December 2003

Test Performance and the Medium: Unearthing Differences That Make a Difference

Jennifer Nicholson
Washington State University

Andrew Hardin
Washington State University

Darren Nicholson
Washington State University

TEST PERFORMANCE AND THE MEDIUM: UNEARTHING DIFFERENCES THAT MAKE A DIFFERENCE

Jennifer Nicholson[1]
Washington State University
jnicholson@wsu.edu

Andrew Hardin
Washington State University
ahardin@wsu.edu

Darren Nicholson
Washington State University
dbnicholson@wsu.edu

Abstract

This study will attempt to provide an explanation for the disparity that exists between the literature that purports no difference between individuals' performance when taking paper-and-pencil versus computer-based tests, with the literature that purports that high levels of computer anxiety will influence computer-based performance. The question is asked, how can the findings that show high levels of computer anxiety negatively influence performance co-exist with findings that show that there is no difference between an individual's performance when taking a paper-and-pencil versus a computer-based exam? This research posits that although there may be no difference in test performance between pencil-and-paper and computer-based tests for simple questions, there will be a test performance difference for questions that are complex. Thus, we also ask what is the emergent phenomenon that is occurring that causes computer anxiety to reduce performance on complex questions administered via the computer? To explain this phenomenon, we will draw upon the cognitive psychology literature, specifically the literature regarding cognitive capacity, to show that some of an individual's cognitive capacity will be utilized by emotional arousal (e.g. computer anxiety), thereby increasing the amount of cognitive load on an individual's working memory. The increase in cognitive load on working memory, namely computer anxiety, should not influence an individual's computer-based test performance for simple questions, however, it is posited that it will influence an individual's computer-based test performance for questions that are complex.

Introduction

With the pervasiveness of computers has come the increased use of computerized testing in both educational and organizational settings. Although paper-and-pencil testing will never become extinct, computerized testing is becoming the norm in both Internet-based distance education courses and traditional courses supplemented with web-based materials and exercises (Klass and Crothers, 2000). With this inevitable shift in test administration mode, it has thus become critical to examine the differences, if any, that computerized testing may have on an individual's test-taking performance. It would be foolish to operate under the assumption that as long as identical tests are given, it does not matter what administration mode is used.

It is important at the outset to note the different terms used to describe tests that are administered via computer. Computerized testing is the most general term and encompasses any type of testing using a computer. Computer-based testing involves using a computer to administer the same test as that given in a paper-and-pencil format; whereas computer-adaptive testing administers test questions based on the examinee's answer to the previous question or questions. This study is primarily concerned with

[1]All authors contributed equally.

computer-based testing, not with computer-adaptive testing. Some commonly cited and well-established advantages to computer-based testing include reduced testing time (Bunderson, Inouye, and Olsen, 1989; English, Reckase, and Patience, 1977; Green, 1988; Olsen, Maynes, Slawson, and Ho, 1986; Wise and Plake, 1989), increased test security (Grist, Rudner, and Wise, 1989), the ability to provide instant scoring (Bugbee, 1992; Bugbee and Bernt, 1990; Kyllonen, 1991; Mazzeo and Harvey, 1988), and the ability to schedule computer-based tests more easily than paper-and-pencil tests (Bugbee, 1992; Hambleton, Zaal, and Pieters, 1991; Wise and Plake, 1989). Despite the well-known advantages associated with computer-based testing, the question still remains as to whether computer-based testing is equivalent to traditional paper-and-pencil administration (i.e. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985). However, this question has been difficult to answer.

Although several studies have found no difference in performance between paper-and-pencil and computer-based tests (Baird and Silvern, 1992; Olsen, Maynes, Slawson, and Ho, 1986; Sykes et al, 1991; Ward, Hooper, and Hannafin, 1989; Wilson, Genco, and Yager, 1985; Wise and Plake, 1989), others studies have found a difference (Bunderson, Inouye, and Olsen, 1989; Mazzeo and Harvey, 1988). Even studies that have gone so far as to compare the performance and anxiety levels of examinees across the two administration modes have resulted in mixed findings (Chin, Donn, and Conroy, 1991; Johnson and Johnson, 1981; Llabre et al, 1987). Hence it may be necessary to try and uncover other underlying factors that may affect an individual's performance when taking a test in these two different administration modes.

The evaluation process in education has been criticized due to the high levels of anxiety that may arise from the importance placed on academic examinations (Shermis and Lombard, 1998). The relationship between anxiety and computer-based testing was investigated by Hedl, O'Neil, and Hansen (1973), who found that college students taking a computer-based intelligence test had higher levels of test anxiety, before and after the test, than students given the same test via paper-and-pencil. Furthermore, Rocklin and Thompson (1985) found that when students were given a fairly complex paper-and-pencil test, students with lower test anxiety had a tendency to perform better than more anxious examinees. However, when given a fairly simple paper-and-pencil test, examinees with moderate levels of anxiety achieved higher scores than examinees with either low or high levels of anxiety. This anxiety, which may make it difficult for students to demonstrate the skill or mastery of achievement, may further be compounded as computer technology is used as a substitute for traditional paper-and-pencil assessments (Shermis and Lombard, 1998). In fact, the self-efficacy literature, which has only recently looked at the concept of task complexity as it pertains to computer training (Bolt, Killough, and Koh, 2001), has shown that computer self-efficacy leads to computer anxiety, which in turn affects performance. Based on the self-efficacy literature, we thus posit that there will be a difference between performance on computer-based tests and performance on paper-and-pencil tests due to systematic individual differences of computer anxiety. However, we further propose that these differences will occur only under conditions of increased task[2] complexity. To explain this phenomenon we draw upon the cognitive psychology literature regarding working memory, cognitive load, and performance. Thus, the purpose of this paper is to investigate the effects of computer anxiety and task complexity on computer-based test performance in hopes of unearthing the phenomenon contributing to the inconsistent findings surrounding paper-and-pencil and computer-based exam performances.

The rest of the paper is organized as follows: first, the literature on computer self-efficacy, computer anxiety, and performance is briefly discussed; second, the literature on working memory, cognitive load, and their relationship to performance is explained; and, finally , the methodology that will be utilized in this study is developed.

Computer Self-Efficacy, Computer Anxiety, and Performance

Self-efficacy, a crucial variable within the Social Cognitive Theory framework as proposed by Bandura (1982, 1986, 1997), has been used in an assortment of areas by researchers attempting to better understand the influence of personal efficacy beliefs on human agency. Included among the areas of research in which self-efficacy has been studied is information systems. Seminal works within the information systems literature include a study by Compeau and Higgins (1995), who developed a computer self-efficacy construct based on the earlier work of Harrison and Rainer (1992), and a study by Marakas, Yi, and Johnson (1998) who during an extensive literature review of self-efficacy, posited various antecedents and mediators in the computer self-efficacy – performance relationship. Many definitions have been offered for the computer self-efficacy construct, however for the purpose of this study, computer self-efficacy (CSE) will be referred to as defined by Compeau and Higgins (1995):

---

[2]The word task is used to refer to the administration of a set of simple and complex test questions.

Computer self-efficacy, then, refers to a judgment of one's capability to use a computer. It is not concerned with what one has done in the past, but rather with judgments of what could be done in the future.

Among the many mediators posited by Marakas et al (1998), was computer anxiety. Specific studies involving computer anxiety (CA) include Compeau and Higgins (1995), a study in which CA was found to be a moderator between CSE and computer usage, and Johnson and Marakus (2000), a study in which CA was found to be both an antecedent to CSE and a mediator in the CSE – performance relationship. The following definition by Harrison and Rainer (1992) will be used for CA in the current study:

Computer anxiety may be defined as a response to interaction or anticipation of interaction with automated data or information processing systems.

Both CSE (Johnson and Marakas 2000) and CA (Shermis and Lombard 1998) have been shown to influence performance. Johnson and Marakas (2000) found, during a study designed to further establish the causal ordering of variables within the CSE – performance relationship, that CSE perceptions were significantly related to spreadsheet task performance. Shermis and Lombard (1998) in a study designed to determine the influence of CA during educational testing, found that low levels of CA were positively related to both math and reading test performance.

Figure 1 represents the previously established relationships among computer self-efficacy, computer anxiety, and performance.

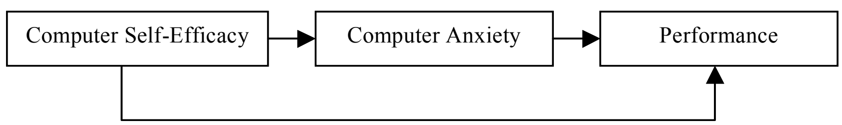

Figure 1.

Some authors have identified specific variables that moderate the CSE – performance relationship. Bolt, Killough and Koh (2001) for example, identified task complexity as a moderator between CSE and performance, and more precisely, found that computer self-efficacy was positively related to performance when task complexity was high. Although the Bolt et al (2001) study used a linear programming task rather than computer based test performance as the dependent variable, the results of the study clearly establish task complexity as an important variable for consideration by researchers interested in studying the CSE – performance relationship. These relationships are shown in Figure 2

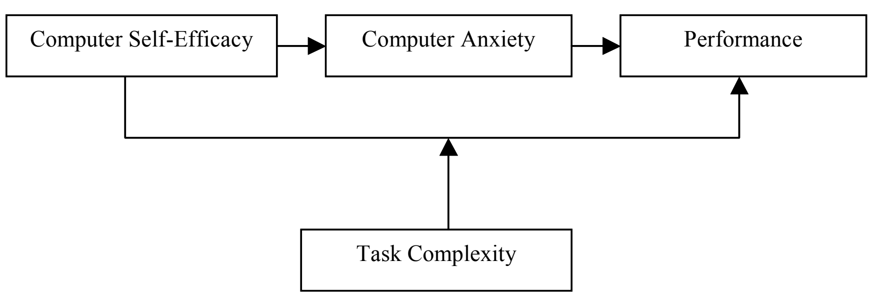

Figure 2.

The relationships identified above are supported by previous empirical research. However, no studies exploring the potential role of task complexity on the relationship between computer anxiety and performance was found. We propose that task complexity will also moderate the relationship between computer anxiety and performance. In support of this relationship, the following section explores the connection between anxiety, task complexity, and performance in greater detail

Working Memory, Cognitive Load, and Performance

Working memory, as defined by cognitive psychologists, refers to "a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning and reasoning" (Baddeley, 1986). Working memory, commonly referred to as short-term memory, cognitive capacity, blackboard of the mind, and mental scratch pad, is characterized by its limited storage capacity and quick turnover and is set apart from the larger capacity and archival memory system traditionally referred to as long-term memory (Baddeley, 1986; Goldman-Rakic, 1987; Reddy, 1980). Working memory is intimately related to where and how we direct our attention to think about things, or to process information. The biggest limitation of working memory is its capacity to deal with no more than about seven elements (plus or minus two) of information simultaneously (Miller, 1956). With this limitation in mind, Baddeley (1986) discusses the concept of d*isplacement/interference* in working memory. Displacement/interference refers to the repercussions involved when additional new items enter an individual's working memory – existing items tend to become harder to access, can become displaced by new information, and can result in decreased cognitive efficiency.

The displacement or interference of additional new items can be conceptualized as placing an additional cognitive load on working memory. Cognitive load refers to the total amount of mental activity imposed on working memory at an instance in time (Sweller, 1988, 1994). The major factor that contributes to cognitive load is the number of elements or chunks (Miller, 1956) that need to be attended to. Kahneman (1973) found that harder or more difficult tasks, as opposed to easier tasks, place greater demands on an individual's mental or cognitive capacity. Moreno and Bodenhausen (1999) found that the effortfulness of information processing tasks clearly determines whether or not the imposition of a cognitive load is likely to create disruptions in cognitive capacity. Easterbrook (1959) found that reduced performance occurred when individuals performed complex tasks in combination with high arousal states such as high emotionality or anxiety. Eysenk (pg. 341, 1984) identified that high arousal affects performance by reducing an individual's attentional control, accuracy, short-term memory (working memory), and retrieval efficiency. Deffenbacher (1978) found that non-anxious people were able to attend to the task far better than anxious individuals. The findings above suggest that anxiety acts as a cognitive load, or adds to the existing cognitive load, on working memory, resulting in reduced performance on tasks requiring more of a person's cognitive capacity (Eysenk, 1984; Kahneman, 1979).

In this paper, we contend that differences exist between paper-and-pencil and computer-based testing performances but that these differences will only manifest themselves in conditions when an individual has high computer anxiety and task complexity is high. As discussed above, simple tasks demand very little of one's working memory (Kahneman, 1979) leaving surplus "space" for maintaining awareness of arousal states like computer anxiety. However, as tasks become increasingly complex, there is less space available in working memory for maintaining awareness of arousal states like computer anxiety, and displacement (Baddeley, 1986) of needed information (chunks) that is pertinent to task performance may occur, thereby reducing test performance (Baddeley, 1986). The affect of task complexity on the relationship between computer anxiety and performance is identified in Figure 3 below.

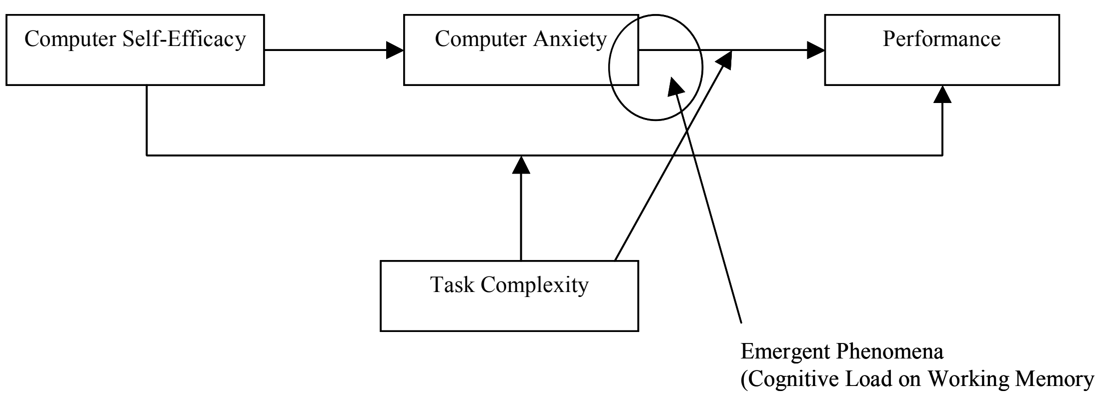

Figure 3.

With this in mind we posit the following hypotheses:

H1 There will be no significant difference on performance between complex paper-and-pencil tests and complex computer-based tests for individuals with low computer anxiety.

H2 There will be a significant difference on performance between complex paper-and-pencil tests and complex computer-based tests for individuals with high computer anxiety.

H3 There will be no significant difference on performance between simple paper-and-pencil tests and simple computer-based tests for individuals with either low or high computer anxiety.

Methodology

Design

A 2 x 2 x 2 between-subjects Factorial design (Figure 4) involving three independent variables, namely, medium (paper-and-pencil and computer-based tests), computer anxiety (high and low), and task type (simple and complex) will be employed.

		Paper-and-Pencil		**Computer-Based**	
		Simple	Complex	Simple	Complex
CA	Low				
	High				

Figure 4.

Sample

A sample of 100 students enrolled in an introductory management information systems course[3] will be used for the pilot test. Following pilot testing, 250 students enrolled the following semester in the same introductory management information systems course will then be utilized to collect the primary data for the study. The introductory management information systems course was chosen, as it consists of students who are majoring in all areas of business and who have varying levels of computer experience.

Procedure

A pilot study will first be conducted in the 2003 summer session on introductory management information students utilizing the procedure discussed below. The objective of the pilot study will be to ensure the validity of the instruments employed (computer self-efficacy, computer anxiety, and perceived effort), confirm the previously established relationships (Figures 1, 2, and 3), and to identify any potential problems with the procedures used in the study. Following pilot testing and analysis, the primary data will then be collected during the following semester from students enrolled in the same introductory management information systems course.

Subjects will be given a pre-test survey to measure computer anxiety and computer self-efficacy. Computer self-efficacy will be measured using items adapted from an existing scale developed by Murphy, Coover, and Owen (1989) while computer anxiety will be measured using the same four items that Compeau, Higgins, and Huff (1999) adapted from Heinssen et al.'s (1987) Computer Anxiety Rating Scale. Dyads will be formed by conducting an inter-quartile split (top and bottom 25 percent) – based on an individual's score on the computer anxiety scale. Subjects from each dyad will then be randomly assigned to one of four treatment groups – simple paper, complex paper, simple computer and complex computer. For the simple task, subjects will be given five simple questions to answer; whereas for the complex task, subjects will be given five complex questions to answer.[4]

[3]The introductory management information systems course is a mandatory course for all undergraduate majors in the school of business.

[4]A complex task is defined as one that requires an individual to deal with five or more pieces of information simultaneously in order to identify a correct solution, whereas, a simple task is defined as one that requires an individual to deal with two or fewer pieces of information simultaneously. Simple and complex questions were drawn from a bank of general IQ questions. Initial pilot testing during the Fall 2002 semester was performed to discern subjects' (sophomore information systems students) perceptions of the question – simple or complex – and

These questions will be administered to subjects in one of two mediums, paper-and-pencil or computer-based. The dependent variable, performance, will be operationalized as the ratio of correct answers to the total number of answers on the exam instrument (100 would represent the highest performance, whereas, 0 would indicate the lowest performance). The success of the simple/complex task manipulation will be measured after each task has been completed using an instrument developed and psychometrically tested by the authors to measure an individual's perception of the effort that it took to answer the questions they were given.

Conclusion

Previous research has resulted in mixed findings as to the affects of computer anxiety on test performance. The current study will contribute to the existing literature by identifying the conditions under which computer anxiety will affect an individual's performance on a computer-based test. We expect to see a negative effect on performance for subjects exhibiting high levels of computer anxiety when answering complex, computer-based questions.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Baddeley, A. (1986). *Working Memory.* London: Oxford University Press.

Baird, W.E., and Silvern, S.B. (1992). "Computer Learning and Appropriate Testing: A First Step in Validity Assessment," *Journal of Research on Computing in Education*, 25(1), 18-27.

Bandura, A. (1982), "Self-Efficacy Mechanisms in Human Agency," *American Psychologist,* 37:2, 122-147.

Bandura, A. (1986), *Social Foundation of Thought and Action: A Social Cognitive Theory,* Prentice Hall: Englewood Cliffs.

Bandura, A. (1997), *Self-Efficacy: The Exercise of Control,* W.H. Freeman and Company, New York.

Bolt, M.A., Killough, L.N., and Koh, H.C. (2001), "Testing the Interaction Effects of Task Complexity in Computer Training Using the Social Cognitive Model," *Decision Sciences,* 32:1, 1-20.

Bugbee, A.C. (1992). *Examination on demand: Findings in ten years of testing by computer 1982-1991*. Edina, MN: TRO Learning.

Bugbee, A.C., and Bernt, F.M. (1990). "Testing by Computer: Findings in Six Years of Use 1982-1988," *Journal of Research on Computing in Education*, 23(1), 87-100.

Bunderson, C.V., Inouye, D.K., and Olsen, J.B. (1989). The four generations of computerized educational measurement. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.), (pp. 367-407). New York: American Council on Education – Macmillan.

Compeau, D.R. and Higgins, C.A. (1995), "Computer Self-Efficacy: Development of a Measure and Initial Test," *MIS Quarterly,* 19:2, 189-211.

Compeau, D., Higgins, C.A., and Huff, S. (1999), "Social Cognitive Theory and Individual Reactions to Computing Technology: A Longitudinal Study," *MIS Quarterly*, 23:2, 145-158.

Deffenbacher, J. L. (1978). "Worry, Emotionality, and Task-Generated Inferences in Test Anxiety: An Empirical Test of Attention Theory," *Journal of Educational Psychology,* 70, 248-254.

Easterbrook, J. A. (1959). "The Effect of Emotion on Cue Utilisation and the Organisation of Behaviour ," *Psychological Bulletin,* 66, 183-201.

English, R.A., Reckase, M.D., and Patience, W.M. (1977). "Application of Tailored Testing to Achievement Measurement," *Behavior Research Methods and Instrumentation*, 9, 158-162.

Eysenck, M. W. (1984). *A Handbook of Cognition Psychology*. London: Lawrence Erlbaum Associates, Ltd., Publishers.

Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In F. Plum and V. Mountcastle, Eds., *Handbook of Physiology.* Bethesda, MD: American Physiological Society, pp. 373-417.

Green, B.F. (1988). Construct validity of computer-based tests. In H. Wainer and H.I. Braun (Eds.), *Test Validity* (pp. 77-86). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grist, S., Rudner, L., and Wise, L. (1989). Computer adaptive tests. *ERIC Clearinghouse on Tests, Measurement, and Evaluation* (Vol. 107). Washington DC: American Institute for Research. (ERIC ED 315 425)

---

to further develop the 'perceived effort' instrument used in this study.

Hambleton, R.K., Zaal, J.N., and Pieters, J.P.M. (1991). Computerized adaptive tests: Theory, applications, and standards. In R.K. Hambleton and J.N. Zaal (Eds.), *Advances in educational and psychological testing* (pp. 341-366). Boston: Kluwer.

Harrison, A.W. and Rainer, R.K. (1992), "An Examination of the Factor Structures and Concurrent Validities for the Computer Attitudes Scale, the Computer Anxiety Rating Scale, and the Computer Self-Efficacy Scale," *Educational and Psychological Measurement,* 52, 735-746.

Hedl, J.J., O'Neil, H.F., and Hansen, D.N. (1973). "Effective Reactions Toward Computer Based Intelligence Testing," *Journal of Consulting and Clinical Psychology*, 40, 217-222.

Johnson, R.D. and Marakas, G.M. (2000), "Research Report: The Role of Behavioral Modeling in Skill Acquisition: Toward Refinement of the Model," *Information Systems Research,* 11:4, 402-417.

Kahneman, D. (1973) *Attention and Effort.* Englewood Cliffs: Prentice Hall.

Klass, G. and Crothers, L. (2000). "An Experimental Evaluation of Web-Based Tutorial Quizzes," *Social Science Computer Review*, 18 (4), 508-515.

Kyllonen, P.C. (1991). "Principles for Creating a Computerized Test Battery," *Intelligence*, 15(1), 1-15.

Marakas, G.M., Yi, Y.M. and Johnson, R.D. (1998), "The Multilevel and Multifaceted Character of Computer Self-Efficacy: Toward Clarification of the Construct and an Integrative Framework for Research," *Information Systems Research,* 9:2, 126-163.

Mazzeo, J., and Harvey, A.L. (1988). *The equivalence of scores from automated and conventional educational and psychological tests* (College Board Report No. 88-8). New York: College Entrance Examination Board.

Miller, G.A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review, 63*, 81-97.

Moreno, K. N. and Bodenhausen, G. V. (1999) "Resisting Stereotype Change: The Role of Motivation and Attentional Capacity in Defending Social Beliefs," *Group Processes and Intergroup Relations*, 2(1), 5-16.

Murphy, C.A., Coover, D., and Owen, S.V. (1989). "Development and Validation of the Computer Self-Efficacy Scale," *Educational and Psychological Measurement*, 49, 893-899.

Olsen, J.B., Maynes, D.D., Slawson, D., and Ho, K. (1986). *Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Reddy, E. R. (1980). Machine models of speech perception. In R. A. Cole, Ed., *Perception and Production of Human Speech.* Hillsdale, NJ: Erlbaum pp. 215-242.

Rocklin, T., and Thompson, J.M. (1985). "Interactive Effects of Text Anxiety, Test Difficulty, and Feedback," *Journal of Educational Psychology*, 77, 368-372.

Shermis, M.D., and Lombard, D. (1998). "Effects of Computer-Based Test Administrations on Test Anxiety and Performance," *Computers in Human Behavior*, 14(1), 111-123.

Sykes, R.C., Heydeman, H., Whitaker, M., Lonsberg, D., Kelly, R., and Haynes, B. (1991, April). *Assessing the effects of computer administration on scores and parameter estimates using IRT models*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Sweller, J. (1988). "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science,* 12, 257-285.

Sweller, J. (1994). "Cognitive Load Theory, Learning Difficulty and Instructional Design," *Learning and Instruction,* 4, 295-312.

Ward, T.J., Jr., Hooper, S.R., and Hannafin, K.M. (1989). "The Effect of Computerized Tests on the Performance and Attitudes of College Students," *Journal of Educational Computing Research*, 5(3), 327-333.

Wilson, F.R., Genco, K.T., and Yager, G.G. (1985). "Assessing the Equivalence of Paper-and-Pencil vs. Computerized Tests: Demonstration of a Promising Methodology," *Computers in Human Behavior*, 1, 265-275.

Wise, S.L., and Plake, B.S. (1989). Research on the effects of administering tests via computers. *Educational Measurement: Issues and Practice*, 8(3), 5-10.