

Article

"Student Choice between Computer and Traditional Paper-and-Pencil University Tests: What Predicts Preference and Performance?"

Karen Lightstone et Steven M. Smith

Revue internationale des technologies en pédagogie universitaire / International Journal of Technologies in Higher Education, vol. 6, n° 1, 2009, p. 30-45.

Pour citer cet article, utiliser l'information suivante :

URI: <http://id.erudit.org/iderudit/039179ar>

DOI: 10.7202/039179ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : info@erudit.org

Student Choice between Computer and Traditional Paper-and-Pencil University Tests: What Predicts Preference and Performance?

Karen **Lightstone**
karen.lightstone@smu.ca

Steven M. **Smith**
steven.smith@smu.ca

Saint Mary's University
Halifax, Nova Scotia

Recherche scientifique avec données empiriques

Abstract

Two studies were conducted to assess meta-cognitive and individual difference influences on students' choice of writing tests in paper-and-pencil or computer-administered format. In Study 1, university students chose the test format for an accounting exam (paper-and-pencil or computer). In Study 2, students disclosed their reasons for their choice of test format, predicted their scores on the first test and provided confidence ratings for their predictions. The results of both studies show that the reasons for choosing a computer vs. a paper-and-pencil test format differ, and that both choice and performance can be explained to some extent by individual difference and meta-cognitive factors.

Key words

computer-based testing; paper-and-pencil testing; meta-cognitive influences; test format

Résumé

Deux enquêtes ont été menées afin de mesurer l'apport des différences métacognitives et des différences individuelles dans une tâche où un étudiant doit choisir de compléter un examen en format papier ou administré par ordinateur. Dans la première enquête, des étudiants universitaires ont choisi le format d'un examen de comptabilité (format papier ou administré par ordinateur). Dans la deuxième enquête, on a demandé aux étudiants d'expliquer leur choix, de prédire les résultats de leurs examens, et de fournir leur niveau de confiance quant à l'exactitude de leurs prédictions. Les résultats des deux enquêtes ont montré que le choix d'un examen en format papier est associé à un raisonnement différent que celui associé au choix d'un examen administré par ordinateur. De plus, les résultats ont montré que le choix du format de l'examen et la performance aux examens peuvent être, en partie, expliqués par des facteurs individuels et métacognitifs.

Student Choice between Computer and Traditional Paper-and-pencil University Tests: What Predicts Preference and Performance?

The migration of many graduate school entrance examinations from paper-and-pencil to computer format has resulted in a flurry of other computer-administered tests. The Graduate Record Examination (GRE) is probably the most well known graduate school exam to offer a computer-based format, beginning in 1992. Since then the Graduate Management Admission Test (GMAT) and Medical College Admission Test (MCAT) have become computer-based, although the Scholastic Aptitude Test (SAT) and Law School Admission Test (LSAT) are still written with paper and pencil. The trend toward computer-based testing began in 1982 when The American College in the U.S. began offering computerized tests to its students. The American College is a distance-education institution that provides financial services training. Before it began offering computerized tests, students could write the tests only twice a year and had to wait six weeks for their results (Bugbee & Bernt, 1990).

The ease with which computers and the Internet can provide testing, combined with the inherent flexibility in terms of item type and scoring, have made computer-assisted testing a popular choice by educators and others (Anderson, 2003; Zenisky & Sireci, 2002). For example, computer-administered testing is widely used to assess second language learning (e.g., Chalhoub-Deville & Deville, 1999; Chapelle, 2001; see Wainer et al., 2000, for a review) and to select personnel in the military and corporate worlds (Anderson, 2003). This popularity is not surprising, as the benefits of using computers for testing are well documented: flexible assessment times and locations (Bugbee, 1996; Bugbee & Bernt, 1990; Peterson & Reider, 2002; Wood, 1984), immediate feedback to candidates (Bugbee, 1996; Bugbee & Bernt, 1990; Peterson & Reider, 2002; Rabinowitz & Brandt, 2001; Wise & Plake, 1990; Wood, 1984), shorter time to administer to large numbers of test takers (Rabinowitz & Brandt,

2001; Wise & Plake, 1990), the ability to collect additional information about respondents (Wise & Plake, 1990; Wood, 1984) and more recently, low administration cost (Peterson & Reider, 2002).

Despite the popularity of computer-based testing, relatively little research has been done on the differences between computer and paper-and-pencil testing. Moreover, the research that has been conducted has generally focused on one of two approaches to testing procedures: a computer-*based* approach, or a direct copy of paper-and-pencil tests but delivered via computer (which is the focus of this paper); and a computer-*adapted* approach, which changes the nature of the questions according to the test taker's response (e.g., becoming more difficult with each correct response).

Interestingly, little research has focused on computer-*based* testing, perhaps because most people believe that if the items are identical, then the testing mode is irrelevant. Contrary to this theory, Bugbee and Bernt (1990) found that test takers at The American College performed better on computerized tests than their counterparts who took paper-and-pencil tests. Similar results were found by Parshall and Kromrey (1993) in their examination of the pilot test of the computerized GRE. Ployhart, Weekley, Holtz and Kemp (2003) also found better performance on Web-based than paper-and-pencil tests, as did a nineteen-month study of students at a U.K. university (Bocji & Greasley, 1999).

On the other hand, Green, Bock, Humphreys, Linn and Rekase (1984) concluded that students tend to perform better on paper-and-pencil examinations, and Russell (1999) as well as Russell and Haney (1997) found that computer-administered mathematics tests underestimated student ability. Smith and Caputi's research (2004) focused on attitudes toward computerized testing. They devised a scale to assess test takers' reactions to a computerized versus a traditional paper-and-pencil environment. Factor analysis revealed two significant factors regarding perceptions about computer-based testing: (1) ease of use and (2) confidence in compu-

ter-based testing (Smith & Caputi, 2004, p. 417). They also found that practicing computerized tests before the actual testing was an important factor in test taker confidence. However, Glowacki, McFadden and Price (1995) and Baird and Silvern (1992) found no significant differences between the two test modes. Finally, Potosky and Bobko (2004) found that test equivalency usually depends on the content (e.g., math, verbal or spatial) and timing (timed versus un-timed) of the test. Thus, it is not clear which testing format is superior, or whether there is any measurable difference.

Other researchers have suggested potential explanations for why results vary. Wood (1984) proposed that computerized tests may be less supervised, allowing greater opportunities for cheating. Alternatively, computerized testing may be affected by the student's ability to use and feel comfortable with the computer or testing software (Gershon & Bergstrom, 1991; Lee, 1995; Wood, 1984). Third, several researchers have raised concerns about whether the various testing modes are examining the same items (Lee, 1995; Peterson & Reider, 2002; Wise & Plake, 1990; Wood, 1984). Fourth, Bugbee and Bernet (1990) reported that some students complained about the usability of computer equipment and its performance. Indeed, the Educational Testing Service lost the right to administer the GMAT due to technical glitches (Merritt, 2003). Finally, previous studies have sometimes failed to use "real" test takers (Ployhart et al., 2003). Ployhart et al. argue that "theoretical and practical reasons thus require an examination of Web-based tests in the actual sample and context for which they are ultimately to be used" (2003, p. 735).

Recognizing the concerns over the move to computer-based testing, the Insurance Institute of America (IIA) and the American Institute for Chartered Property Casual Underwriters (CPCU) made the transition from paper-and-pen testing to computer-delivered testing at a very cautious pace (Oakes, 1999). They began with a six-month pilot test offering exams on computer that were identical to the paper exams, except that students would

enter their essay answers by computer. Over a subsequent two-year period, they moved entirely to computers and changed the exam format from essay to multiple choice and short answer. Oakes, Senior Vice President of Examinations, indicated that during the phase-in period they did receive some requests from exam takers, particularly those over 50 years old, who either were not computer literate or could not type fast enough to complete the exam in the allotted time period (D. Oakes, personal communication, May 8, 2006). However, no data was kept on the impact on students of the move to computers. Nonetheless, the IIA and CPCU exams are not mandatory. Therefore, it is not known whether some potential students were dissuaded by the computer-only environment.

Therefore, despite the fact that computer-based testing has been conducted for many years, a number of questions remain. Does the performance of some students vary as a function of test mode? Are some students disadvantaged by computer-based testing compared to their computer confident peers? To our knowledge, no studies have explicitly focused on the impact of meta-cognitive aspects on computer-based test performance. We have attempted to address these two issues in the present research.

The Role of Meta-Cognition

Broadly, meta-cognition refers to what we know about our own cognitive processes (Flavell, 1979; Hacker, 1998). A number of meta-cognitive processes could be at play when students make decisions about the test format or make predictions about their test performance. For instance, meta-cognitive factors may play a role when people evaluate their mastery over the to-be-tested material (i.e., make judgments about knowing), assess the advantages and disadvantages of computer versus pencil-and-paper tests for the specific material, and make a choice that presumably leads to an optimal outcome.

In addition to these meta-cognitive factors, individual personality differences (i.e., individual differences) could also play a role in meta-cognitive processes, choice of test format and performance. Individual differences are distinct from meta-cognitive factors, as meta-cognition refers to people's reflections on their own cognitive processes, whereas individual differences relate to more stable personality dimensions.

Specifically, three individual difference measures that could complement meta-cognitive measures include the Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984), the Personal Need for Structure Scale (Neuberg & Newsome, 1993) and the Need to Evaluate Scale (Jarvis & Petty, 1996). The Need for Cognition (NC) scale (Cacioppo et al., 1984) is well validated and widely used for assessing people's preference for engaging in effortful cognitive tasks. People with high need for cognition (high NC) prefer complex, cognitively demanding tasks, whereas individuals with low need for cognition (low NC) prefer simpler, more straightforward tasks. This may translate into preferences for certain types of test formats. For instance, to the extent that individuals perceive that computer-based tests involve more cognitive resources, those with high NC may be more likely to choose computer-based tests and be more comfortable with that format compared to those with low NC.

The Personal Need for Structure (PNS) scale was developed to assess individuals' preferences for a simplified structure. Specifically, people with high PNS tend to view objects and situations in simple rather than complex ways. In addition, people with high PNS prefer to apply existing structures (e.g., schemas, scripts) to new situations, and they are uncomfortable in situations where such structure is lacking (Leone, Wodglin, & Wallace, 1999; Neuberg & Newsome, 1993). The PNS scale is designed to assess people's preferences for order and predictability in their personal and social worlds. Thus, as with Need for Cognition, PNS may play a role in people's decisions about test taking. To the extent that individuals have high PNS, they may be more likely to choose a test format with which they are more comfortable or have more experience.

The Need to Evaluate Scale (NE; Jarvis & Petty, 1996) is designed to assess an individual's chronic tendency to engage in evaluative responding. This well-developed and validated scale has demonstrated that people whose scores reflect a high need to evaluate are more likely to report attitudes toward a variety of topics, have more evaluative thoughts during the day, and make evaluations when encountering new social objects (Jarvis & Petty, 1996). People with high NE may be more likely to choose computer-based tests because they tend to provide immediate evaluations, unlike paper-and-pencil tests.

Overview of Research

The purpose of the research described here was threefold. First, we wanted to assess the meta-cognitive reasons students give for choosing a particular test format. We suspect that test format choice reflects an estimation of the extent to which the material has been mastered (i.e., meta-comprehension; Maki & Berry, 1984) and a judgment as to which test format would maximize performance. This judgment probably consists of an evaluation of computer skills, an assessment of comfort with each testing format, and an assessment of comfort with change. A careful look at the meta-cognition and education literature reveals little empirical research in this specific area, and to our knowledge, this paper is the first to explore the issue of meta-cognitive factors in test-format choice.

A second goal of this research was to explore the role of individual differences in people's choices to write exams in either paper-and-pencil or computer-administered format. It has been demonstrated that performance can differ across test administration formats. Thus, we examined the personality characteristics Need for Cognition, Personal Need for Cognition, and Need to Evaluate. These measures have been previously tested and are well documented in the literature as valid and reliable. In order to maximize the utility and predictive ability of our research findings, we ensured that identical condi-

tions were applied to the two studies. Thus, all tests were proctored to ensure that cheating would not be an issue. In addition, practice tests on computer were available to all students to ensure that they were familiar with the system. All students had at least minimal computer skills, and the computer hardware was standardized. We also used identical multiple-choice questions for both computer-based and paper-and-pencil tests. Furthermore, the questions were presented in the same order and computerized test takers were free to return to previous questions and change their answers. Finally, all test takers were university students enrolled in a second-year undergraduate financial accounting course. The course is a requirement for all commerce students regardless of their intended area of concentration. We feel that these conditions provided a viable and defensible approach to studying the preference for, as well as potential performance differences in, computer versus paper-and-pencil test administration.

In Study 1, students were given the opportunity to write two exams in either paper-and-pencil or computer format and to describe the meta-cognitive reasons for their decision. In Study 2, students in a subsequent offering of the same accounting course chose the test format for two exams, described the meta-cognitive reasons for their decisions, predicted their scores and assessed their confidence in these predictions. We also obtained actual test scores for this sample, allowing us to measure the accuracy of their predictions and evaluate the effectiveness of their meta-cognitive processes.

STUDY 1

Methods

Participants

The participants were 162 students enrolled in an Introductory Accounting course (out of 191 registered). They completed both course exams and completed questionnaires measuring their meta-cognitive processes. Data on gender was not collected, but of the students registered in the class, 116 were male and 75 were female. Nonetheless, in an attempt to determine whether paper versus computerized test takers differed, the total population (i.e., all 191 students registered in the course) was examined. Students who switched formats were not included in this examination because they would appear in both groups. No significant differences in grades were noted between pencil-and-paper and computerized test takers. Gender had no significant effect on test scores either. However, t-tests indicated that women scored better on the assignment component of the course ($M = 17.51$ for males and $M = 18.90$ for females; $t = 2.44$, $p < .05$) as well as on the final grade ($M = 64.21$ for males and $M = 68.27$ for females; $t = 1.97$, $p < .05$). Because assignment grades are based on completion of the work, this suggests that women complete more assignments than males do.

Materials and Procedure

Participants' reasons for choosing the paper-and-pencil versus computer-based exam were assessed at the end of the course during teaching evaluations. The students wrote two exams during the course, and students had the option of writing a paper-and-pencil or computer-based version. The course instructor told the students that she was studying the computer-based exam process. Students were told that the computer-based exam would have the same content as the paper-and-pencil exam, and that the marking would be done by computer and double checked by the instructor. Students had to sign up for spaces in university computer labs in order to write the exam by computer. They were told that

the choice to write by computer was completely optional, but that because many professional exams are computer-based, this format would provide useful practice.

At the end of the course (and after having written the two exams), students were asked to anonymously indicate which version of the test they chose for the first exam and the reason for that choice. Next, participants were asked to indicate their choice for the second exam, their reasons for that choice, and their reasons for changing from paper-and-pencil to computer (or vice versa) for the second exam (if applicable).

Results

For the first exam, 89 participants chose the paper-and-pencil exam, whereas 73 chose the computer-based exam. For the second exam, eight students who wrote the paper-and-pencil exam switched to a computer exam. Twenty-two students who wrote the computer exam switched to a paper-and-pencil exam. Four students who wrote the first exam did not write the second exam and their data were excluded from our analyses.

Coding of open-ended responses

In all, 158 participants provided a total of 205 reasons for choosing the paper-and-pencil over the computerized version for the first exam, and 150 participants reported 151 reasons in all for their choice for the second exam. Although 30 students changed their exam format choice from the first to the second exam, only 17 provided reasons for that change. Their reasons were categorized by the themes reflected in their responses. Statements were coded by two independent raters. The 373 statements describing students' reasons for their choices fell into approximately 20 categories.

Reasons for choosing the exam format

The reasons that students provided for their choice for the first exam are summarized in Table 1 (columns 1 and 2). We separated the reasons provided by those who chose the paper-and-pencil format from those who chose the computerized format. As illus-

trated in Table 1, the most common response (37%) for those who chose the paper-and-pencil format ($n = 89$) was greater comfort with this arrangement than with computers. An additional 24% of respondents mentioned that they were uncomfortable with computers. A further 20% expressed concerns about potential technical problems or computer glitches. Thus, the vast majority of comments reflected comfort with the test format and concerns about working with computers. Other concerns centered on the ability to understand the items (7%) or make changes (6%). Finally, a number of students (16%) believed, incorrectly, that the computer format would preclude them from using scrap paper to perform calculations (Table 1).

A majority of the 73 students who completed the computer-based exam noted the benefits of receiving immediate feedback on their test performance (55%) as the reason for choosing this test format (Table 1 – columns 3 & 4). A substantial percentage of students (40%) mentioned that the novelty of computer-based exams was appealing. Students also indicated that the computer exam was easier to follow (10%), that they were more comfortable with computers (14%) and that it was good preparation for future exams, such as the GRE and GMAT (10%).

For the second exam, of the 151 reasons provided, 73 were “same as above.” In other words, people were reiterating their reasons for choosing the first exam. Many students who actually wrote reasons simply repeated the reason they had provided for the first exam. Finally, many of the reasons provided were actually reasons for switching from paper-and-pencil to computer (or vice versa). In other words, participants did not provide a reason for their choice for the second exam, but rather why they switched formats. For those who did not switch, the pattern of reasons was very similar to that found in Table 1, so we have not presented these results in detail.

Reasons for change

As indicated above, 30 participants switched test formats from the first to the second exam, and they reported a number of reasons for doing so. Some students reported that their poor performance on the first test prompted them to try a change of format in the hope that the format change might produce better test performance. Other students reported wanting to change because the computer format provided faster feedback or because they were more familiar with the paper-and-pencil format.

Discussion

Overall, the reasons provided for choosing each test format were quite different, and they shed some light on the psychological processes that occur when students assess their mastery of the material to be tested, evaluate their comfort with computer technology and estimate which test format will optimize their test performance. However, despite these insights, a number of other factors are likely to influence the choice of test format and performance. Importantly, in Study 1, we did not assess actual student performance on the exams, differences in test format or meta-cognitive factors, or individual differences. We therefore conducted a second study to confirm the results of Study 1, and in an attempt to explore the potential role of some additional individual difference factors as well as the impact of these factors on performance.

STUDY 2

Overview

The purpose of Study 2 was threefold: 1) to examine further meta-cognitive aspects of participants' decisions to choose paper-and-pencil versus computer-based testing; 2) to explore the potential roles of individual difference variables in these decisions and in performance; and 3) to explore whether predictions of test performance and actual test performance varied across the students who provided reasons for their choice of test format. Thus, in Study 2, participants disclosed their reasons for the choice of test format, predicted their scores on the first test, and provided confidence ratings for their predictions. Participants also allowed us to collect their actual test scores in order to determine the accuracy of their predictions. Finally, participants completed questionnaires based on specific individual difference measures, namely the Need for Cognition Scale, the Personal Need for Structure Scale, and the Need to Evaluate Scale. We predicted that people with high need for cognition would find a computer-based exam more challenging and interesting than a paper-and-pencil exam. Alternatively, low NC participants would prefer the more familiar paper-and-pencil tests. Since the PNS scale is designed to assess people's preferences for order and predictability in their personal and social worlds, we predicted that personal need for structure would predict test format choice such that people with high PNS would prefer the paper-and-pencil exam to the computer-based exam. With respect to the Need to Evaluate Scale, we predicted that people who score highly on this measure would be more likely to provide specific reasons for their choice of using the paper-and-pencil versus the computer-based exam. We also predicted that they would be more likely to take the computer exam because of the rapid feedback.

Method

Participants

A total of 88 out of 105 students who were registered in an Introductory Accounting course (48 male, 37 female and three unreported) participated in the first part of Study 2. All 88 participants provided data on predicted and actual grades, along with confidence ratings about their prediction. Seventy-one of the students in this group provided individual difference data. A subset of 43 participants provided additional information on their predicted grades, actual grades and judgments of confidence (20 male and 23 female).

Materials and Procedure

Data were collected in several phases. After the first test, participants were asked to estimate their test grade (from 0% to 100%). In addition, participants were asked to indicate how confident they were that their estimate was accurate (also on a scale from 0% to 100%). In the second phase, participants completed a questionnaire addressing their reasons for choosing the test format. Finally, participants were asked to complete a series of personality measures.

The Need for Cognition Scale is a balanced 18-item scale that has been well developed and validated in previous research (Cacioppo et al., 1984). Participants responded to items such as, "I only think as hard as I have to," (reverse scored) and, "I prefer complex to simple problems," on a 1 (Extremely Uncharacteristic) to 5 (Extremely Characteristic) Likert-type scale. Negatively keyed items were reverse scores and responses were averaged. High scores reflect a high need for cognition. Scores ranged from 1.94 to 4.78 with $M = 3.30$ and $SD = 0.62$. Observed internal reliability was good ($\alpha = .79$).

The Personal Need for Structure scale is a frequently used, validated and balanced 12-item scale with four reverse keyed items (Neuberg & Newsome, 1993). Participants responded to items such as, "I enjoy being spontaneous," (reverse scored) and, "I hate to change my plans at the last minute," on a 1

(Strongly Disagree) to 6 (Strongly Agree) Likert-type scale. All items were averaged to obtain the overall score. PNS scores ranged from 1.77 to 5.50 with $M = 3.67$ and $SD = .85$. Observed reliability of the scale was good ($\alpha = .79$). High scores reflect a strong personal need for structure.

The Need to Evaluate scale is a reliable and valid 16-item scale with six negatively worded items (Jarvis & Petty, 1996). Participants responded to questions such as, "I prefer to avoid taking extreme positions," (reverse scored) and, "I form opinions about everything," on a 1 (Extremely Uncharacteristic) to 5 (Extremely Characteristic) Likert response scale. All items were averaged to obtain the overall score. High scores reflect a strong need to evaluate. Need to Evaluate ranged from 1.94 to 4.44 with $M = 3.25$ and $SD = .59$. Observed reliability of the scale was good ($\alpha = .80$).

Results

Of the 71 participants who provided reasons for their choices, 50 chose the paper-and-pencil exam, whereas 21 chose the computer-based exam. Only four participants changed from paper-and-pencil to computer or vice-versa. Data were coded as in Study 1. The 71 participants reported a total of 78 reasons for their choice of exam format for the first midterm.

Reasons for choosing the exam format

The reasons provided for the choice for the first exam are summarized in Table 2. As in the first study, we separated reasons for choosing the paper-and-pencil exam (columns 1 & 2) from reasons for choosing the computer-based exam (columns 3 & 4). As shown in Table 2 (columns 1 and 2), the reasons cited for choosing the paper-and-pencil exam are similar to those in Study 1. Students reported being more comfortable with this format, that they liked that they could do hand calculations and make changes and that they had concerns about technical issues. Students who chose the computer-based exam did so primarily because they could

have immediate access to their grades, for the novelty, to prepare for the future, and because they were comfortable with computers.

Participant's reasons for their choice for the second exam mostly paralleled their reasons for choosing the first exam format. As a result, 47 of the 71 total reasons provided were "same as previous." Only four participants switched formats between exams, and they did so for novelty reasons or because they were dissatisfied with their performance on the first exam and reasoned that a change in test format might contribute to better performance on the second exam.

Prediction of grades

As described above, 78 participants provided their predicted grades on the first midterm and their confidence in their predictions, and allowed us to collect data on their actual grades in the course. Interestingly, participants who chose the computer-based exam predicted higher grades (percentage correct on the test; $M = 78\%$) than those who took the paper-and-pencil exam ($M = 71\%$, $F_{(1,77)} = 3.84$, $p = .05$). Although participants who wrote the computer-based exam were no more confident ($M = 71\%$) than participants who wrote the paper-and-pencil exam ($M = 69\%$, $F_{(1,77)} = 0.28$, $p = .59$), they did score better on the exam ($M = 83\%$ versus 75% for computer and paper-and-pencil respectively, $F_{(1,77)} = 7.09$, $p < .01$). As in Study 1, gender has no effect on exam scores.

An interesting question that arises is why did students who chose computer-based tests perform better than those who chose paper-and-pencil tests? It is important to remember that the exams were identical (i.e., they had the identical questions and grading scheme). Therefore, the content of the exam cannot be the cause of any differences in grade. Does it have to do with the method of taking the exam? Or are they simply better students? Although students could choose the format for the midterm, all students had to complete hand-in written assignments as part of the overall course requirements (as was the case for the students in Study 1).

Interestingly, when the marks on the assignments were compared, students who wrote the computer-based midterm did no better on the assignments ($M = 27.63$) than those who wrote the paper-and-pencil exam ($M = 27.11$, $F_{(1,77)} = .35$, $p = .56$).

The final exam grades also suggest that students who wrote the computer-based exam did not have superior mastery over the material than those who chose the paper-and-pencil exam. The final exam was a paper-and-pencil exam administered during the formal final examination period. Again, as with the assignments, final exam scores did not differ significantly ($F_{(1,77)} = 0.38$, $p = .54$; $M = 65\%$ versus 67% for paper-and-pencil versus computer-based exams, respectively). Although there were no gender differences in test scores (see above), there were significant gender effects on assignments. As in Study 1, t-tests indicated that, although women scored better on assignments ($M = 29.65$ for females and $M = 25.30$ for males; $t = -2.81$, $p < .05$), gender was not a significant predictor on any of the other dependent measures (all $p > .09$). Therefore, gender cannot explain our results. Although not conclusive, these findings suggest that the test-taking format may be important for test performance, independent of ability.

Personality as a predictor of decisions

As described above, 43 participants provided data on individual difference measures and the reasons for their decisions. We conducted a regression analysis with the three individual difference measures (Need for Cognition, Need for Evaluation and Personal Need for Structure) as predictors¹, and participant's choice of test format as the dependent variable (see Table 3 for correlations among all predictor and dependent measures). Surprisingly, only participant's Need for Cognition predicted choice of test format ($B = .014$, $SE = .007$, $t = 2.19$, $p < .04$; see Table 4). As predicted, the positive and significant standardized regression coefficient indicates that participants with high Need for Cognition were more likely to select the computer-based exam. However, contrary to our prediction, personal Need for Structure did not predict choice, nor did Need to Evaluate (Table 3).

In order to assess the impact of individual differences on the reasons for choice of exam format, we conducted a median split on each of the three individual difference variables and then examined the reasons listed for their choice of exam format by participants who scored high versus low on the specific measure. Providing some support for our predictions, participants with high Need to Evaluate provided more reasons for their choice (49) than participants with low Need to Evaluate (36). In addition, participants with high Need to Evaluate were more likely to indicate that they wanted to experience the novelty of the computer-based exam (eight versus none). There was only one apparent difference in the reasons provided in terms of Need for Structure: participants with high Need for Structure were more likely to indicate that their comfort with the situation was the key deciding factor in their choice of format for the midterm (53% versus 34% for low PNS participants). Finally, there was one apparent difference in terms of Need for Cognition. Supporting our prediction that participants with high Need for Cognition would choose the computer-based exam, only high NC participants indicated that they chose the computer exam format because of the novelty (9% versus 0% for low NC individuals).

Individual differences in prediction of grades

A series of regression analyses were conducted with the three individual difference measures as predictors and predicted grade, confidence or actual grade as the dependent measure. Importantly, because choice of exam format was a predictor of grades, choice of exam format was included as a control variable in all three regression equations (i.e., any effect of the individual difference variable would be independent of exam format choice).

Only Need for Cognition proved to be a reliable predictor. Specifically, Need for Cognition was a significant and positive predictor of actual grade ($B = .004, SE = .002, t = 2.08, p < .05$). The significant and positive standardized regression coefficient indicates that higher Need for Cognition scores were predictive of higher grades (Table 5). Need for Cognition was also a significant predictor of anticipa-

ted grades ($B = .003, SE = .002, t = 1.99, p = .05$, indicating that higher Need for Cognition was also related to higher predicted grades (Table 6). However, Need for Cognition did not significantly predict participant's confidence ($B = .004, SE = .002, t = 1.64, p = .11$; see Table 7)). Need to Evaluate and Personal Need for Structure did not predict any of the dependent variables (Tables 5–7)).

General Discussion

Summary and Implications

The purpose of these two studies was first to assess the meta-cognitive reasons students have for choosing a particular test format, and second to explore the role of individual differences in students' choices to write exams in either paper-and-pencil or computer-administered format, as well as the impact of this choice on performance. In our sample, students who opted for computer-based tests did so because they perceived the immediate feedback (i.e., grade) to be an important advantage over paper-and-pencil tests, which take longer to grade. These students also said that they were comfortable with computers and were willing to try something new. Students who opted for paper-and-pencil exams expressed discomfort with computer technology or concerns that technical glitches might affect their performance. Of the students who changed from one test format to another, some thought that trying a new test format might contribute to superior performance on the second exam.

Study 2 allowed us to examine whether students' choices of test format were associated with their test performance. We found that students who took the computer-based test had a higher mean score than students who took the paper-and-pencil test. However, this does not seem to reflect a difference in the quality of the students who chose to write paper-and-pencil versus computer-based exams, as the assignment and final exam marks did not differ between the groups. This difference in mean score could reflect the greater comfort of students who

wrote in the computer-based environment. There are a number of possible explanations for this. It is clear from both studies that students who wrote the paper-and-pencil version were uncomfortable with a computer-based approach, and many indicated that they did not trust computers. We might assume that the opposite was true for students who preferred computer-based tests. This comfort may have resulted in less anxiety and better overall performance. This may help explain why some studies found a computer-based advantage in terms of grades.

The results of Study 2 on individual differences as a predictor of exam format choice are intriguing. As we predicted, Need for Cognition predicted the choice of test format. Specifically, people with high Need for Cognition were more likely to indicate that they had chosen a computer-based exam for novelty reasons. In addition, Need for Cognition predicted test performance. Although not specifically predicted, this result is unsurprising, as previous work has shown a weak but significant association between grades and Need for Cognition (Cacioppo et al., 1984) as well as self-efficacy, grade point average (GPA) and Need for Cognition (Elias & Loomis, 2002). This further reinforces our supposition that people with high Need for Cognition are more comfortable with novel situations, and may therefore perform better overall. One noteworthy finding from our research is that one of the primary and consistent reasons participants gave for choosing computer-based testing was comfort with and faith in the reliability of computers.

Unanswered Questions and Future Directions

As with any study, there are some questions that remain unanswered. First, the applicability of our results to a general population is questionable as we assessed only students in an accounting course. However, this is the group who will be targeted with the most online testing in future (e.g., GMA, GRE). A second more minor limitation is that in Study 1 we asked students to provide their reasons

for choosing a test format only after their second exam. It would have been better to ask this question after each exam, which may have resulted in a greater number of reasons provided. Nonetheless, the diversity of reasons given suggests that this methodological issue did not have a significant negative impact on results.

Nonetheless, these unanswered questions, as well as the above-mentioned issue of the generalizability of second-year accounting students provide fruitful ground for future research. Although we feel that our results on the role of meta-cognitions and individual differences in choice of test taking format and performance are intriguing, there are many directions that remain to be explored.

Overall, the two studies reported here suggest that concerns about computer-based testing may be unnecessary. As people become more comfortable with this test taking format, there may be little or no performance decrement. In fact, some people (such as those with high Need for Cognition or who enjoy novelty) may actually find that their performance is enhanced when they take computerized tests. Thus, one direction for future study would be to assess the general population's comfort level with computers. Indeed, as long as the material is held constant across test-taking modalities, and as long as people trust the computer, there is no reason for computer-based testing not to be more widely used.

Table 1: Number and percentage of respondents who indicated specific reasons for their choice for Exam 1 by paper-and-pencil versus computer: Study 1.

Reason	P&P		Computer	
	Number ^a	Percentage ^b	Number	Percentage
More comfortable/familiar	33	37%	10	14%
Dislike computers	21	24%	-	-
Easier to follow	6	7%	7	10%
Can do hand calculations	13	16%	-	-
Can make changes	5	6%	-	-
Concern with Tech issues	18	20%	-	-
Access to computer room	7	8%	-	-
Concerns about marking	3	4%	-	-
Less stress	1	1%	3	4%
Preparation for future	-	-	7	10%
Opportunity to cheat	-	-	1	1%
Immediacy of grades	-	-	40	55%
Novelty	-	-	30	40%

a = of 89 paper-and-pencil respondents and 73 computer respondents;

b = percentage rounded; percentages do not add up to 100 as some respondent provided more than one reason.

Table 2: Number and percentage of respondents who indicated specific reasons for their choice for Exam 1 by paper-and-pencil versus computer: Study 2.

Reason	P&P		Computer	
	Number ^a	Percentage	Number ^b	Percentage
More comfortable/familiar	24	48%	9	52%
Dislike computers	1	2%	-	-
Easier to follow	5	10%	2	10%
Can do hand calculations	6	12%	-	-
Can make changes	6	12%	-	-
Concern with Tech issues	7	14%	-	-
Access to computer room	2	4%	1	5%
Concerns about marking	-	-	-	-
Less stress	1	2%	-	-
Preparation for future	-	-	2	10%
Immediacy of grades	-	-	9	42%
Novelty	-	-	3	15%

a = of 48 paper-and-pencil respondents and 21 computer respondents

b = percentage rounded; percentages do not add up to 100 as some respondents provided more than one reason

Table 3: Correlations among predictors and outcome variables: Study 2

	Need for Structure	Need to Evaluate	Need for Cognition	Gender	Exam Choice	Actual Grade	Predicted Grade	Grade Confidence
Need for Structure	1.00							
Need to Evaluate	.12	1.00						
Need for Cognition	.18	.32*	1.00					
Gender	.09	.26	.22	1.00				
Exam Choice	.11	-.13	.25	-.29	1.00			
Actual Grade	-.04	.01	.30*	.00	.20	1.00		
Predicted Grade	-.14	-.01	.24	-.15	.03	.72**	1.00	
Grade Confidence	-.16	.02	.25	.23	-.09	.26	.43**	1.00

Note: N = 43; *p < .05; **p < .01

Table 4: Coefficients, significance levels, and effect size for individual differences as predictors of choice of exam format.

Factor	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i> =	<i>Eta</i> ²
Need for Structure	.055	.065	0.84	.31	.02
Need to Evaluate	-.114	.131	-0.87	.75	.02
Need for Cognition	.014	.007	2.19	.04	.12

Table 5: Coefficients, significance levels and effect size for individual differences as predictors of actual grade.

Factor	<i>B</i>	SE	<i>t</i>	<i>p</i> =	Eta ²
Need for Structure	-.007	.017	-0.39	.70	.00
Need to Evaluate	-.020	.032	-0.64	.52	.01
Need for Cognition	.004	.002	2.08	.04	.11

Table 6: Coefficients, significance levels and effect size for individual differences affecting grade predictions.

Factor	<i>B</i>	SE	<i>t</i>	<i>p</i> =	Eta ²
Need for Structure	-.017	.016	-1.03	.31	.03
Need to Evaluate	-.009	.030	-0.32	.75	.00
Need for Cognition	.003	.002	1.99	.05	.10

Table 7: Coefficients, significance levels and effect size individual differences as predictors of confidence in grade.

Factor	<i>B</i>	SE	<i>t</i>	<i>p</i> =	Eta ²
Need for Structure	-.033	.022	-1.46	.15	.06
Need to Evaluate	-.021	.041	-0.51	.61	.01
Need for Cognition	.004	.002	1.64	.11	.07

References

- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, 11, 121-136.
- Baird, W. E., & Silvern, S. B. (1992). Computer learning and appropriate testing: A first step in validity assessment. *Journal of Research on Computing in Education*, 25, 18-28.
- Bocij, P., & Greasley, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 1, 1-16.
- Bugbee, A. C., Jr. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28, 282-290.
- Bugbee, A. C., Jr., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education*, 23, 87-101.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48, 306-307.
- Chalhoub-Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273-299
- Chapelle, C. A. (2001). *Computer application in second language acquisition*. Cambridge, UK: Cambridge University Press.
- Elias, S. M., & Loomis, R. J. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance. *Journal of Applied Social Psychology*, 32, 1687-1702.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.
- Gershon, R. C., & Bergstrom, B. (1991, April). *Individual differences in computer adaptive testing: Anxiety, computer literacy, and satisfaction*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.
- Glowacki, M. L., McFadden, A. C., & Price, B. J. (1995, November). *Developing computerized tests for classroom teachers: A pilot study*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED391471). Retrieved from Education Resources Information Center [ERIC] Web site: <http://www.eric.ed.gov>
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347-360.
- Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1-24). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jarvis, W. B. G., & Petty, R. E. (1996). The need to evaluate. *Journal of Personality and Social Psychology*, 70, 172-194.
- Lee, M. (1995, November 13). Technology (A special report): In the classroom – Examined Options. True or false: Standardized tests can be easily put on computers. *Wall Street Journal* (Eastern ed.), p. R26.
- Leone, C., Modglin, K., & Wallace, H. M. (1999). The need for closure and the need for structure: Interrelationships, correlates, and outcomes. *Journal of Psychology*, 133, 553-562.
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 663-679.
- Merritt, J. (2003, December 29). Why the folks at ETS flunked the course: A tech-savvy rival service will soon be giving B-school applicants their GMATs. *Business Week*, p. 48.
- Neuberg, S. L., & Newsome, J. T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology*, 65, 113-131.
- Oakes, D. R. (1999). Institutes moving to computerized exams. *National Underwriter*, 103, 42.

- Parshall, C. G., & Kromrey, J. D. (1993, April). *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. (ERIC Document Reproduction Service No. ED363272). Retrieved from Education Resources Information Center [ERIC] Web site: <http://www.eric.ed.gov>
- Peterson, B. K., & Reider, B. P. (2002). Perceptions of computer-based testing: A focus on the CFM examination. *Journal of Accounting Education, 20*, 265-284.
- Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56*, 733-752.
- Potosky, D., & Bobko, P. (2004). Selection testing via the internet: Practical considerations and exploratory empirical findings. *Personnel Psychology, 57*, 1003-1034.
- Rabinowitz, S., & Brandt, T. (2001). *Computer-based assessment: Can it deliver on its promise? Knowledge brief*. San Francisco, CA: WestEd. (ERIC Document Reproduction Service No. ED462447). Retrieved from Educational Resources Information Center [ERIC] Web site: <http://www.eric.ed.gov>
- Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives, 7*, 472-522.
- 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). Retrieved from the publication's Web site: <http://epaa.asu.edu/epaa/v5n3.html>
- Smith, B., & Caputi, P. (2004). The development of the attitude towards computerized assessment scale. *Journal of Educational Computing Research, 31*, 407-422.
- Wainer, H., Dorans, N. J., Eignor, D., Flaughner, R., Green, B. F., Mislavy, R. J., et al. (2000). *Computer adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wise, S. L., & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development, 23*, 3-11.
- Wood, S. (1984). Computer use in testing and assessment. *Journal of Counseling and Development, 63*, 177-179.
- Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education, 15*, 337-362.

Endnotes

- 1 Although gender was initially included in all regression analyses, it did not achieve statistical significance in any analysis, thus it is not reported here.