ASW SI Editorial

Paper-based vs computer-based writing assessment: divergent, equivalent or complementary?

Writing on a computer is now commonplace in most post-secondary educational contexts and workplaces, making research into computer-based writing assessment essential. This special issue of Assessing Writing includes a range of articles focusing on computer-based writing assessments. Some of these have been designed to parallel an existing paper-based assessment, others have been constructed as computer-based from the beginning. The selection of papers addresses various dimensions of the validity of computer-based writing assessment use in different contexts and across levels of L2 learner proficiency. First, three articles deal with the impact of these two delivery modes, paper-baser-based or computer-based, on test takers' processing and performance in large-scale high-stakes writing tests; next, two articles explore the use of online writing assessment in higher education; the final two articles evaluate the use of technologies to provide feedback to support learning.

Delivery mode is an important feature of a writing task which may potentially influence test takers' processing and performance. While research to date on the impact of delivery mode on writing test scores has been inconclusive, the three articles in this issue examine the comparability of the paper-based and computer-based versions of high-stakes writing tests and arrive at similar results. Brunfaut, Harding and Batty examine the impact of delivery mode on L2 test takers' performance on an independent and an integrated writing task within the Trinity College London Integrated Skills in English (ISE) test suite across three proficiency levels (CEFR B1-C1[1]). They report that delivery mode had no discernible effect on test scores, apart from the integrated task at B1, where the paper-based mode was slightly easier. Barkaoui and Knouzi look at the writing of second language users on the TOEFL iBT[2] independent writing task under the two delivery modes. They compare the performance of pre-admission (low proficiency group) and university students (high proficiency group), finding no significant effect of delivery model on students' computer-based and paper-based writing scores. Chan, Bax and Weir investigate the comparability of paper-based and computer-based IELTS Academic Task 2. They use both quantitative and qualitative data to examine the equivalence of the two delivery modes in relation to their cognitive and scoring validity and conclude that test takers' writing scores under the PB and CB conditions were not meaningfully different.

Additionally, the three studies examine the impact of several variables of other test taker characteristics on their computer-based writing test scores (including keyboard skills, perceptions of computer-based assessment and familiarity with computers). They report no significant impact of these variables on final test scores but note the test takers' preference for computer-based assessment. Two of the articles further investigate the possible impact of delivery mode on L2 test takers' processing and the linguistic characteristics of their performance. Chan et al. observe some differences in test takers' processes of planning, generating texts and revising under the two conditions. Barkaoui and Knouzi find that the delivery mode impacted on some characteristics of the writing: length of essay, lexical complexity, cohesion, and content. Although these differences in processing and linguistic features were not reflected

---

[1] Common European Framework of Reference (CEFR) for Languages: B1 (Independent User—Threshold), B2 (Independent User—Vantage) and C1 (Proficient User—Effective) levels

[2] Test of English as a Foreign Language (TOEFL) iBT is the version of TOEFL delivered via the Internet

in the test scores, the authors caution against the interpretation of full equivalence between the two delivery modes.

The second focus of this issue is to explore practices of computer-based writing assessment in higher education contexts. Writing on a computer in real-life situations, where writers are likely to have access to a wide range of materials (text, audio and video), writing tools and support, is clearly different from writing for a typical paper-based essay writing test. Computer-based assessment would seem to provide a more authentic task environment and is increasingly common in tests of English language skills for entry to higher education and formative assessment of academic writing skills. Kim, Bowles, Yan and Chung examine the comparability of an online version (delivered via the Moodle learning management system) and a campus-based version of an English Placement Test (EPT). Both versions involve process-oriented integrated writing (Reading-Listening-Writing), incorporating multiple draft submission and an online discussion forum (online) or face-to-face peer review (on-campus). Students wrote slightly longer texts in the online than in the on-campus version, but their writing was largely comparable in terms of its linguistic (lexical complexity, syntactic complexity and accuracy) and rhetorical (use and integration of source, progression of ideas, and argument effectiveness) features. The study demonstrates the use of computer-based assessment in an institutional setting and suggests ways to further enhance the validity of the online version. Fernando experiments with a formative academic literacy assessment through the Mahoodle learning platform. This involves scaffolding key processes of academic literacy: reading materials critically, outlining ideas, producing a draft and working with feedback to revise the text. To investigate the usefulness of the assessment, the author provides a fine-grained qualitative analysis of a range of data, including genre analysis of outlines and essays with feedback, semiotic analysis for student-generated digital artefacts, and thematic analysis for questionnaires and follow-up interviews. She argues that the interactive features of the computer-based academic literacy assessment enabled language instructors to shift the focus of the assessment from an evaluation of textual features and technicalities of textual composition to assessing students' engagement with complex academic literacy processes.

The final two articles explore how technologies (including keystroke logging, eye-tracking and automated evaluation of written text) can improve on our current methods of evaluating writers' performance[3]. In particular, they discuss how these technologies can support assessment for learning by providing more meaningful feedback to students. Ranalli, Feng and Chukharev-Hudilainen conduct an experimental study to explore how keystroke logging and eye-tracking technologies can be incorporated in formative assessment practice. Building upon previous research, they offer an innovative way of generating visualizations from keystroke and eye-tracking logs for users (i.e. instructors and students) to see how much time was allocated to different processes such as formulation versus revision and to see what type of revision was made at a particular point in task time. They argue that, when supported with instructor's evaluative feedback, visualisations of students' writing processes helped to position students' progress of learning in relation to developmental models of writing and to identify problems with individual processes (i.e. planning, formulation, and revision). While convinced by the benefits of using process-tracing technologies in formative assessment of writing, the authors discuss the support and resources required for this to be feasible for institutions.

In the final article of this Special Issue, Zhang and Hyland compare Chinese students of English's engagement with feedback provided by their teacher and Pigai, an Automated Writing Evaluation (AWE)

---

[3]For other important questions regarding assessing writing through automated scoring systems, readers are referred a previous Special Issue (Volume 18) of Assessing Writing edited by Elliot and Williamson (2013).

program, on their writing over a 16-week semester. The authors examine the complex processes of learner engagement with different types of feedback from the two sources. The authors recognise different strengths in the feedback provided by the teacher and the AWE program and argue that in classroom contexts it might be more effective to integrate the two types of feedback.

This special issue contributes to the research base of the comparability of paper-based and computer-based versions of high-stakes writing tests by examining the impact of delivery mode on writing test scores, test takers' processing and linguistic features of their performance. The issue also contributes to research on computer-based writing assessment use via online platforms in higher education contexts and explores the use of technologies to support assessment. For future research on computer-based assessment, the following questions merit attention.

- As demonstrated by articles in this special issue, computer-based assessment allows more flexibility in creating authentic task features which mirror more closely what test takers are expected to do in the target language domain than a typical paper-based writing test. Research is needed to identify what impact these contextual variables (e.g., access with/without the Internet, writing support tools such as spelling and grammar checks) might have on students' writing processing and performance.
- While computer-based writing has become the norm in most post-secondary contexts, the work of child psychologists argues for the value of hand writing as against typing in facilitating writing and reading development in young children; For example, James (2012) found that, in comparison to typing and tracing experiences, only handwriting letters fully activated processing in brain regions essential for writing and reading. Similar research on L2 writers across proficiency levels in the two modes would be a welcome addition to our knowledge base.
- The use of process-tracing technologies, such as keystroke-logging and eye-tracking, has great potentials for formative writing assessment, but the data generated can be enormous. To increase the feasibility of adopting these technologies in assessment contexts, research is needed to specify connections between writers' processing and final products. These findings can in turn suggest what data provided by these technologies would be most useful to inform assessment decisions.
- The final insight that arises from the articles is that human judgements and automated analyses in evaluating students' writing processing and performance may well complement each other. More research is needed to propose possible ways to integrate both teacher and AWE feedback in the same writing assessment.

It is our hope that this special issue provides readers an overview of the latest research on computer-based assessment so as to foster discussion and inform decisions regarding its development and use in various contexts.

Reference

James, K. H. (2012). The effects of handwriting experience on functional brain development in pre-literate children, *Trends in Neuroscience and Education, 1*(1), 32-42.