



# Issues in transitioning from the traditional blue-book to computer-based writing assessment

Nathaniel J. Hunsu\*

*Writing Assessment Program, Washington State University, Pullman, WA 99164-4530, USA*

---

## Abstract

With the increasing popularity of computers in learning and instruction, it is evident that students are no longer doing much pen-and-paper writing in the classroom. Since most academic work is done on computers, handwriting any piece of academic work is tacitly becoming foreign to students. As such, their placement writing assessments, too, should be administered on computers. However, the decision to transition from the traditional blue book to computer-based writing assessment requires a careful understanding of issues that affect students and raters and that college writing programs must be quipped to manage. This article discusses such critical issues necessary for making informed transitioning and suggests ways to ensure tests administered in both modes are comparable. © 2015 Elsevier Inc. All rights reserved.

---

Most U.S. colleges and universities have established requirements for freshmen entering into first-year composition courses and have developed some sort of placement mechanism for placing students into appropriate writing courses to prepare them for writing in their major disciplines. Variants include the use of students' scores in the writing components of such standardized tests as the SAT/ACT, locally developed placement tests (handwritten or taken online), and self-placements. A recent survey of the writing programs of member institutions listed on the website of the Association of American Universities (AAU) revealed that about 60% of universities rely on a system to place students into first-year composition courses (2012). While many of these do first-year composition placement based on SAT/ACT scores, only a few rely exclusively on such standardized tests for composition placement purposes. Most of them administer locally developed writing tests for international students, as well as for students whose performances are low on the SAT. Some of these schools, however, exclusively use unstandardized writing tests developed by the school to determine students' placement into writing courses. It appears that even schools that use scores from the SAT/ACT rely on samples of at least some, if not all, students' writings to make their placement decision. Some administrators require students to submit samples of past writings or take an impromptu placement writing test (i.e., blue-book exams).

While samples of students' writing remains the only tangible way to assess writing ability, presentation formats of writing samples have evolved considerably with the infusion of technology with learning. Technology has changed the landscape considerably. From desktops to laptops and now iPads, it is needless to say that students no longer do extensive handwriting. Most instructors require academic work to be word-processed rather than handwritten (Alexander, Bartlett, Truell, & Ouwenga, 2001; Wolfe, Bolton, Feltovinch, & Niday, 1996). With more states deemphasizing learning

---

\* Tel.: +1 509 335 3459.

E-mail address: [nat.hunsu@wsu.edu](mailto:nat.hunsu@wsu.edu)

handwriting in the elementary language arts curriculum, it is certain that handwriting skill is destined to lose its place for the more desired typing skill.

Since most students in the United States work with computers (DeBell & Chapman, 2006), handwriting any piece of academic work is becoming foreign. It seems instinctive that their placement writing assessments, too, should be administered on computers; at least it is intuitive to think that the construct validity of handwritten assessments for a lot of students seem compromised. It may be suggested that placement examinations, which are meant to demonstrate writing abilities, might be measuring psychomotor skills and psychological constructs, too, as a result of the discomfort that ensues under timed examination conditions for students who find no comfort in handwriting. In fact, H.K. Lee (2004) reported that habitual “typers” expressed embarrassment and frustration when required to handwrite these placement examinations. Such frustrations might confound the construct writing tests are poised to measure and dampen students’ confidence about the test, especially students with poor penmanship.

To address the situation, some administrators of writing assessment now offer computer-based impromptu writing tasks. While others have not made this transition for varied reasons, it is important to critically examine the need for migrating from a pen-and-paper to computer-based writing assessment, especially for high stake purposes such as freshmen placement writing assessments. This article makes a case for writing programs that still administer handwritten placement examinations to strongly consider transitioning from handwritten tests to computer-based impromptu writing placement examinations. It will examine factors that affect students’ performance on computer-administered writing tests and how raters’ judgments of students’ written products are affected by the mode in which they are presented. Additionally, it will suggest how to assuage these effects, as well as the side effects of transitioning to computer-based assessment for those students with minimal typing skill.

## 1. The need for transitioning

Just over a few decades ago, the use of computers in assessing students’ essays in direct placement writing assessments was less desirable, particularly because computers were not widely used, and their use would introduce construct irrelevant variants into the assessment of students’ writing (Maulan, 2004). Predicting the future role of computers in writing assessments, Donald Powers, Mary Fowles, Marisa Farnum, and Paul Ramsey (1994) reflected that as computer usage becomes more common in learning and assessment, “familiarity and comfort with computers will undoubtedly become a less important issue in assessment” than they have been in the past (p. 220). Computers are more common in educational settings today, and their proliferation in U.S. schools has defied the restraints of gender, ethnicity, and the level of education. The National Assessment of Education Progress (NAEP) reported that at least 89% of fourth graders who took the NAEP reading assessment in 2009 reported they have access to computers at home, and 91% have access to computers in school for reading and language arts activities (NCES, 2011). These findings, and other data on computer usage, suggest that students begin to use computers at an early age and go on in life using computers on many fronts. In light of this research, it is irrefutable that computers have become the way to learn, and studies on the role of technology in learning concur that computers facilitate teaching, facilitate the writing process, and motivate students to write by providing them with useful revision tools, thus enhancing students’ written products (Li, 2006; Wolfe et al., 1996).

Some studies have raised concerns that administering students’ writing assessments on computers really measures more constructs than are required in writing assessment (Maulan, 2004; McDonald, 2002). However, the pervasion of computers in schools reverses the concerns educators and assessors once had about the use of computers in writing assessment. With computers being extensively used for most students’ writing, projects, and assignments (Alexander et al., 2001), many students have become adept computer users. This situation calls researchers to question the fairness and validity of results of writing assessments of students who are adroit computer users constrained to handwrite their essays (Way, Davis, & Strain-Seymour, 2008; Lee, 2004; Russell & Haney, 1997). Having students who are so dependent on typing to handwrite, especially under timed writing conditions, may indeed be confounding their writing ability with other abilities. Put another way, is the system punishing students for neglecting their handwriting practices? Samuel Messick (1994) noted that for assessments to remain authentic, test administrators must provide students with the same tasks, contexts, time, and resources that parallel those they use in the real world. In fact, several studies have found that constraining students who are accustomed to writing with a computer to handwrite their essays undermines their productivity (Way et al., 2008; Russell & Plati, 2001; Russell & Haney, 1997).

A number of writing assessments have transitioned, or are at some phase of transitioning, from pen-and-paper to computer-based assessment of students' writing. Among such are writing assessments offered by the ETS. Similarly, more than 50,000 eighth and twelfth graders took the NAEP Writing Computer-Based Assessment (WCBA) pilot tests in 2011 (NCES, 2011). With this changing trend in the way students' writings are assessed, it is pertinent that colleges that still offer handwritten placement exams strongly consider upgrading their mode of assessing students' writing.

Besides the question of authenticity and the construct validity of pen-and-paper writing assessments in a computer era, transitioning online proffers other benefits for students and college writing programs. Several studies report instances where students found word processors enabled them to make revisions that improved the quality of their writings more easily than they could when they handwrote their essays (Li, 2006). Also, computer-based testing saves costs associated with printing and shipping test paper materials (Way, Davis, & Fitzpatrick, 2006) and eliminates the constraints of time and geographical location in test administration (Alexander et al., 2001). Hence, it seems logical that transitioning to computer-based writing assessment is the right way to go.

It should be noted, however, that transitioning to computer-based writing assessment genders alternative concerns that writing program administrators need to be aware of and prepare for. Particularly important is how students' individual differences impact their performances on word-processed timed essay examinations. In a circular logic, asking all students to type their timed essays may undermine the fairness of such a test, as we still have some students who may be disadvantaged because of uneven familiarity with computers (Chen, Whiteb, McCloskeyc, Sorouid, & Chune, 2011; Maulan, 2004). Hence, where written tests are to be administered in both handwritten and computer modes, as suggested by some studies (Russell & Tao, 2004b; Wolfe et al., 1996), validity, as well as comparability, of tests taken in both modes must be of interest (McDonald, 2002; Alexander et al., 2001). Secondly, the writing program needs to understand how students' demographic characteristics might affect their writing processes and the quality of their written works, as well as the influence of computer prints and handwriting on raters' decision. Bearing student differences in mind will, consequently, guide in providing appropriate accommodation for test takers. Lastly, administrators should be concerned about logistics and technology usability issues, especially if considering an online transition that would facilitate glitch-free writing assessments.

### *1.1. Validity of computer-based writing assessments*

How test performances are interpreted and used has always been of concern to education stakeholders. Hence, administrators strive to ensure valid use of tests and are severely criticized where that is not case. American Educational Research Association (AERA) defines validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA, APA, NCME, 1999, p. 9). Because placement writing examinations are presumed to evaluate where students are in their writing in order to place them into an appropriate writing course, it is important that students' performance on such writing tests fairly demonstrate their writing ability. Hence, factors that work against this intent undermine the validity the test and thus the subsequent inference derived from the test. To demonstrate validity in the way writing assessments are used will espouse issues of fairness as well as appropriateness, as indicated in the National Council of Teachers of English and the Council of Writing Program Administrators general positions on writing assessment (CWPA, 2014). As such, it is desired that writing tests, irrespective of the mode in which they are administered, are indeed measuring writing ability.

The question of validity is perhaps the underlying tone cutting across inquiries on the comparability of pen-and-paper versus computer-based assessment in earlier studies. Test validity is important because it affects the meaning placed on test results (Messick, 1995). Similarly in writing assessment, it is important not only to know that writing tests are appropriate to the tested population, but that the way the test is administered does not arouse constructs not intended to be measured by the writing test. Researchers are concerned if computer-based assessment of students' writing is an authentic measure of their writing performance. Given that computers are now accepted as tools in classroom engagement, students have become more dependent on them (Maulan, 2004). Having such students do all their schoolwork on the computer except the writing assessment undermines the validity of the assessment itself (Lee, 2004; Maulan, 2004; Messick, 1994). There is evidence to show that adept computer users are more likely to perform better on computer-based writing assessments than they would were they to handwrite (Way et al., 2008). Besides, ESL students who are not conversant with writing English alphabets are disadvantaged when they are required to handwrite their essays (Lee, 2004). Hence, requiring such people to handwrite might be capping their potential and introducing undue frustration into the testing hall. The standard posits that "fairness requires that all examinees be given

a comparable opportunity to demonstrate their standing on the construct(s) the test is intended to measure” (AERA, APA, & NCME, 1999, p.74). In fairness to these groups, computer-based writing assessment may be a valid measure of their writing performance.

Conversely, research reveals that students who are not skilled with word processors are worse off when they are required to type their timed writing assessment (Wolfe et al., 1996). Hence, the decision to administer the test in only one mode can result in test bias and place either group at a disadvantage (Maulan, 2004; Chen et al., 2011). Since many students are proficient using computers and word processors, computer-based writing assessment for them is a valid measure of their writing ability. On the contrary, pen-paper-based assessment will remain for the foreseeable future because U.S. colleges still enroll ESL students, adults, and members of minority population groups who do not have adequate computer experience and may not feel comfortable taking computer-based writing assessments (Chen et al., 2011; Lee, 2004). So validity of computer-based writing assessment depends on the student-group in question. Hence, to remain in the center of the game, writing programs may be compelled to provide writing exam candidates a choice to write their essays in the format they find most comfortable for them. Having the two modes of administration coexist, however, introduces a potential problem for those who score student essays. As much as possible, where writing assessment might be offered in both modes, it is essential that students’ performances are comparable to their writing ability, irrespective of the mode in which they write their assessment essays (Way et al., 2006).

## 2. Comparability of computer-based and pen-and-paper-based assessment

The literature reports mixed findings on the comparability of paper-and-pen and computer-based writing assessment. Some studies reported that students did better on handwritten essays than on word processors (Breland, Lee, & Muraki, 2005; Wolfe & Manalo, 2004; Bridgeman & Cooper, 1998; Powers et al., 1994; Arnold et al., 1990). Some studies found that students who used word processors performed better than those who handwrote their essays (Russell & Plati, 2001; Russell & Haney, 1997), while others reported mixed findings or found no differences between the two modes (Boulet, McKinley, Rebbecchi, & Whelan, 2007; Horkay, Bennett, Allen, Kaplan, & Yan, 2006; MacCann, Eastment, & Pickering, 2002). These inconclusive findings might have much to do with the characteristic of participants included in the studies. Given that students use computers more than ever before, it may be safe to suppose that keyboarding skill is less of a concern to the larger population, which perhaps might explain why this issue is not of particular concern in recent literature. It is a disservice to our practice, however, to neglect the fact that students who show up at college do not have equal access to technology, especially international students. Despite earlier findings, scrutinizing available comparability studies suggest factors that may be responsible for differences in students’ performance across modes. These factors as categorized by Michael Russell and Wei Tao (2004a) might be due to composition and presentation mode. *Composition mode* refers to the mode in which students compose their essays (handwriting or typing), while *presentation mode* refers to how written products are presented to raters for scoring (handwritten or typed).

Disparities in comparability of students’ performance due to mode of composition are a result of students’ individual and demographic differences, while presentation mode effects are a result of the influence of the mode of presentation (handwritten or typed) on raters’ holistic impression of the writing. Other factors, such as technology-related factors, can only affect comparability of students’ performances on either modes of composition by exploiting their individual differences.

### 2.1. Comparability studies and student-related factors

Computer-administered writing tests would only be tenable as valid alternatives to handwritten tests if they were deemed to be equally valid in measuring the intended construct: writing performance. Hence, it is important to understand how the two would compare—if a student had the opportunity to write in either mode—and how the factors might impair students’ abilities to type as comparably as they would handwrite their placement essays in order to alleviate their impacts.

A highlight of extant studies suggests that students’ individual differences and demographic characteristics play a major role in how students’ typed and handwritten essays compare. Individual differences include constructs such as: computer effects, keyboarding expertise, English language proficiency, reading proficiency, and writing ability, just to mention a few. Demographic differences include students’ ethnicity, age group, gender, and level of education.

Different studies examined how these factors moderate the influences typing and handwriting have on students' writing processes and the quality of essay products. The next section highlights some of these factors discussed in the literature.

### 2.1.1. Computer effects and performance

Students' computer attitudes are typically shaped by their experience, familiarity, anxiety, or confidence with computers (McDonald, 2002) and are factors that inform how they choose to take written assessments when given a choice. Edward W. Wolfe and Jonathan R. Manalo (2004) noted that inequality in access to computers and familiarity with word processors might account for low confidence and high level of computer anxiety common with minorities groups, older individuals, and others who may have limited experience with computers. Computer anxiety and level of experience with computers correlate with performance on computer-based test (Marcoulides, 1988). Thus research has observed that candidates from developing countries and other groups with minimal or no computer exposures would prefer to handwrite their essays when given the choice (Wolfe & Manalo, 2004).

It is unclear how computer anxiety influences students' writing ability. However, Michael W. Eysenck (1988) argued that anxiety reduces working memory capacities when the tasks involved are cognitively demanding. Computer anxiety, however, diminishes with more experience and familiarity with computers. Earlier, Michael W. Gos (1996) demonstrated that exposure to computers in itself was insufficient, and that the quality of the exposure was more important. Similarly, Sumarni Maulan (2004) found no significant relationship between computer familiarity and performance on computer-based tests in a study examining the effect of test-delivery medium on students' performances. Understandably, students may have varied experiences with computers depending on what they use them for, and computer familiarity may not necessarily be synonymous with typing ability. However, keyboarding proficiency is found to strongly influence students' performance on computer-administered writing tests. Students with good keyboarding skills are advantaged while those with poor skills are disadvantaged (Maulan, 2004; McDonald, 2002).

### 2.1.2. Keyboarding skills and performance

The literature provides interesting findings on how keyboarding skills interfere with students' performances. Researchers have used keyboarding skill (e.g. Whithaus, Harrison, & Midyette, 2008; Wolfe & Manalo, 2004) as well as word processing (Wolfe et al., 1996; Harrington, Shermisa, & Rollinsa, 2000) to express students' abilities to compose essays using the computer, and the two constructs seem quite related. While earlier studies argued that word processors in themselves make no difference in students' writing (Wolfe et al., 1996), these studies seemingly concurred that level of experience with word processors and keyboarding skill might interfere with students' writings. Some studies noted that keyboarding skills interfere with the production of texts, with students who have low keyboarding skills being adversely affected. Students with medium to high levels of experience with keyboards or word processors were not much affected (Wolfe & Manalo, 2004; Lee, 2004; Wolfe et al., 1996). Wolfe and Manalo (2004) argued that this phenomenon may be pronounced for ESL students with low English language proficiency and keyboarding skills, as they face a "double translation" dilemma translating their cognitive processes into the English language and subsequently into corresponding keyboard strokes. This challenge exerts considerable extraneous cognitive load on the student that may interfere with text production, especially under a timed writing placement examination condition.

Since level of experience with word processors and keyboards may interfere with text production, writing test administrators should make testing decisions with the awareness that students who lack the necessary typing competence may perform poorly on the computer, and those who are adept typers may perform below par on their handwritten test. This becomes consequential if we bear in mind that students perceive placement examinations to be high stakes in nature because the placement might mean an additional credit hour, which may reflect on students' academic transcripts, add additional time and money, and carry an accompanying stigma (Moss, 1995). Since we know that level of keyboarding skill and word-processing experience might influence both students' writing process and the quality of written products, it may be expedient to provide informed guidance to students who enroll for placement writing examinations, thus enabling students with limited keyboarding skill to elect whether to type or write their tests.

### 2.1.3. Influence of keyboarding/word-processors on process and quality of writing

To investigate the influence of word processors, researchers considered both the quantity and quality of written products (Lee, 2002). Earlier studies concluded that handwritten essays appeared apparently longer than word-processed essays (Powers et al., 1994), but word-processed essays were found to be longer than handwritten essays based on numbers of characters and word counts (Russell & Haney, 1997), perhaps because fast typers do more typing than



they could handwrite within the same time frame. Students with extended experience with word processors produced longer texts with significantly higher quality than they would have if they had handwritten their essays, (Harrington et al., 2000), while students with less experience were found to struggle with the cognitive demand of translating their ideas into keyboard strokes (Wolfe & Manalo, 2004) and might have written poorer essays (Powers et al., 1994). Research has found that students with less experience performed better when they handwrote their essay than when they typed on word processors (Wolfe et al., 1996; Russell, 1999; Russell & Haney, 1997). Compared to handwritten essays, word-processed essays were reported by extant studies to be neater (Wolfe et al., 1996) and seemed to have a more formal tone and weaker voice than handwritten essays, as perceived by raters (Lee, 2002; Wolfe et al., 1996).

Researchers have also considered the quality of revision that students make on their written works across modes and seemed to agree that word processors enhance students' capacity to review their work at levels not possible on paper. Students could easily copy, cut, paste, and totally delete written works with ease, which pen-and-paper based administration does not provide (Lee, 2004). Students are found to do more revisions when they type than they would when they handwrite their essay. However, it seems some students only do superficial revisions of words and characters. The research reported that increased revisions in typed essays do result in increased quality of written products, but most revisions are at surface level. Most students seem to pay less attention to sentence and structural level revisions (Lee, 2002; Wolfe et al., 1996). This inability to make higher order revisions may be a result of anxiety and tension associated with timed impromptu writing.

On the contrary, Jiang Li (2006) reported that students do higher level cognitive processing when they type on computers, and hence do more higher-level revisions when using computers than they would when they handwrote their essays. However, it is important to note that the students involved in that study already had considerable keyboarding experience. Observations from studies comparing the quality of handwritten and typed essays suggest that when English language ability is held constant, then keyboarding skill is a major influence on students' written works (Horkay et al., 2006).

Research has established that level of keyboarding skill and word-processor experience are key influences on students' performances on computer administered writing assessments. Fluency and English language ability are equally important determinants of the quality of written works. However, when English language ability was held constant, Hunter Breland, Yong-Won Lee, and Eiji Muraki (2005) found no relationship between students' performance on different essay prompts and mode in which tests were administered for 83 writing prompts on the TOEFL, thus suggesting that no prompt in itself is biased towards any particular mode of test administration. Although familiarity, or rather unfamiliarity, with a given prompt might interfere with a writer's performance on an essay, we may deduce that no prompt will favor handwriting above typing and, as such, will not affect the comparability of students' performances across modes.

Although research is divided on the comparability of the quality of handwritten and typed essays, some findings seem to uphold that students perform slightly better on handwritten essays than on computers (Breland et al., 2005; Russell & Tao, 2004b; Powers et al., 1994). For example, in the study by Breland et al. (2005), for all the 83 prompts examined over a 3-year period, handwritten essays were found to have scored higher than typed essays. Could this be a result of subtle differences across the two modes that impair rater impression?

It is unclear whether this effect is due to the students' writings themselves or to the raters' general impressions of the essays written in the two modes. It should be noted, however, that language proficiency, understanding of the prompt, and level of experience with word processors are key factors that influence students' performances in writing. Despite the changing characteristics of students entering our schools, it is obvious that these factors will still influence some category of test takers. With increasing computer accessibility, however, it is doubtless that keyboarding skill will continue to be an issue. Handwriting, however, might be the new issue for students' writing placement tests. Accommodating those placement writing test takers who might struggle with typing then might require that administrators give handwritten tests to students who so elect in order to give all test takers a fair choice. This means that raters may sometimes face the dilemma of rating the two kinds of written products within the same scoring session. Unfortunately, the unsuspecting rater may not recognize subtle differences in these formats and how they may influence judgment. Writings presented in both modes are different in systematic ways, as noted earlier, and such differences are found to interfere with raters' perceptions and judgments of students' written works in subtle ways, and may proffer some explanation for disparity in students' performances across modes.

#### 2.1.4. Effects of presentation mode on rater's judgment

Several studies have reported rater bias as a reason for disparity in the comparability of students' handwritten and typed essays (Russell & Tao, 2004a; Russell & Tao, 2004b; Powers et al., 1994; Arnold et al., 1990). Voiza Arnold et al. (1990) first observed that essays converted to word-processed forms received lower scores than in their original handwritten format. Power et al. (1994) transcribed word-processed essays to handwritten formats and vice versa and found that handwritten essays were scored higher irrespective of the mode in which they were initially composed. Russell and Tao (2004b) reported the same scenarios in a replica study.

Researchers have proposed various factors suspected for instigating rater bias in handwritten and word-processed essays. Raters may be biased because of the apparent length of handwritten essays (Powers et al., 1994); raters may have higher expectations for word-processed essays, seeing them as final products instead of as drafts written under impromptu timed conditions (Breland et al., 2005). Arnold et al. (1990) reported that raters in their studies might have been more lenient while scoring handwritten and harsher while scoring word-processed essays because they empathized with writers' efforts to make revisions—such as cancellations and cross-outs, which are lost when essays are word-processed—a phenomenon they termed as Reader Empathy Assessment Discrepancies Effects. More so, raters may have been more distracted by mechanical errors (typos) in typed essays, but may have given the writer a benefit of the doubt when scoring handwritten essays (Powers et al., 1994). Some raters also reported feeling closer to the writer in handwritten essays because handwritten essays had “stronger voice” than word-processed ones (Russell & Tao, 2004a).

Some studies examined the extent to which text appearances might have affected rater's perception. Russell and Tao (2004b), in a study of the effect of apparent length of essay on raters' perception, used single and double-spaced word-processed essays. They found that extending the length of typed essays did not eliminate rater bias. The study also reported that raters scored typed essays in scripted print higher than plain print, but less than handwritten essays. To mitigate rater bias due to presentation mode effect, Russell and Tao (2004b), as well as Powers et al. (1994), provided supplemental training to raters in their follow-up studies. Powers et al. (1994) found that such training diminished the effect. Hence, rater impression of the essay, coupled with students' individual differences, might be contributing to the differences in comparison across the two modes.

### 3. Closing the gap in comparability across modes

With typing becoming rather natural to students, just as handwriting used to be, it is increasingly necessary for administrators to consider providing placement essays in modes that are fair to all concerned. That fairness strongly suggests migrating to computer-based writing assessments, which seem to hold validity for most students. While changing to computer-based testing is important for writing assessments to remain authentic for the larger body of students, administrators might sometimes need to accommodate students in the minority—those not accustomed to typing, especially under timed testing conditions. As noted earlier, administering writing tests in only one mode may sideline test takers in the other category. Hence, despite the urgency of offering computer-based writing assessments, writing programs may sometime be constrained to offer handwritten placements tests options to accommodate students with minimal typing skills. While only a few students may require such accommodation, such offering would demonstrate that administrators are sensitive to the diverse nature of their student population, and are careful to ensure that the mode of administering placement exams and the mode of presenting students' written works to raters leaves their scores across the two modes as comparable as possible.

Given that administering placement writing assessment in paper-and-pen format only may restrain the capabilities of a majority of students who are adept computers users (Russell & Haney, 1997), administrators should consider transitioning to computer-based or even online-writing placement examinations, where such moves have not been made, in the interests of the larger student population. However, since colleges still enroll students with minimal or no computing skills, placement administrators might need to accommodate this category of students in order to provide contextually valid tests for all, thus giving students a choice in how they write their placement tests.

Where such choice is allowed, it may be helpful to inform students about how their computer skills, or the lack thereof, may affect their potential to write under a timed writing condition (Wolfe et al., 1996), thus empowering them to make informed decisions. Maulan (2004) inferred that a wrong decision on the mode a written test is administered might bias performance on the test. This could also mean that students are advised to practice their typing before they

sit for a computer-based written test. Test providers should give students the opportunity to familiarize themselves with the testing interface on computer-based tests.

Consequently, writing programs choosing to administer computer-based writing assessments would need to re-train raters who evaluate students' writings in both modes in order to mitigate the presentation mode effects, as suggested in the literature. Supplemental training sessions may highlight the following recommendations:

- Raters should be aware that typed essays appear shorter than handwritten essays. However, this effect does not translate into more word counts or longer and more complex sentences. Keeping to scoring rubrics (predetermined writer expectations) will keep raters' minds on sentence construction, paragraphing, support and integration, and organization, which are the essence of good writing. To reduce bias due to the apparent length of students' essay, raters should ignore the apparent length and appearance of typed essays.
- Training should notify that raters tend to give handwritten essays a benefit of the doubt when punctuations and spellings are not clear. On the contrary, typos in typed essays are more evident and can distract raters from the main idea(s) of the writer.
- Raters should keep in mind that typed essays of placement examinations are only first drafts, at best, written under timed condition. Hence, raters should make discretionary decisions between typos and mechanical errors, which are essential parts of scoring rubrics. Being stringent on typos might be tantamount to punishing students for their lack of typing skills.
- Training should notify raters of the apparent difference in formalities and voice or tone between handwritten and computer prints.
- The relative ease of reading computer prints may influence raters' general impression. Hence, it is important to keep the scoring rubrics (or pre-determined expectations) in mind while scoring students' essays.
- Training should highlight that raters may have a bias for handwritten essays because students' efforts to review are visible in handwritten essays, while the same efforts are lost when essays are typed.

Lee (2004) suggested that trainers should provide both handwritten and typed essay samples for training sessions, and Powers et al. (1994) posited that raters should give careful attention to low quality essays during training sessions because typed, poorly-written essays are most harshly rated. Russell and Tao (2004b) suggested that keeping a mental count of the number of mechanical errors observed while carefully reading responses during re-training sessions may help raters think carefully about factors that influenced their rating compared with their decisions at earlier pre-training reading sessions. Such reflective scoring is poised to help them better understand how the training has influenced the way they view essays presented in handwritten or computer print formats. Writing programs should also provide raters with some background demographic information of writers. Carl Whithaus, Scott B. Harrison, and Jeb Midyette, (2008) reported that semantic markers and cues that identify some ESL and LD are lost when they take computer-administered writing tests.

### 3.1. Other concerns with transitioning

Apart from student and rater factors that may influence computer-based assessments, tests administrators should be familiar with other factors vital to ensuring that the validity of computer-administered writing assessments is not compromised. It is important to thoughtfully consider the technical details and logistical implication of providing glitch-free online placement writing assessment tests, for instance. A thorough discussion of the technical implications of mode migration is beyond the scope of this article. However, test administrators would have to ensure testing platforms are thoughtfully considered. Testing platform interfaces should not stimulate construct irrelevant variants (e.g. extraneous cognitive loads engendered by navigating unfamiliar testing interfaces). Interfaces should be user friendly and easy to learn. Brent Bridgeman, Mary Lennon, and Altamese Jackenthal (2003) found that some students in their study reported that scrolling pages interfered with their tasks. Replacing screen-scrolling systems with a paging system where students could view snapshots of the pages they are working on might be an alternative to screen scrolling (Way et al., 2008; Higgins, Russell, & Hoffmann, 2005). The testing platform should provide students with such editing capabilities as the cut, copy, and paste functions—features that simulate their natural writing environment. Way et al. (2008) suggested these capabilities should integrate both known shortcuts and icons of these editing features that computer savvies are used to. Lastly, security concerns must be factored into any migration moves. Test administrators would want to



ensure that those taking tests are who they claim to be when considering online writing test administration. Unlike big testing companies that require candidates to take tests in designated centers all over the world, writing assessments administrators must find other means to offer supervised tests when they have to administer writing assessments online.

#### 4. Future exploration of transitioning issues

To ensure computer-based written assessments are as valid as paper-and-pen based tests used to be, research in the future should explore other factors that may impact students' writing capabilities when using computers—especially when they are writing under timed examination conditions. Students vary in the amount of pre-writing planning they do. When handwritten tests are administered, students could scribble, underline, or highlight points of interest. However, computer-based administrations might eliminate those possibilities. Research needs to explore how these changes impact students' ability to compose essays on computers and whether these changes significantly hamper essay planning and the quality of the written product.

Earlier research has shown that keyboarding skill can disrupt the writer's cognitive process, especially under timed testing conditions. Research is needed to determine what typical level of typing speed is sufficient before the mechanical action of typing begins to interfere with the writing process, and if this depends on other factors, such as the level of English language proficiency and the familiarity with the writing prompt. Knowing how the cognitive load associated with these factors interacts could provide informed guidance to candidates of writing assessment, especially those with poor English language and word-processing abilities.

A number of studies have identified that raters are partially influenced by mechanical errors because they are more evident in typed essays. However, more research is needed to understand how different types of grammatical errors other than spelling errors, which are most reported in existing literature, influence raters in order to better understand how to re-train raters.

#### 5. Conclusion

As technology pervades most aspects of daily living, writing with computers has become the norm and is influencing the way we learn. It is only natural to think that the way we assess learning or learning-related abilities should reflect how such learning has evolved. There is no other time than now when computer-based assessment of writing is more desirable than handwritten assessment. With the decline of the validity of handwritten writing assessments, and in the face of a growing need to administer students' writing placement examinations on computers, it is important for those who administer writing assessments to note that the need to transition to computer-administered writing tests far outweighs the desire to adhere to traditional practices or political whims that might have stalled this vital change all along. Given the present way learning and assessment are largely administered across the nation, paper-and-pen writing assessment seems to have lost its validity for most students, and the question we now face is not if we need to make that change or where the change is required, but what we need to know and put in place in order to ensure that computer-administered writing assessments have the same authenticity handwritten assessments used to have.

#### Acknowledgements

The author is grateful to Diane Kelly-Riley, director of the Writing Program, University of Idaho, and Sola Adesope of the Washington State University for their insight during the preparation of this manuscript, and Mabel Hunsu for proofreading the manuscript.

**Nathaniel Hunsu** is a graduate assistant at the Washington State University writing assessment program. He is currently a doctoral student of Educational Psychology. His research interests include educational measurement and assessments, the impact of technology on students' writing, writing portfolio and assessments. He is currently working on the impact of keyboarding skills on ESL writers in high-stakes writing assessments.

#### References

Alexander, Melody W., Bartlett, James E., Truell, Allens D., & Ouwenga, Karen. (2001). [Testing in computer technology course: An investigation of equivalency in performance between online and paper and pencil methods.](#) *Journal of Career and Technical Education, 18*(1), 69–80.

- American Educational Research Association (AERA); American Psychological Association (APA); & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: AERA Publications.
- Arnold, Voiza; Legas, Julia; Obler, Susan.; Pacheco, Mary Ann; Russell, Carolyn; & Umbdenstock, Linda. (1990). Do students get higher scores on their word-processed paper? A study of bias in scoring hand-written vs. word-processed papers. *The Educational Resources Center*. Whitter, CA Rio Hondo College.
- Association of American Universities (AAU). (2012). Retrieved from <https://www.aau.edu/home.aspx>
- Boulet, John R., McKinley, Danette, Rebecchi, Thomas, & Whelan, Gerald. (2007). Does composition medium affect the psychometric properties of scores on an exercise designed to assess written medical communication skills? *Advances in Health Sciences Education*, 12(2), 157–167.
- Breland, Hunter, Lee, Yong-Won, & Muraki, Eiji. (2005). Comparability of TOEFL CBT essay prompts: Response-mode analyses. *Educational and Psychological Measurement*, 65(4), 577–595.
- Bridgeman, Brent, & Cooper, Peter. (1998, April). *Comparability of scores on word-processed and handwritten essays on the Graduate Management Admission Test*. Paper presented at the annual meeting of the American Educational Research Association. San Diego, CA. Abstract retrieved from <http://eric.ed.gov/?id=ED421528>
- Bridgeman, Brent, Lennon, Mary L., & Jackenthal, Altamese. (2003). Effects screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education*, 16(3), 191–205.
- Council of Writing Program Administrators (CWPA). (2014). *NCTE-WPA white paper on writing assessment in colleges and universities*. Retrieved from <http://wpacouncil.org/whitepaper>
- Chen, Jing, Whiteb, Sheida, McCloskeyc, Michael, Sorouid, Jaleh, & Chune, Young. (2011). Effects of computer versus paper administration of an adult functional writing assessment. *Assessing Writing*, 16(1), 49–71.
- DeBell, Matthew, & Chapman, Chris. (2006). *Computer and Internet use by students in 2003*. Washington, DC: National Center for Education Statistics.
- Eysenck, Michael W. (1988). Anxiety and attention. *Anxiety Research*, 1(1), 9–15.
- Gos, Michael W. (1996). Computer anxiety and computer experience: A new look at an old relationship. *The Clearing House*, 65(5), 271–276.
- Harrington, Susanmarie, Shermisa, Mark D., & Rollinsa, Angela L. (2000). The influence of word processing on English placement test results. *Computer and Composition*, 17(2.), 197–210.
- Higgins, Jennifer; Russell, Michael. & Hoffmann, Thomas. (2005). Examining the effect of computer-based passage presentation on reading test performance. *Journal of Technology, Learning, and Assessment*, 3(4). Retrieved from <http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1657/1499>
- Horkay, Nancy, Bennett, Randy E., Allen, Nancy, Kaplan, Bruce A., & Yan, Fred. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 5(2). Retrieved from <http://files.eric.ed.gov/fulltext/EJ843858.pdf>
- Lee, Young-Ju. (2002). A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessment Writing*, 8(2), 135–157.
- Lee, H. K. (2004). A comparative study of ESL writers' performance in paper-based and a computer-delivered writing test. *Assessment Writing*, 9(1), 4–26.
- Li, Jiang. (2006). The mediation of technology in ESL writing and its implications for writing and its implications for writing assessment. *Assessment Writing*, 11(1), 5–21.
- MacCann, Robert, Eastment, Benjamin, & Pickering, Samantha. (2002). Responding to free response examination questions: Computer versus pen and paper. *British Journal of Educational Technology*, 33(2), 173–188.
- Marcoulides, George A. (1988). The relationship between computer anxiety and computer achievement. *Journal of Educational Computing Research*, 4(2), 151–158.
- Maulan, Sumarni. (2004). Language testing using computers: Examining the effect of test-delivery medium on students' performance. *Journal of e-Language & Teaching*, 1(2), 1–14.
- McDonald, Angus S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computer and Education*, 39(3), 299–312.
- Messick, Samuel. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher*, 23(2), 13–23.
- Messick, Samuel. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Moss, Pamela A. (1995). Themes and variations in validity theory. *Educational Measurement: Issues and Practice*, 14(2), 5–13.
- National Center for Education Statistics (NCES). (2011). *NAEP writing computer-based assessment: An overview for Grade 4*. Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/about/schools/2012grade4wcbabrochure.pdf>
- Powers, Donald E., Fowles, Mary E., Farnum, Marisa, & Ramsey, Paul. (1994). Will they think less of my handwritten essay if others word process theirs? Effect on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement*, 31(3), 220–223.
- Russell, Michael. (1999). Testing writing on computers: A follow-up study comparing performance on computer and on paper. *Educational Policy Analysis Archives*, 7(20)
- Russell, Michael, & Haney, Walt. (1997). Testing writing on computers: An experiment comparing students' performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis Archives*, 5(3).
- Russell, Michael, & Plati, Tom. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Retrieved from <http://www.bc.edu/research/intasc/PDF/ComputerVsPaperStateWriting.pdf>
- Russell, Michael, & Tao, Wei. (2004a). Effects of handwriting and computer-print on composition scores: A follow-up to Power, Fowles, Farnum, & Ramsey. *Practical Assessment, Research & Evaluation*, 9(1). Retrieved from <http://pareonline.net/getvn.asp?v=9&n=1>

- Russell, Michael, & Tao, Wei. (2004b). The influence of computer-print on rater scores. *Practical Assessment, Research & Evaluation*, 9(10). Retrieved from <http://pareonline.net/getvn.asp?v=9&n=10>
- Way, Walter D.; Davis, Laurie L.; & Fitzpatrick, Steven. (2006, April). *Score comparability of online and paper administrations of the Texas assessment of knowledge and skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA. Retrieved from [http://images.pearsonassessments.com/images/tmrs/Score\\_Comparability\\_of\\_Online\\_and\\_Paper\\_Administrations\\_of\\_TAKS\\_03\\_26\\_06\\_final.pdf](http://images.pearsonassessments.com/images/tmrs/Score_Comparability_of_Online_and_Paper_Administrations_of_TAKS_03_26_06_final.pdf)
- Way, Walter D.; Davis, Laurie L.; & Strain-Seymour, Ellen. (2008). The validity case assessing direct writing by computer: A Pearson assessments & information white paper. Retrieved from [http://images.pearsonassessments.com/images/tmrs/tmrs\\_rg/TheValidityCaseforOnlineWritingAssessments.pdf?WT.mc\\_id=TMRS\\_The\\_Validity\\_Case\\_for\\_Assessing\\_Direct](http://images.pearsonassessments.com/images/tmrs/tmrs_rg/TheValidityCaseforOnlineWritingAssessments.pdf?WT.mc_id=TMRS_The_Validity_Case_for_Assessing_Direct)
- Whithaus, Carl, Harrison, Scott B., & Midyette, Jeb. (2008). Keyboarding compared with handwriting on high-stakes writing assessment: Students choice of composing medium, raters' perception, and text quality. *Assessment Writing*, 13(1), 4–25.
- Wolfe, Edward W., Bolton, Sandra, Feltovich, Brian, & Niday, Donna M. (1996). The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessment Writing*, 3(2), 123–147.
- Wolfe, Edward W., & Manalo, Jonathan R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning & Technology*, 8(1), 53–65.