



**Title of Project:**

Investigating the Authenticity of Paper- and  
Computer-Based ESL Writing Tests

**Researcher:**

Mingxia Zhi  
[zhimingxia1990@gmail.com](mailto:zhimingxia1990@gmail.com)



Mingxia Zhi

*Current Affiliation*

Northside Independent School District

*Doctoral Institution*

The University of Texas at San Antonio

**Research Supervisor:**

Dr. Becky Huang  
The University of Texas at San Antonio

**TIRF Research Topic Investigated:**

Language Assessment

---

**Final Report**

**Motivation for the Research**

The use of technology in the administration of language assessments has received extensive research attention. Because the majority of students' writing and learning activities are now mediated by computers (Endres, 2012; Kohler, 2015), many researchers have argued for using computer-based (CB) ESL writing tests, resting on the assumption that CB writing is more authentic than paper-based (PB) writing task (Cheung, 2016; Kohler, 2015; Lee, 2004; Lessien, 2013). Such assumptions, however, are supported by little empirical evidence. Little research has investigated the authenticity of English as a second language (ESL) writing assessments delivered in PB and CB formats. The authenticity of a test is defined as "the degree of correspondence of the characteristics of a given language test task to the features of a target language use (TLU) task" (Bachman & Palmer, 1996, p. 23). The correspondence of physical characteristics of the test task to TLU task is influential on test takers' language use (Bachman & Palmer, 2010). Test authenticity is important because it relates to the domain of generalization of a test (Bachman & Palmer, 2010) and has the potential effect on test-takers' perceptions of the test (perceived relevance). This study filled the research gap by examining the effect of test modality (PB and CB) on the authenticity of ESL academic writing tests.

**Research Questions**

Informed by Liu's (2005) conceptual model of five aspects of authenticity (TLU task, test task, learner interpretation, learner involvement, and learner response), this dissertation study

compared the authenticity of CB and PB ESL writing tests. The research questions (RQs) are the following:

1. To what extent do second language (L2) writing processes reported in daily English Writing tasks correspond to writing processes reported in the CB and PB writing tests?
2. Does test-takers' computer familiarity correlate with their CB and PB test essay scores?
3. What are the test-takers' perceived connections between the CB and PB writing tests and daily English writing tasks?
4. Does test-takers' perceived authenticity of the task medium correlate with their perceived validity and preferences of the testing mode?

### **Research Methodology**

This study employed the embedded correlational model of mixed methods design (Creswell & Plano Clark, 2011). Sixty International ESL students ( $N = 60$ ) were recruited using convenient sampling and snowball sampling methods. Data collection was carried out in three stages (Figure 1). Prior to the study, the participants were recruited using a pre-survey. Then, the participants were grouped into four groups, and group equivalence was attained by matching the participants on gender, age range, score of the standardized writing tests (TOEFL/IELTS), and self-ratings of computer familiarity. The counter-balanced measures are used to solve the issue of potential carryover effect, fatigue effect, and context effect (Price, Chiang, & Jhangiani, 2015). During the study, the participants first completed a one-minute online typing test and then completed the two assigned writing tasks. After the study, a post-survey and an interview were conducted with the participants to capture their perceptions of the two writing test tasks. To analyze the writing samples, the handwritten essays were transcribed verbatim into a typed form to ensure the same presentation of the essays to the rater. Each essay was rated by at least two raters using a holistic scale and an analytic rating scale. For RQ 1, bivariate correlation analysis (Spearman rho) was conducted to examine the relationships between the ratings on the process items in the pre-survey and the ratings on the cognitive writing questionnaire items in the CB and PB tasks, respectively. Correlation coefficients were compared with an updated version of Steiger's Z test (Steiger, 1980) using a web-based calculator for testing the statistical significance of the difference between dependent correlations (Hoerger, 2013). For RQ 2, Spearman's rank order correlation test was used to test the bivariate correlations between the survey Likert-scale items, such as computer use, and the essay scores. Two multiple regression analyses were conducted to predict the scores of CB-and PB-generated essays, respectively, using the typing speed, typing accuracy, after controlling for the standardized writing test scores. For RQ 3, descriptive statistical analysis was conducted with the ratings on post-survey item. For RQ 4, bivariate Spearman's rank order correlation analysis was employed between the ratings on post-survey item. The two cycle coding method was used to analyze the qualitative data, with In-Vivo coding for the first cycle and focused coding for the second cycle.

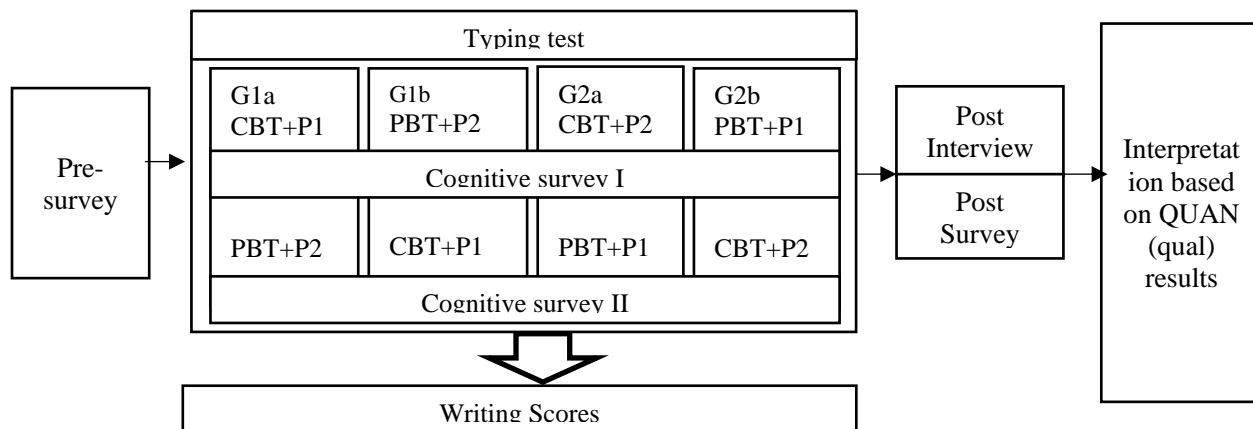


Figure 1. Illustration the data collection procedure.

## Summary of Findings

**RQ 1.** Results suggested that both mediums of writing tests, to different degrees, yielded writing processes that were consistent with the TLU writing processes. Further test of the statistical significance of the difference between dependent correlations using Hoerger's (2013) calculations indicated that the CB writing only yielded higher correlations than the PB writing test on three items: thinking of genre knowledge during planning ( $Z_H = 2.68^{**}$ ,  $p = .007$ ), making pauses during writing for reviewing ( $Z_H = 2.189^*$ ,  $p = .028$ ), and making pauses during writing for revising ( $Z_H = 3.328^{***}$ ,  $p < .001$ ). Therefore, the quantitative results suggested significant but limited evidence for a higher degree of authenticity of the CB writing test. Qualitative results suggested that more participants perceived the CB writing process to be more similar to TLU writing process, thanks to a frequent use of computer for studying and work. Further analysis revealed that many test-takers were "forced" by the PB medium to constantly check during composing due to the difficulties of conducting major revisions after writing. This might have contributed to the low correlation relationship between the PB writing and TLU tasks in the survey item -- d4 "I pause after a few sentences or a whole paragraph to read it again". Taken together, the results partially supported the claim that the CB writing test corresponded more to the TLU tasks than the PB writing task.

**RQ 2.** The results showed that the computer availability at school significantly and positively correlated with the CB essay holistic scores, specifically, with the grammar and accuracy scores. Nevertheless, computer availability at school variable did not correlate with the PB essay holistic scores. To further examine the relevancy of typing proficiency in the L2 CB and PB writing assessment, I conducted two multiple regression analyses to examine the effect of test-takers' typing speed and accuracy on their CB-and PB-generated writing scores. Results suggested typing accuracy variable significantly and positively predicted the CB essay scores, but not PB essay scores, after controlling for their standardized writing scores. Qualitative results corroborated with the quantitative results in that the majority of the participants perceived an influence of their computer skills on their CB writing performance. Most test-takers reported this influence to be positive, for the reason that CB writing was a more familiar medium for more test-takers. Computers also provided writers with more time and convenience for revising. Nonetheless, some students felt they were penalized by this test medium due to a lack of typing

skills in English. A small group of participants did not see a significant influence. Taken together, the results demonstrated an effect of computer familiarity and typing skills on test-takers' CB writing outcomes in this study. In particular, it was the computer availability and use at school, as well as the typing speed that made a statistical difference.

**RQ 3.** Both survey data and interview responses reported more participants perceived a higher correspondence between CB writing and TLU tasks compared to the PB test, indicating a higher perceived authenticity of the CB writing condition for assessing L2 writing. Only eight percent of the participants perceived the two tests as equally authentic. Further qualitative analysis revealed that although most students reported using computers in college studies, there was a discrepancy in assignment practices between U.S. colleges and schools in other countries. Some students shared that electronic assignments had not become a common practice in their home countries. Though having personal access to computers, computers were not always available or frequently used at schools. The results, though in favor of CBT for assessing writing, also brought concerns with the fairness of using computers for administering high-stakes writing test with test-takers from diverse backgrounds.

**RQ 4.** Results showed that test-takers who perceived the PB test more authentic rated it more valid, prefer to take the PB test, and rated CB test less valid. The ratings on the perceived authenticity of the CB writing, on the other hand, did not yield significant correlations to the perceived validity of CB or PB writing assessment. The perceived performance on the test was found to significantly and positively correlate with the perceived validity of the testing medium. Finally, typing speed significantly negatively correlated with the preference to the PB test. This means that the faster one typed, the less likely he/she preferred to take the PB test for writing. In the post interview, the participants explained that when they were more comfortable with the test medium, they could engage in a more routinized writing process, just like when they wrote in real-life academic English tasks in schools. Hence, they might perform better in the test. Taken together, the results supported a correlation relationship between the perceived authenticity of the testing medium and the consequential validity of the test.

### **Implications**

This study uncovered both advantages and limitations of the CB writing test. The CB test was generally more authentic than the PB test for the majority of the test-takers. However, it also brought concerns in the construct relevance of computer literacy, fairness, and its consequence. The findings of this study have implications for test developers and university test administrators. First, test developers should consider the characteristics of the test-takers when developing high-stakes tests. Considering the various backgrounds of the ESL learners from different countries in the world, it is important to offer tests in more than one medium to adapt for all test-takers to avoid test biases. The digital divide within the test-taker population may be a reflection of the divide in socio-economic status (SES) of the international ESL learners. Second, when interpreting the test results, school admission personnel should keep in mind that the construct tested in CB and PB modes of the test may be different. Other than English language proficiency, the student candidates' computer literacy skills may also be tested by the CB test of writing. In addition, test purpose is another factor to consider when choosing the medium for ESL writing tests. Different test mediums may be chosen for high-stakes English proficiency tests for college admission purposes and local in-house ESL assessments for diagnostic purposes. Finally, when comparing the equivalence of the two mediums of L2 writing tests, one key concern from the participants' perspective was the time. For test-takers with higher computer



skills, typing seemed to earn more time for more deliberate planning and revisions. Similarly, for students who were more comfortable with handwriting, the lower their typing skills were, the lower their essay scores were on the CB test. It was perceived to be a key reason for lower self-rated performance on the test when participants had less time to plan, review, and revise due to a disadvantage in typing. Allowing more time in both testing conditions may mitigate this issue. More research is needed to investigate the ideal duration of the test with both objective measures and perspectives from the test-takers.

### References

- Alibakhshi, G., Kiani, G. R., & Akberi, R. (2010). Authenticity in ESP/EAP selection tests administered at Iranian universities. *The Asian ESP Journal*, 6(2), 64-92.
- Ameera R. A. (2019). Impact of digital technology on economic life in India—a special reference to digital divide in rural-urban India. *International Journal of Science, Engineering, and Management*, 4(6), 83 – 88.
- Anthony, J. J. (2009). Classroom computer experiences that stick: Two lenses on reflective timed essays. *Assessing Writing*, 14(3), 194–205.
- Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36-44.
- Aydin, S. (2006). The effect of computers on the test and inter-rater reliability of writing tests of ESL learners. *TOJET: The Turkish Online Journal of Educational Technology*, 5(1), 75-81.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671-704.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy, & Practice*, 18(3), 279–293.
- Barkaoui, K. (2014). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL-iBT writing tasks. *Language Testing*, 31(2), 241-259.
- Barkaoui, K. (2015). Test takers' writing activities during the TOEFL iBT® writing tasks: A stimulated recall study. *ETS Research Report Series*, 2015(1), 1-42.
- Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal*, 100(1), 320-340.
- Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 1(2), 1-15.
- Bernhardt, S. A., Edwards, P. R., & Wojahn, PG. (1989). Teaching college composition with computers: A program evaluation study. *Written Communication*, 6(1), 108-133.





- Breland, H. M. (1983). The direct assessment of writing skill: A measurement review. *ETS Research Report Series, 1983(2)*.
- Bridgeman, B., & Carlson, S. B. (1984). Survey of academic writing tasks. *Written Communication, 1(2)*, 247-280.
- Brown, A. (1993). The role of test-taker feedback in the test development process: Test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing, 10(3)*, 277-301.
- Brown, H. D., & Abeywickrama, B. (2010). *Language assessment: Principles and classroom practices*. Pearson Education. Inc.
- Cardell, C. D., & Nickel, P. M. (2003). *Computer proficiency testing in higher education: Impetus and implementation*. Paper presented at the 84<sup>th</sup> Annual meeting of the American Educational Research Association, Chicago, IL.
- Chambers, L. (2011). Composition and revision in computer-based written assessment. *Research Notes, 1(43)*, 25-32.
- Chan, S., Bax S., & Weir. C. (2017). Researching participants taking IELTS Academic Writing Task 2 (AWT2) in paper mode and in computer mode in terms of score equivalence, cognitive validity and other factors. *IELTS Research Reports Online Series, No. 4*. British Council, Cambridge English Language Assessment, & IDP: IELTS Australia.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Ernst Klett Sprachen.
- Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Sage.
- Chen, P., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. SAGE Publications Inc.
- Chen, W., & Wellman, B. (2004). The global digital divide-within and between countries. *IT & Society, 1(7)*, 39-45.
- Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment, 16(2)*, 104-122.
- Cheung, Y. L. (2012). Critical review of recent studies investigating effects of word processing-assisted writing and pen-and-paper writing on the quality of writing and higher level revisions. *Procedia-Social and Behavioral Sciences, 46(2012)*, 1,047-1,050.
- Cheung, Y. L. (2016). A comparative study of paper-and-pen versus computer-delivered assessment modes on students' writing quality: A Singapore study. *The Asia-Pacific Education Researcher, 25(1)*, 23-33.



- Choi, I. C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295-320.
- Collier, R., & Werier, C. (1995). When computer writers compose by hand. *Computers and Composition*, 12(1), 47-59.
- Compare TOEFL® Scores. (n.d.). Retrieved January 22, 2019, from <https://www.ets.org/toefl/institutions/scores/compare/>
- Coniam, D. (2006). Evaluating computer-based and paper-based versions of an English-language listening test. *ReCALL*, 18(2), 193-211.
- Cooper, P. L. (1984). The assessment of writing ability: A review of research. *ETS Research Report Series*, 1984(1), i-46.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Sage Publications.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). American Council on Education.
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621-694). American Council on Education.
- Cushing Weigle, S. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-287.
- Daiute, C. (1986). Physical and cognitive factors in revising: Insights from studies with computers. *Research in the Teaching of English*, 20(2), 141-159.
- Dalton, D. W., & Hannafin, M. J. (1987). The effects of word processing on written composition. *The Journal of Educational Research*, 80(6), 338-342.
- Dey, I. (1999) *Grounding grounded theory: Guidelines for qualitative inquiry*. Academic Press.
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge University Press.
- East, M. (2016). *Assessing foreign language students' spoken proficiency*. Springer.
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer? *Language Testing*, 19(4), 347-368.
- Educational Testing Services. (2009). *Official guide to the TOEFL test* (3<sup>rd</sup> edition). McGraw-Hill Companies.





- Educational Testing Services. (2012). *Official guide to the TOEFL test* (4<sup>th</sup> edition). McGraw-Hill Companies.
- Endres, H. (2012). A comparability study of computer-based and paper-based writing tests. *Research Notes*, 49(2012), 26-33.
- Heck, R. H., & Crislip, M. (2001). Direct and indirect writing assessments: Examining issues of equity and utility. *Educational Evaluation and Policy Analysis*, 23(3), 275-292.
- Feng, L., Lindner, A., Ji, X. R., & Joshi, R. M. (2017). The roles of handwriting and keyboarding in writing: A meta-analytic review. *Reading and Writing*, 32(1), 33-63.
- Ferris, D., & Hedgcock, J. (2013). *Teaching L2 composition: Purpose, process, and practice*. Taylor and Francis.
- Flores, B. B., Casebeer, C. M., & Riojas-Cortez, M. (2011). Validation of the early childhood ecology scale-revised: A reflective tool for teacher candidates. *Journal of Early Childhood Teacher Education*, 32(3), 266-286.
- Flores, B. B., Clark, E. R., Guerra, N. S., Casebeer, C. M., Sánchez, S. V., & Mayall, H. J. (2010). Measuring the psychosocial characteristics of teacher candidates through the Academic Self-Identity: Self-Observation Yearly (ASI SOY) Inventory. *Hispanic Journal of Behavioral Sciences*, 32(1), 136-163.
- Frase, L. T., Kiefer, K. E., Smith, C. R., & Fox, M. L. (1985). Theory and practice in computer-aided composition. In S. W. Freedman (Ed.). *The acquisition of written language*. Ablex Publishing Corporation.
- Gambell, T. J. (1991). University education students' self-perceptions of writing. *Canadian Journal of Education/Revue canadienne de l'éducation*, 16(4), 420-433.
- Hardy, C., Heeler, P., & Brooks, D. (2006). Are high school graduates technologically ready for post-secondary education? *Journal of Computing Sciences in Colleges*, 21(4), 52-60.
- Hawisher, G. E. (1987). The effects of word processing on the revision strategies of college freshmen. *Research in the Teaching of English*, 21(2), 145-159.
- Hayes, J. R. (1996). *The science of writing*. Lawrence Erlbaum.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In I. W. Gregg & E.R. Steinberg (eds.), *Cognitive processes in writing* (pp. 3-30). New York, NY: Routledge.
- Hoa, N. T. N. (2018). Test takers' attitudes to the test content of the two listening tests: IELTS and TOEFL iBT. *VNU Journal of Foreign Studies*, 34(4).
- Hoekje, B., & Linnell, K. (1994). "Authenticity" in language testing: Evaluating spoken language tests for international teaching assistants. *Tesol Quarterly*, 28(1), 103-126.



- Hoerger, M. (2013). Z<sub>H</sub>: An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from [http://www.psychmike.com/dependent\\_correlations.php](http://www.psychmike.com/dependent_correlations.php)
- Hoomanfar, M. H., & Meshkat, M. (2015). Writing on a computer and using paper and pencil: Is there any difference in the internal cognitive processes? *GEMA Online® Journal of Language Studies*, 15(2), 17-31.
- Huang, B. H. (2013). The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech. *System*, 41(3), 770-785.
- Hunsu, N. J. (2015). Issues in transitioning from the traditional blue-book to computer-based writing assessment. *Computers and Composition*, 35(1), 41-51.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for applying Fusion Model to language assessment. *Language Testing*, 26(1), 31-73.
- Jang, E. E., Wagner, M., & Park, G. (2014). Mixed methods research in language testing and assessment. *Annual Review of Applied Linguistics*, 34(1), 123-153.
- Johnson, M. A. (1988). Word processing in the English as a second language classroom. In: J. L. Hoot & S. B. Silvern (Eds.), *Writing with computer in the early grades* (pp. 107-121). New York, NY: Teachers College Press.
- Joram, E., Woodruff, E., Bryson, M., & Lindsay, P. H. (1992). The effects of revising with a word processor on written composition. *Research in the Teaching of English*, 26(2), 167-193.
- Kalikorha, C., Strauss, P., & Smedley, F. (2009). The perceptions of first-year undergraduate Malawian students of the essay writing process. *Africa Education Review*, 6(1), 37-54.
- Kimmons, R., Darragh, J. J., Haruch, A., & Clark, B. (2017). Essay composition across media: A quantitative comparison of 8th grade student essays composed with paper vs. Chromebook. *Computers and Composition*, 44(1), 13-26.
- Kitchin, D. A. (1991). *Case study of ESL community college students using computer-based writing tools in composition course* (Unpublished doctoral dissertation). University of San Francisco, California.
- Kohler, B. (2015). Paper based or computer-based essay writing: Differences in performance and perception. *Linguistic Portfolios*, 4(1), 13.
- Kurth, R. J. (1987, April). *Word processing and composition revision strategies* (ERIC Document Reproduction Service No. ED 283195). Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.



- Lam, F. S., & Pennington, M. C. (1995). The computer vs. the pen: A comparative study of word processing in a Hong Kong secondary classroom. *Computer-Assisted Language Learning*, 8(1), 75-92.
- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*. Routledge.
- Laurier, M. (2000). Can computerised testing be authentic? *ReCALL*, 12(1), 93-104.
- Lee, Y. J. (2002). A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessing Writing*, 8(2), 135-157.
- Lee, H. K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing*, 9(1), 4-26.
- Lee, H. K. (2008). The relationship between writers' perceptions and their performance on a field-specific writing test. *Assessing Writing*, 13(2), 93-110.
- Lee, Y. J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research*, 1(4), 366-389.
- Lee, A. Y., & Wang, K. J. (2019). Young people's media use and social participation in Hongkong: A perspective of digital use divide. In E. Morrell & J. Rowsell (Eds.), *Stories from inequity to justice in literacy education: Confronting digital divides* (pp. 130-148). Routledge.
- Lee, H., & Winke, P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing*, 30(1), 99-123.
- Lessien, E. (2013). *The effects of typed versus handwritten essays on students' scores on proficiency tests* (Unpublished M.A. TESOL thesis). Michigan State University.
- Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. *Language Testing*, 17(1), 43-64.
- Li, J. (2006). The mediation of technology in ESL writing and its implications for writing assessment. *Assessing Writing*, 11(1), 5-21.
- Li, Z. (2013). The issues of construct definition and assessment authenticity in video based listening comprehension tests: Using an argument-based validation approach. *International Journal of Language Studies*, 7(2).
- Li, J., & Cumming, A. (2001). Word processing and second language writing: A longitudinal case study. *International Journal of English Studies*, 1(2), 127-152.



- Lichtenstein, N. (1996). *The effect of word processing on writing achievement* (ERIC Document Reproduction Service No. ED 394146). Research Report, M.A. Project, Kean College of New Jersey.
- Liu, H. M. (2005). An investigation of methods for assessing authenticity in computer-assisted language learning and assessment. *Retrospective Theses and Dissertations*. 269(1). Iowa State University.
- Lutz, J. (1987). A study of professional and experienced writers revising and editing at the computer with pen and paper. *Research in the Teaching of English*, 21(4), 398-421.
- Manchon, R. M., Murphy, L., & de Larios, J. R. (2005). Using concurrent protocols to explore L2 writing processes: Methodological issues in the collection and analysis of data. In P. K. Matsuda & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 191-205). Routledge.
- Mangen, A., & Velay, J. L. (2010). Digitizing literacy: Reflections on the haptics of writing. In M. H. Zadeh (Ed.). *Advances in haptics* (pp. 385-401). InTech.
- McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*, 39(3), 299-312.
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). American Council on Education & National Council on Measurement in Education.
- Messick, S. (1994). Standards-based score interpretation: Establishing valid grounds for valid inferences. *ETS Research Report Series*, 1994(2), 291-305.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8.
- Messick, S. (1996). Validity and washback in language testing. *ETS Research Report Series*, 1996(1).
- Moore, T., & Morton, J. (1999): Authenticity in the IELTS Academic Module Writing Test: A comparative study of Task 2 items and university assignments. *IELTS Research Reports*.
- Moore, T., Morton, J., Hall, D., & Wallis, C. (2015). Literacy practices in the professional workplace: Implications for the IELTS reading and writing tests. *IELTS Research Reports Online Series*, 46(1), 1-46.



- Nelson, K., Courier, M., & Joseph, G. W. (2011). Teaching tip an investigation of digital literacy needs of students. *Journal of Information Systems Education*, 22(2), 95-109.
- Neuwirth, C. M., Haas, C., & Hayes, J. (1990). Does word processing improve students' writing? A critical appraisal and assessment. *Final Report to FIPSE*.
- Norris, J. T., Pauli, R., & Bray, D. E. (2007). Mood change and computer anxiety: A comparison between computerized and paper measures of negative affect. *Computers in Human Behavior*, 23(6), 2875-2887.
- Noubandegani, P. A. (2012). Students' perceptions of computerized TOEFL test. *Language Testing in Asia*, 2(2), 73-101.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners*. Boston, MA: Addison-Wesley Publishing Company
- Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing*, 23(1), 17-13.
- Ostovar-Namaghi, S. A. (2011). A comparative study of test tasks and target use tasks. *Theory and Practice in Language Studies*, 1(5), 525-529.
- Owston, R. D., Murphy, S., & Wideman, H. H. (1992). The effects of word processing on students' writing quality and revision strategies. *Research in the Teaching of English*, 26(3), 249-276.
- Owston, R. D., & Wideman, H. H. (1997). Word processors and children's writing in a high-computer-access setting. *Journal of Research on Computing in Education*, 30(2), 202-220.
- Pennington, M. C. (1993). A critical examination of word processing effects in relation to L2 writers. *Journal of Second Language Writing*, 2(3), 227-255.
- Petrić, B., & Czár, B. (2003). Validating a writing strategy questionnaire. *System*, 31(2), 187-215.
- Plata, M. (2008). Looking beyond undergraduates' attitude about a university-wide writing requirement. *Journal of Instructional Psychology*, 35(4), 365-376.
- Poonpon, K. (2011, April). *Synergy of mixed method approach to development of ESL speaking rating scale*. Paper presented at Doing Research in Applied Linguistics Conference, Bangkok, Thailand.
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Education Measurement*, 31(3), 220-233.





- Powers, D. E., Kim, H. J., Yu, F., Weng, V. Z., & VanWinkle, W. (2009). The TOEIC® speaking and writing tests: Relations to test-taker perceptions of proficiency in English. *ETS Research Report Series*, 2009(1).
- Price, P. C., Chiang, I. C. A., & Jhangiani, R. (2015). *Research methods in psychology*. BC campus, BC Open Textbook Project.
- Richards-Mealy, L. (2018). *A quantitative study to analyze new student digital literacy knowledge and skills at a four-year institution* (Doctoral dissertation). Northcentral University).
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104.
- Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7(20), 1–47.
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis Archives*, 5(3), 1–20.
- Russell, M., & Tao, W. (2004a). Effects of handwriting and computer-print on composition scores: A follow-up to Powers, Fowles, Farnum, & Ramsey. *Practical Assessment, Research & Evaluation*, 9(1), 1-9.
- Russell, M., & Tao, W. (2004b). The influence of computer-print on rater scores. *Practical Assessment, Research & Evaluation*, 9(10), 1-13.
- Saldaña, J. (2016). *The coding manual for qualitative researchers*. Sage.
- Sarbakhshian, B., & Saeidi, M. (2016). The comparison of typed and handwritten essays of Iranian EFL students in terms of length, spelling, and grammar. *The Journal of Applied Linguistics*, 9(19), 104-118.
- Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*, 4(2), 38-59.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.
- Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing* (Vol. 26). Cambridge University Press.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-24.





- Shirzad, M., & Shirzad, H. (2017). The effect of computer literacy on the participants' writing ability in TOEFL iBT. *Theory and Practice in Language Studies*, 7(2), 134-139.
- Shohamy, E. (1982). Affective considerations in language testing. *The Modern Language Journal*, 66(1), 13-17.
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14(3), 340-349.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Routledge.
- Shohamy, E., Gordon, C. M., & Kraemer, R. (1992). The effect of raters' background and training on the reliability of direct writing tests. *The Modern Language Journal*, 76(1), 27-33.
- Snyder, I., & Prinsloo, M. (2007). Young people's engagement with digital literacies in marginal contexts in a globalised world. *Language and Education*, 21(3), 171-179.
- Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31(4), 378-388.
- Sommers, E. A. (1985). The effect of word processing and writing instruction on the writing processes and products of college writers (ERIC Document Reproduction Service No. ED 269762).
- Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, 18(4), 463-481.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251.
- Stiggins, R. J. (1982). A comparison of direct and indirect writing assessment methods. *Research in the Teaching of English*, 16(2), 101-114.
- Sullivan, N., & Pratt, E. (1996). A comparative study of two ESL writing environments: A computer-assisted classroom and a traditional oral classroom. *System*, 24(4), 491-501.
- Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). The relationship between computer familiarity and performance on computer-based TOEFL test tasks. *TOEFL Research Reports 61: ETS Research Report 98-8*. Educational Testing Service.
- Taylor, C., Kirsch, I., Eignor, D., & Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning*, 49(2), 219-274.
- Teichman, M., & Poris, M. (1989). Initial effects of word processing on writing quality and writing anxiety of freshman writers. *Computers and the Humanities*, 23(2), 93-103.



- Understand and Explain the IELTS Scores. (n.d.). Retrieved from <https://takeielts.britishcouncil.org/teach-ielts/test-information/scores-explained>
- Van Dijk, J. A. G. M. (2002). A framework for digital divide research. *Electronic Journal of Communication, 12*(1), 2.
- Van Dijk, J. A. (2006). Digital divide research, achievements and shortcomings. *Poetics, 34*(4-5), 221-235.
- Veal, L. R., & Hudson, S. A. (2009). Direct and indirect measures for large-scale evaluation of writing. In B. Huot & P. O'Neill (Eds.), *Assessing writing: A critical sourcebook* (pp. 290-296). Bedford/St. Martin's.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave.
- Weir, C., O'Sullivan, B., Jin, Y., & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS writing component: Effects and impact. *IELTS Research Reports, 7*(1), 1-37.
- Wolfe, E. W., & Manalo, J. R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning & Technology, 8*(1), 53-65.
- Wolfe, E. W., & Manalo, J. R. (2005). An investigation of the impact of composition medium on the quality of TOEFL writing scores. *TOEFL® Research Report. RR-72. ETS RR-04-29*. ETS Research Report Series.
- Womble, G. (1984). Process and processor: Is there room for a machine in the English classroom? *English Journal, 73*(1), 34-37.
- Whithaus, C., Harrison, S. B., & Midyette, J. (2008). Keyboarding compared with handwriting on a high-stakes writing assessment: Student choice of composing medium, raters' perceptions, and text quality. *Assessing Writing, 13*(1), 4-25.
- Widdowson, H. (1979). *Explorations in applied linguistics*. Oxford University Press.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. Jossey-Bass.
- Williamson, M. M., & Pence, P. (1989). Word processing and student writers. In B. K. Britton (Ed.), *Computer writing environments: Theory, research, and design* (pp. 93-127). Lawrence Erlbaum Associates.
- Wu, W. M., & Stansfield, C. W. (2001). Towards authenticity of task in test development. *Language Testing, 18*(2), 187-206.
- Xie, Q. (2011). Is test taker perception of assessment related to construct validity? *International Journal of Testing, 11*(4), 324-348.



- Zhi, M. (2017, March). *Washback revisited: Test-taker perceptions and individual differences in test preparation*. Poster presented at the American Association for Applied Linguistics Conference, Portland, OR.
- Zou, X. L., & Chen, Y. M. (2016). Effects of test media on different EFL test-takers in writing scores and in the cognitive writing process. *Technology, Pedagogy and Education*, 25(1), 79-99.