyari.com