Annotation for literature related to online testing

Alexander, M. W., Bartlett, J. E., & Truell, A. D., & Ouwenga, K. (2001). Testing in a computer technology course: An investigating of equivalency in performance between online and paper and pencil methods. *Journal of Career and Technical Education, 18,* 69-80.

Abstract:

This experiment sought to examine the equivalence of online and paper and pencil testing methods as related to student performance in a computer technology course. Test score and completion time were the dependent variables that were used to assess students' performance. The study utilized a quasi-experimental design. Test scores were not significantly different on the variables of pretest, age, class standing, ethnicity, and gender. The findings showed that test scores were equivalent in both groups; however, time to complete the test was significantly different between the groups. The online testing group completed the test in less time than the paper and pencil group. The exploration of class standing did reveal that freshmen were the only group that took significantly less time to complete the online test. The study supports the online test method did not effect score as result of age, class level, and gender.

Key points:

- This study suggests no difference in the administration mode (CBT vs PPT) for age, class level, and gender.

Ames, P. C. (2003). Gender and learning style interactions in students' computer attitudes. *J. Educational Computing Research, 28*(3).

Abstract:

University students' attitudes toward computers were assessed as a function of learning style. Analyses of responses provided by 232 students to a learning style assessment instrument and a computer attitude survey revealed that specific learning styles were associated with an affinity for (liking of), confidence in, and anxiety about the use of computers. Within those learning styles, gender differences were discovered when students manifested a clearly dominant style. The findings indicate that computer-based or computer-assisted instruction may not be appropriate for all students and that curriculum modifications to account for learning style differences may increase the effectiveness of and reduce the aversion to computers in the classroom. Additional research into the relationship between learning styles and computer attitudes may also provide assistance relative to increasing the enrollment of females in technology-oriented courses of study.

Key point(s):

- Learning style: Gregorc defined four distinct learning styles – Abstract Sequential (AS), Abstract Random (AR), Concrete Sequential (CS), and Concrete Random (CR). He proposed that all learners had the ability to acquire and process data using all of these qualities but that each person had a preferred or natural style that is used most often.

Bennett, R.E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment, 6*(9).

Abstract:

This article describes selected results from the 2001 Math Online (MOL) study, one of three field investigations sponsored by the National Center for Education Statistics (NCES) to explore the use of new technology in NAEP. Of particular interest in the MOL study was the comparability of scores from paper- and computer-based tests. A nationally representative sample of eighth-grade students was administered a computer-based mathematics test and a test of computer facility, among other measures. In addition, a randomly parallel group of students was administered a paper-based test containing the same math items as the computer-based test. Results showed that the computer-based mathematics test was significantly harder statistically than the paper-based test. In addition, computer facility predicted online mathematics test performance after controlling for performance on a paper-based mathematics test, suggesting that degree of familiarity with computers may matter when taking a computer-based mathematics test in NAEP.

Key points:

- Computer facility predicted online mathematics test performance after controlling for performance on a paper-based mathematics test, suggesting that degree of familiarity with computers may matter when taking a computer-based mathematics test in NAEP

Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L.(2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly, 23*, 271-282.

Abstract:

The purpose of this study was to compare the effects of computer-based test accommodations to a noncomputer-based test accommodation and to no accommodation on mathematics performance assessment (PA) scores for secondary students with learning disabilities. Over four weeks, each student (N = 81) was tested on four parallel PAs, each time under a different condition: (a) standard administration (SA), (b) teacher-read (TR), (c) computer-read (CR), and (d) CR with video (CRV). PA alternate forms and order of conditions were counterbalanced across classrooms. After each PA administration, a student questionnaire assessed perceptions of the benefits of the testing condition. Results indicated that providing a reader, either human or computer, increases scores, but no significant difference was seen among TR, CR, and CRV. Implications are discussed with respect to reading accommodations, computerized testing, and the effects of reading accommodations on students with differing reading achievement levels.

Key points:

- The results revealed that computer-based test accommodations were as much effective as human based accommodations for students with learning disabilities. This is one of

evidences that computer-based test accommodation is suited to student with learning disabilities.

Cassady, J. C. & Gridley, B. E. (2005). The effects of online formative and summative assessment on test anxiety and performance. ***Journal of Technology, Learning, and Assessment, 4*****(1).**

Abstract:

This study analyzed the effects of online formative and summative assessment materials on undergraduates' experiences with attention to learners' testing behaviors (e.g., performance, study habits) and beliefs (e.g., test anxiety, perceived test threat). The results revealed no detriment to students' perceptions of tests or performances on tests when comparing online to paper-pencil summative assessments. In fact, students taking tests online reported lower levels of perceived test threat. Regarding formative assessment, findings indicate a small benefit for using online practice tests prior to graded course exams. This effect appears to be in part due to the reduction of the deleterious effects of negative test perceptions afforded in conditions where practice tests were available. The results support the integration of online practice tests to help students prepare for course exams and also reveal that secure web-based testing can aid undergraduate instruction through improved student confidence and increased instructional time.

Key points:

- Students’ testing behaviors (e.g., performance, study habits) and beliefs (e.g., test anxiety, perceived test threat) were studied under both the online and paper-pencil assessment mode. The results revealed no detriment to students' perceptions of tests or performances on tests when comparing online to paper-pencil summative assessments. It is interesting to find that students taking tests online reported lower levels of perceived test threat.

Clariana, R. and Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. ***British Journal of Educational Technology, 33*****(5).**

Abstract:

This investigation seeks confirm several key factors in computer-based versus paper-based assessment. Based on earlier research, the factors considered here include content familiarity, computer familiarity, competitiveness, and gender. Following classroom instruction, freshman business undergraduates (N=105) were randomly assigned to either a computer-based or identical paper-based test. ANOVA of test data showed that the computer-based test group outperformed the paper-based test group. Gender, competitiveness, and computer familiarity were not related to this performance difference, though content familiarity was. Higher-attaining students benefited most from computer-based assessment relative to higher-attaining students under paper-based testing. With the current increase in computer-based assessment, instructors and institutions must be aware of and plan for possible test mode effects.

Key Points:

- ANOVA of test data showed that the computer-based test group outperformed the paper-based test group
- Gender, competitiveness, and computer familiarity were not related to this performance difference, though content familiarity was.
- Higher-attaining students benefited most from computer-based assessment relative to higher-attaining students under paper-based testing.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. ***Journal of Technology, Learning, and Assessment, 3*****.**

Abstract:

Standards-based reform efforts are highly dependent on accurate assessment of all students, including those with disabilities. The accuracy of current large-scale assessments is undermined by construct-irrelevant factors including access barriers, a particular problem for students with disabilities. Testing accommodations such as the read-aloud have led to improvement, but research findings suggest the need for a more flexible, individualized approach to accommodations. The current pilot study applies principles of Universal Design for Learning to the creation of a prototype computer-based test delivery tool that provides students with a flexible, customizable testing environment with the option for read-aloud of test content. Two contrasting methods were used to deliver two equivalent forms of a National Assessment of Educational Progress United States history and civics test to ten high school students with learning disabilities. In a counterbalanced design, students were administered one form via traditional paper-and-pencil (PPT) and the other via a computer-based system with optional text-to-speech (CBT-TTS). Test scores were calculated, and student surveys, structured interviews, field observations, and usage tracking were conducted to derive information about student preferences and patterns of use. Results indicate a significant increase in scores on the CBT-TTS versus PPT administration for questions with reading passages greater than 100 words in length. Qualitative findings also support the effectiveness of CBT-TTS, which students generally preferred over PPT. The results of this pilot study provide preliminary support for the potential benefits and usability of digital technologies in creating universally designed assessments that more fairly and accurately test students with disabilities.

Key Points:

- It was found that a significant increase in scores on the CBT-TTS versus PPT administration for high school students with learning disabilities. This is one of evidences that computer-based test accommodation is more effective than PPT administration.

Gallagher, A., Bridgeman, B., & Cahalan, C. (2000). ***The effect of computer-based tests on racial/ethnic, gender, and language groups*** **(GRE Board Professional Report No. 96–21P). Princeton, NJ: Education Testing Service.**

Abstract:

This study examined data from several national testing programs to determine whether the change from paper-based administration to computer-based tests (CBTs) influences group differences in performance. Performance by **gender, racial/ethnic, and language groups** on the GRE General Test, GMAT, SAT, and TOEFL was analyzed to ensure that the change to CBTs does not pose a disadvantage to any of these subgroups, beyond that already identified for paper-based tests. Although all differences were quite small, some consistent patterns were found for some racial/ethnic and gender groups. African American examinees and, to a lesser degree, Hispanic examinees appear to benefit from the CBT format. However, for some tests, the CBT version negatively impacted female examinees. Analyses by gender within race/ethnicity revealed a similar pattern, though only for White females. Analyses for groups based on language showed no consistent patterns, but results indicate that the computer-based TOEFL has increased impact for some language groups—especially Chinese and Korean groups.

Key Points:

- Performance by **gender, racial/ethnic, and language groups** on the GRE General Test, GMAT, SAT, and TOEFL was analyzed to ensure that the change to CBTs does not pose a disadvantage to any of these subgroups, beyond that already identified for paper-based tests
- Although all differences were quite small, some consistent patterns were found for some racial/ethnic and gender groups
- African American examinees and, to a lesser degree, Hispanic examinees appear to benefit from the CBT format
- However, for some tests, the CBT version negatively impacted female examinees
- Analyses by gender within race/ethnicity revealed a similar pattern, though only for White females
- Analyses for groups based on language showed no consistent patterns, but results indicate that the computer-based TOEFL has increased impact for some language groups—especially Chinese and Korean groups.

Goldberg, A., & Pedulla, J.J. (2002). Performance differences according to test mode and computer familiarity on a practice GRE. *Educational and Psychological Measurement, 62*(6), 1053-1067.

Abstract:

Ideally, test performance is unrelated to the mode under which one administers the test. This study investigated relationships between test mode (paper-and-pencil vs. computerized-with-editorial-control, and computerized-without-editorial-control) and computer familiarity (lower, moderate, and higher) with test performance on the Graduate Record Exam (GRE). The GRE test was administered to 222 undergraduate students who were stratified by gender, then randomly assigned to a test mode group. With self-reported GPA as a covariate in a MANCOVA, the authors found that examinees in the paper-and-pencil group outperformed the computerized-without- editorial-control group on all subtests. The computerized-with-editorial-control group outperformed the computerized-without-editorial-control group on the Analytical subtest only. The authors also found a significant main effect for computer familiarity on the Analytical and

Quantitative subtests. A significant interaction between computer familiarity and test mode on the Quantitative subtest confounded the main effect for that subtest. The subtests were dramatically more speeded in the computerized forms. Results emphasize the importance of evaluating time constraints when converting exams from paper-and-pencil to computer-delivery.

Key Points:

- This study investigated relationships between test mode (paper-and-pencil vs. computerized-with-editorial-control, and computerized-without-editorial-control) and **computer familiarity** (lower, moderate, and higher) with test performance on the Graduate Record Exam (GRE).
- The authors found that examinees in the paper-and-pencil group outperformed the computerized-without- editorial-control group on all subtests.
- The computerized-with-editorial-control group outperformed the computerized-without-editorial-control group on the Analytical subtest only.
- The authors also found a significant main effect for computer familiarity on the Analytical and Quantitative subtests. A significant interaction between computer familiarity and test mode on the Quantitative subtest confounded the main effect for that subtest.

Johnson, G. M. (2007). Learning style under two web-based study conditions. *Educational Psychology, 27* (5).

Abstract:
A sample of 48 college students prepared for in-class examinations using two web-based study conditions. The A condition used web-based study groups and the B condition used web-based quizzes. The Index of Learning Styles positioned students on four dimensions of learning style (active-reflective, visual-verbal, sequential-global, and sensing-intuitive). Students who were more active than reflective expressed a preference for face-to-face study groups rather than online study groups and for online quizzes rather than pencil-and-paper quizzes. Students who were more visual than verbal expressed a preference for online quizzes rather than online study groups. Such preferences were validated by decreased achievement in the less-preferred study condition. At college level, students are aware of their learning style and understand the conditions that facilitate their mastery of course content. Instructional applications of web-based technology may provide mechanisms for more consistently accommodating student learning style in higher education.

Key point(s):

- This paper positioned students on four dimensions of learning style based on the Index of Learning Styles (ILS) developed by Felder and Silverman (1988): active-reflective, visual-verbal, sequential-global, and sensing-intuitive. [1. active (e.g., learns by doing and enjoys working with others) versus reflective (e.g., learns by thinking and prefers working alone); 2. visual (e.g., prefers to learn with pictures, diagrams, and charts) versus verbal (e.g., prefers written and spoken explanations); 3. sequential (e.g., linear thinking, learns in small steps) versus global (e.g., holistic thinking, learns in leaps); 4. sensing

(e.g., practical, concrete thinker, oriented toward facts) versus intuitive (e.g., innovative, abstract thinker, oriented toward theory)].

- Learning style refers to "characteristic cognitive, affective, and psychological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment" (Keefe, 1979, p. 4). "Each person's style is a combination of various biological and experiential variables that contribute to learning" (Rochford, 2003, p. 667). McKeachie (1995) proposed that "learning styles are preferences and habits of learning that have been learned" and that students "learn strategies that enable them to be effective when taught by methods that are not compatible with their preferred style" (p. 2).

Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing, 6*(1), 1-24.

Abstract:
In addition to the potential that computer-based testing (CBT) offers, empirical evidence has found that identical computerized and paper-and-pencil tests have not produced equivalent test-taker performance. Referred to as the "mode effect," previous literature has identified many factors that may be responsible for such differential performance. The aim of this review was to explore these factors, which typically fit into two categories, participant and technological issues, and highlight their potential impact on performance.

Key Points:

- In his review paper, Leeson (2006) categorized two potential factors associated with test performances in CBT and PPT from two perspectives: participant and technology. Particularly, the participant factors include cognitive processing, ability, computer familiarity, and computer anxiety as well as demographic characteristics of participant such as ethnicity and gender.

Parshall, C., & Kromrey, J. D. (1993*). Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect.* Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta, GA.

Key Points:

- In the study of CBT and PPT versions of Graduate Record Examination (GRE), Parshall and Kromrey (1993) found examinees' **gender**, **race,** and **age** were associated with the test mode although the results varied across the three subscales (verbal, quantitative and analytic) of the GRE. (Note: full paper is not found yet).

Pomplun, M., & Custer, M. (2005). The score comparability of computerized and paper-and-pencil formats for K-3 reading tests. *Journal of Educational Computing Research, 32*(2), 153-166.

Abstract:

This study investigated the equivalence of scores from computerized and paper-and-pencil formats of a series of K-3 reading screening tests. Concerns about score equivalence on the computerized formats were warranted because of the use of reading passages, computer unfamiliarity of primary school students, and teacher versus computer administration of the test. The results indicated that the computerized format produced lower scores than the paper-and-pencil format across all four grades. These difficulty differences could be related to the differences in family income but further research is needed into the causal factors. However, confirmatory factor analysis supported the construct and measurement equivalence of the scores from the two formats.

Key Points:

- Concerns about score equivalence on the computerized formats were warranted because of the use of reading passages, computer unfamiliarity of primary school students.
- These difficulty differences could be related to the differences in family income but further research is needed into the causal factors.

Shermis, M. D., & Lombard, D. (1998). Effects of computer-based test administrations on test anxiety and performance. *Computer in Human Behavior, 14,* 111-123.

Abstract:

This study examined the degree to which computer and test anxiety had a predictive role in performance across three computer-administered placement tests (math, reading, written English). Subjects (72 college undergraduates at a Midwestern university) were measured with the Computer Anxiety Rating Scale, the Test Anxiety Inventory, the Myers-Briggs Type indicator (MBT1), and the three achievement areas. Age and gender were added to the prediction model for completeness. Results showed that age and test anxiety were significant predictors for math performance, with lower values on the two variables associated with better performance. When reading was the outcome variable, age and computer anxiety were statistically significant performance predictors, with older readers faring better and less anxious individuals achieving higher scores. No predictors were statistically significant for the written English essay. The hypothesis that "thinkers" and "intuitives" on the MBTI would have lower anxiety scores was only partially confirmed. Thinkers had lower anxiety scores than "feelers," but there was no difference on the intuitive-sensor dimension. The results suggest that much of what is considered computer anxiety may in fact be a manifestation of test anxiety. It is possible that, by giving examinees more perceived control, anxiety levels can be reduced.

Key Points:

- The results revealed that computer anxiety was significant factor associated with performance in CBT only for math exam (not reading and written English exam).
- For math, age and test anxiety were significant predictors for math performance in computer based test.
- For reading, age and computer anxiety were statistically significant reading performance predictor in CBT. Older students and less anxious student achieved higher scores.

- For written English exam, neither age, test anxiety, nor computer anxiety were statistically significant for the written English essay.

Stowell, J. R., & Bennett, D. (2010). Effects of online testing on student exam performance and test anxiety. *Educational Computing Research, 42*(2), 161-171.

Abstract:

Increased use of course management software to administer course exams online for face-to-face classes raises the question of how well test anxiety and other emotions generalize from the classroom to an online setting. We hypothesized that administering regular course exams in an online format would reduce test anxiety experienced at the time of the exam and improve exam scores. We recruited 69 participants from a psychology course to take classroom- and online-delivered exams, using a counterbalanced crossover design. We found that students who normally experience high levels of test anxiety in the classroom had reduced test anxiety when taking online exams, while the reverse was true for those low in classroom anxiety. Furthermore, the relationship between test anxiety and exam performance was weaker in an online setting than in the classroom. We recommend that instructors evaluate the potential impact of these findings when considering offering examinations online.

Key Points:

- The results showed that high classroom test anxiety was associated with poor performance in the classroom exam ($r = –.57$, $p < .001$, $n = 66$) but it was less strongly associated with exam performance online ($r = –.28$, $p = .02$, $n = 65$). In other words, the relationship between test anxiety and exam performance was weaker in an online setting than in the classroom.

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*, 219–274.

Abstract:

The planned introduction of a computer-based Test of English as a Foreign Language (TOEFL) test raises concerns that language proficiency will be confounded with computer proficiency, introducing construct-irrelevant variance to the measurement of examinees' English-language abilities. We administered a questionnaire focusing on examinees' computer familiarity to 90,000 TOEFL test takers. A group of 1,200 “low-computer-familiar” and “high-computer-familiar” examinees from 12 international sites worked through a computer tutorial and a set of 60 computer-based TOEFL test items. We found no meaningful relationship between level of computer familiarity and level of performance on the computerized language tasks after controlling for English language ability. We concluded that no evidence exists of an adverse relationship between computer familiarity and computer-based TOEFL test performance due to lack of prior computer experience.

Key Points:

- No meaningful relationship between level of computer familiarity and level of performance on the computerized language tasks after controlling for English language ability.
- No evidence exists of an adverse relationship between computer familiarity and computer-based TOEFL test performance due to lack of prior computer experience.

Wang, K. H., Wang, T. H., Wang, & Huang, S. C. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. ***Journal of Computer Assisted Learning, 22*****.**

Abstract:
The purpose of this research was to investigate the effects of formative assessment and learning style on student achievement in a Web-based learning environment. A quasi-experimental research design was used. Participants were 455 seventh grade students from 12 classes of six junior high schools. A Web-based course, named BioCAL, combining three different formative assessment strategies was developed. The formative assessment strategies included Formative Assessment Module of the Web-Based Assessment and Test Analysis system (FAM-WATA) (with six Web-based formative assessment strategies), Normal Module of Web-Based Assessment and Test Analysis system (N-WATA) (only with partial Web-based formative assessment strategy) and Paper and Pencil Test (PPT) (without Web-based formative assessment strategy). Subjects were tested using Kolb's Learning Style Inventory, and assigned randomly by class into three groups. Each group took Web-based courses using one of the formative assessment strategies. Pre- and post-achievement testing was carried out. A one-way ANCOVA analysis showed that both learning style and formative assessment strategy are significant factors affecting student achievement in a Web-based learning environment. However, there is no interaction between these two factors. A post hoc comparison showed that performances of the FAM-WATA group are higher than the N-WATA and PPT groups. Learners with a 'Diverger' learning style performed best followed by, 'Assimilator', 'Accommodator', and 'Converger', respectively. Finally, FAMWATA group students are satisfied with six strategies of the FAM-WATA.

Key point(s):

- Learning styles (review of literature): Learning style is one of the most important factors that affect personal academic competence (Kolb, 1984). Scholars define learning style differently. There is currently no widely accepted definition of what learning style is. Kolb (1976) divided the learning process cycle into four learning modes in term of information perception and processing by learners: concrete experience (CE), reflective observation (RO), abstract conceptualization (AC), and active experimentation (AE).