Assessing Writing xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Assessing Writing

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

Examining the comparability between paper- and computer-based versions of an integrated writing placement test

Ha Ram Kim[a,*], Melissa Bowles[b], Xun Yan[a], Sun Joo Chung[c]

[a] Department of Linguistics, University of Illinois at Urbana-Champaign, 4080 Foreign Languages Building, 707 S. Mathews Avenue, MC-168, Urbana, IL, 61801, USA

[b] Department of Spanish and Portuguese, University of Illinois at Urbana-Champaign, 4080 Foreign Languages Building, 707 S. Mathews Avenue, MC-176, Urbana, IL, 61801, USA

[c] Graduate School of Education, Gachon University, 1342 Seongnamdaero, Sujeong-Gu, Seongnam-Si, Gyeonggi-Do, Republic of Korea

ARTICLE INFO

Keywords:
Process-oriented integrated writing
Peer review
Placement testing
Quality control
Computer-based test

ABSTRACT

The English Placement Test (EPT) is a process-oriented integrated writing placement test for newly-admitted international students at the University of Illinois at Urbana-Champaign. In order to meet student demand, since 2012 the EPT has been administered in both paper-pencil (on-campus) and computer-delivered (online) versions. Both versions feature a two-draft essay writing process and have identical testing procedures except that the online EPT does not have a peer review session, which is built into the on-campus test. This study examined the comparability of the on-campus and online versions, focusing on essay quality and examinee preference among 26 examinees who took both versions within a week, in counterbalanced order. Essay quality was measured in terms of linguistic (complexity, accuracy, fluency) and rhetorical features (integration of sources, progression of ideas, argument effectiveness). No meaningful differences in essay quality were observed between the two versions, although online essays were slightly longer. Post-test questionnaire responses revealed that a majority of test-takers preferred the online version for its convenience. We discussed the advantages and disadvantages of including peer review in writing placement tests, and we concluded by providing recommendations for evaluating comparability as a part of standard quality control practice in local tests.

1. Introduction

With the advancement of technological tools for language teaching and learning, computer-assisted learning has become an important facet of classroom teaching, and computer-based tests have become part and parcel of that trend. Commercial testing software is readily available, and automated scoring tools like the e-rater® are being developed with comparable scoring reliability to human raters (e.g., Enright & Quinlan, 2010; Weigle, 2010). The current generation of language learners has grown up with computers, meaning that computer access and familiarity, which were impediments to computer-based assessment in the past, have largely been overcome (Winke & Fei, 2008). This constellation of factors has triggered an increase in the use of computer-delivered language tests.

Early iterations of computer-based language tests often consisted of selected-response items or short constructed-response items that could be machine-scored. With a shift to develop more authentic language tasks, computerized tests have evolved to include open-ended item types such as essay tasks. These changes have led to an increase in the number of comparison studies between

* Corresponding author.
E-mail addresses: hkim169@illinois.edu (H.R. Kim), bowlesm@illinois.edu (M. Bowles), xunyan@illinois.edu (X. Yan), sjchung@gachon.ac.kr (S.J. Chung).

https://doi.org/10.1016/j.asw.2018.03.006
Received 21 August 2017; Received in revised form 15 March 2018; Accepted 21 March 2018
1075-2935/ © 2018 Elsevier Inc. All rights reserved.

computer- and paper-based tests, especially in the area of writing assessment (e.g., Goodfellow, Lamy, & Jones, 2002; Li, 2000; Winke & Fei, 2008). Previous studies have most commonly examined score comparability, test-taker attitude and motivation, and response mode (i.e., handwritten vs. typed). However, the results of such comparison studies are mixed and questions remain as to whether the two test modes yield comparable test performances. In addition, computers are so commonly used for writing assignments now that writing on a computer is becoming the norm. As computerized testing has become more popular and feasible, the comparability of the two test modes needs to be revisited.

Specifically, this paper examined the comparability of the paper- and computer-based versions of the English Placement Test (EPT) at the University of Illinois at Urbana-Champaign in terms of essay quality and examinee preference. The EPT is a workshop-style integrated writing test (Reading-Listening-Writing) offered to newly-admitted international students whose first language is not English (ESL, English as a Second Language). Whereas most large, high-stakes testing settings use a conventional timed single-draft essay test, the EPT employs a process-oriented writing test with small group discussion and peer feedback between the first and second drafts (see Section 4 for detailed description of the test). Considering its unique nature, this study also provides implications for the use of process-oriented integrated writing tasks in computer-based placement testing.

2. Paper-and-pencil based test vs. computer-based tests

When computer-based tests were first introduced, much of the concern surrounded whether computers would disadvantage examinees who were less familiar with them. Consequently, comparability studies of paper- and computer-based tests have mostly examined whether test-takers score differently based on the delivery mode (e.g., Choi, Kim, & Boo, 2003; Russell & Haney, 1997). Wolfe and Manalo (2004) compared paper-and-pencil and computer-based scores on the writing section of the Test of English as a Foreign Language (TOEFL), finding only a small effect for delivery mode. Interestingly, delivery mode interacted with test-takers' proficiency level: whereas higher proficiency test-takers did not score differently across delivery modes, lower proficiency test-takers performed better on the paper-and-pencil format.

Some writing assessment comparability studies have found no significant differences in mean scores of online and paper writing tasks (e.g., Horkay, Bennett, Allen, Kaplan, & Yan, 2006). In contrast, a few studies have observed that essays produced on a computer were longer and better in quality than paper-based essays (e.g., Lam & Pennington, 1995; Russell & Haney, 1997; Schwatz, Fitzpatrik, & Huot, 1994). Some studies found, to the contrary, that handwritten essays were of better quality than typed ones (e.g., Breland, Lee, & Murake, 2005; Bridgeman & Cooper, 1998; Powers, Fowles, Farnum, & Ramsey, 1994).

Several studies have shown that a writer's computer familiarity moderates the effect of test delivery mode. These include studies in K-12 contexts with children (e.g., Horkay et al. (2006)'s comparability study with the National Assessment of Educational Progress (NAEP)), as well as with university students (e.g., Wolfe, Bolton, Feltovich, & Niday, 1996) and in adult functional writing contexts (e.g., Chen, White, McCloskey, Soroui, & Chun, 2011). Specifically, computer-based assessments seemed to underestimate the writing ability of examinees with little experience using computers. Given the prevalence of computers in the daily lives of college students today, it seems unlikely that familiarity would play a role for the student population in our study.

Most previous studies comparing paper- and computer-based writing tests have focused on writing performance in terms of quality (as measured by scores) and quantity (as measured by text length). They have furthermore tended to focus on holistic aspects of writing with an emphasis on rater behaviors. The results of the studies are mixed and the effect of computer-based testing on writing performance remains inconclusive. Such conflicting results could be due to any number of factors, including scoring bias (whereby typed essays receive higher scores than handwritten ones, or vice versa), differences across examinee populations, task types, and topic difficulty. Differences in research outcomes seem to be influenced by the particular research context and details of the test procedure (Lee, 2004). Therefore, comparability between different test delivery modes should be established "within each research context" (Lee, 2004, p. 7).

3. Comparability studies for language placement tests

Few comparability studies have been conducted on paper-based and computer-based placement tests used in institutional settings. Fulcher (1999), one such early study on ESL writing placement testing, investigated the reliability and validity of the computerized version of a test adapted from a paper-based instrument. A more recent study by Lee (2004) investigated the comparability of ESL writers' performance on the paper-based and computer-based versions of the English Placement Test (EPT) at the same institution where the present study was conducted. However, when Lee (2004) was conducted, the test was very different. Rather than being process-oriented, it was a timed single-draft essay test. Also, in that study, test-takers had access to a computer during the lecture and while they wrote their draft, but the entire test was not actually online. Lee examined analytic essay features and studied rater behavior on handwritten and typed essays. There was no statistically significant holistic score difference between the versions, although the typed essays received significantly higher scores in analytic components (organization, content, use of sources, linguistic expression). Lee noted that this could have been due to practice effects, as test-takers had been recruited from a pool of students that had previously taken the paper-based test. Raters also seemed to have been negatively influenced by examinees' illegible handwriting on the paper-based test. Lee concluded that the benefits of computer-based ESL writing placement testing outweighed its disadvantages. However, practice effects and rater variance could have been confounding variables. Future research should eliminate these confounds by counterbalancing test mode and topic orders and by typing all essays prior to rating. It should also reflect the changes in the study context and test format.

Download English Version:

https://daneshyari.com/en/article/6831503

Download Persian Version:

https://daneshyari.com/article/6831503

Daneshyari.com