

Improving the Uniform CPA Examination

October 2007



American Institute of Certified Public Accountants

Contents

Executive Summary	1
Background	2
CBT Accomplishments	2
Opportunities for Change	3
<i>Task-Based Simulations</i>	3
<i>Written Communication Simulations in BEC</i>	4
<i>Testing Written Communication in the Future</i>	5
<i>Other Simulations in BEC</i>	5
<i>Assessing Additional Communication Skills</i>	6
<i>Testing Windows</i>	6
<i>Length of the Examination</i>	6
Risk Factors Impacting Decisions	7
Assumption	9

Executive Summary

Computer-based testing (CBT) has produced significant improvements to the CPA Examination. It has not yet accomplished all of the improvements that were envisioned and experience has identified additional ones that could be made. A practice analysis is underway that will result in revisions to the content of the CPA Examination. The Board of Examiners (BOE) wants to use the opportunity provided by the practice analysis to make additional improvements to the examination. These improvements will result in higher quality, better customer service, and a more efficient infrastructure supporting the examination. There are a number of potential improvements and some are mutually exclusive.

The most important customer service improvement is to release scores on a predictable and shortened timeline. The recommendations provided in this paper are designed, among other customer satisfaction improvements, to ensure that score reporting can become more predictable and more rapid.

At its September 14-15, 2007 meeting, the BOE authorized AICPA Examinations staff to work on the following projects this fall:

1. Continue the development of task-based simulations,
2. Introduce task-based simulations at the same time as the practice analysis results are introduced,
3. Begin the administration of simulations in BEC by consolidating the administration of constructed response writing prompts in the Business Environment and Concepts (BEC) section at the same time as the practice analysis results are introduced,
4. Determine stakeholder receptivity to objective assessment of communications skills (including, but not limited to, written communication) to replace essay assessment of communications skills. If stakeholders are amenable to objective assessment of communication skills, investigate the feasibility of such assessments for the CPA Exam and, if appropriate, incorporate them into the operational Exam at a future point,
5. Introduce additional task-based simulations in BEC when objective assessment of communication is introduced to replace essay questions,
6. Maintain the current administration schedule (i.e., two months of testing followed by a month in which the test is not administered), and
7. Once the statistical properties and content of the new exam are known, investigate possible reductions in total testing time.

The BOE also requested that input from state boards and other stakeholders be sought to ensure that they support the changes under consideration.

Improving the Uniform CPA Examination

Background

The introduction of computer-based testing (CBT) for the Uniform CPA Examination in April 2004 was the culmination of several years of research and development. The specifications and features of the CBT were determined through a collaborative process among State Boards of Accountancy, the AICPA, NASBA, and Prometric. There were several task forces and committees as well as invitations to comment and exposure drafts that provided input concerning the needs and desires of state boards.

The CPA Exam that resulted from this collaborative work represented several compromises. State boards were not unanimous in the features they wanted to see in the CBT, technology did not permit all desired features to be implemented, and unanticipated results of some work prevented some features from being implemented at that time.

The purpose of this paper is to review the accomplishments of CBT, identify features that could be added or changed, explain the dependencies and conflicts among these possible features, recommend features to be added or changed, and to propose a process for incorporating desired additions and changes to the CBT, with the endorsement and influence of State Boards.

CBT Accomplishments

The CBT offers numerous improvements over its paper-based predecessor in the Exam and in the process of taking the test. Some of the more significant ones are listed and explained briefly in this section.

Introduction of a new section, Business Environment and Concepts (BEC). The 2001 practice analysis indicated that it was important for entry-level CPAs to be able to understand the business reasons for transactions and the attendant accounting implications. The BEC section includes assessment of knowledge of business structure; economic concepts essential to obtaining an understanding of an entity's operations, business and industry; financial management; information technology; and planning and measurement.

Enhanced assessment of skills. The introduction of simulations allowed for authentic assessment and objective scoring of candidates' job-related abilities to conduct research in the authoritative literature, complete financial statements and tax forms, and analyze and interpret data. The simulations also allow candidates to use tools that are similar to those they use in their work (word processor, spreadsheet, and search engine).

Faster score reporting. In the paper-and-pencil era, candidates received their results approximately three months after they took the Exam. On average, the AICPA now releases scores within 23 to 30 days after candidates sit for the Exam, depending on section. In fact, the AICPA releases around 30 percent of scores in less than 20 days.

Computerized scoring of essays. A significant majority of the essays are scored through an automated scoring technology. This technology, which develops scoring algorithms based on a sample of human scores is at least as reliable as human scoring and is less expensive.

Introduction of adaptive administration of multiple-choice testlets. The multiple-choice portion of the Audit, Regulation, and Financial Accounting and Reporting sections of the Exam are administered adaptively. Adaptive testing adjusts the difficulty of the test based on the candidate's performance and allows for shorter and/or more precise tests than traditional ones.

Increased access to the Exam. Each section of the Exam is typically available six days a week, eight months of the year. Previously, each section was administered on one day in May and one day in November.

Professional testing environment. Prometric test centers are much more appropriate settings for professional testing than the large, uncomfortable venues used for paper-based testing.

Platform for innovation. The initial release of CBT was intended to be just that, an initial release. The ability to continue to innovate and change the Exam was a key component to the decision to implement CBT. In fact, improvements to the candidate interface, the written communications portion of the simulation, the authoritative literature and the research questions have been implemented since CBT was introduced.

Increased candidate satisfaction. Candidates' surveys and focus groups consistently indicate satisfaction with the computer-based test.

Opportunities for Change

The improvements offered by CBT have been impressive and important. However, some desired features were not implemented with the initial release and experience has changed some stakeholders' perceptions concerning existing and desired features. In this section, some of the key potential changes are briefly explained.

Task-Based Simulations

The initial conception of simulations was that a large, integrated work sample would provide good fidelity to the work environment. In practice, however, large simulations have significant practical limitations. If a portion of a simulation does not pass committee approval, the entire simulation is lost or delayed. Large simulations cannot be pre-tested which results in lengthy score reporting timelines. Accounting for prior

answers when scoring multiple questions in an integrated simulation can create scoring difficulties. Shorter, independent simulations would have fewer dependencies. They also appear likely to provide all of the benefits of skills assessment in the larger simulations while also providing more practical assessment.

Two of the factors that most influence our score reporting timelines are the inability to pre-test the simulations in their current format and the necessity of human scoring of essays for some candidates.

The concept of smaller simulations organized around a single task would allow the simulations to be pre-tested without an unfair burden being placed on the candidate. These shorter simulations would also be easier to score than current simulations because there would be fewer dependencies. Furthermore, pre-testing would introduce an important quality control allowing errors in content or keys to be identified at the pretest stage.

Written Communication Simulations in BEC

Currently, written communication is assessed in three sections of the Exam (Auditing, Regulation, and Financial Accounting and Reporting). Analysis indicates that individuals perform similarly on the written communication tasks regardless of section¹. This implies that the written communication skill is similar across areas. As a result, consideration can be given to administering the essay prompts all in one section. This would improve the stability of the written communications results. It would also make time available in the other three sections for administration of more task-based simulations.

Each candidate who now takes all four sections of the CPA Exam provides six writing samples, two each in AUD, FAR, and REG. Of these, three are scored and the other three are designated as pre-test questions. The scores on the writing samples are highly correlated across sections. However, within a section, we have only one measurement point related to writing and it contributes ten percent of the total score. If we concentrate the writing samples in one section, we will have multiple samples upon which to base the score, which is better from a measurement perspective. If we were to collect three writing samples, we could score two of them and use the third as a pre-test question.

The benefits of this approach are that the candidate responds to fewer prompts overall, a greater proportion of them contribute to the score, the total volume of samples to be scored is reduced, a more reliable score will result since it will be based on two samples instead of one, and the impact of the time it takes to score close to the cut samples by hand is eliminated from three of the four sections.

¹ Correlations among writing sample scores across sections are roughly the same as correlations of writing samples within sections. They are also higher than the correlations of the writing score with the score on the rest of the simulation.

The elimination of the writing sample from these three sections also provides more time for the administration of the task-based simulations and, as a result, better measurement of other skills. There would also be savings of staff time since fewer prompts would have to be developed and scored. There may also be cost reductions as the volume of essays to be scored declines. The drawbacks are that a greater proportion of one section is devoted to writing and that writing is slightly de-emphasized overall in the Exam.

Testing Written Communication in the Future

The issue of whether to test written communication by requiring candidates to produce a writing sample has always been controversial. A case can be made for moving to objective assessment² of written and, potentially, of other communication skills. However, that case needs to be presented to state boards and other stakeholders and vetted through them. Even if that is done, objective assessment of communication skills will take time to implement because an inventory will need to be built. In the meantime, staff believes we can improve on our current measurement of written communication and reduce its impact on the overall scoring timeline.

The CPA Exam currently has a direct assessment of written communication. That is, the candidate has to provide a sample of his or her writing. However, since the launch of CBT, a number of candidates have copied written material from the authoritative literature or from narrative in the simulation and submitted that as their own writing. Indirect measures of written communications skills such as multiple-choice editing tasks have been shown in many testing programs to be highly representative of actual writing skill, to provide more reliable results than actual essays, and to be inexpensive and faster to score³. This issue has been considered previously and state boards have been mixed in their preference. It might, however, be time to determine whether state boards' opinions have changed, particularly since some Subject Matter Experts working on the skills portion of the practice analysis have suggested that the ability to edit and revise previously written material is a more important entry-level skill than composition.

Other Simulations in BEC

Since the BEC section was new when CBT was launched, there was a lack of clarity about what skills assessment would be most appropriate for it. A number of potential simulation models were developed for BEC, but the uncertainty about the section prevented them from being implemented. Work has continued on this issue, but it remains open. The 2008 practice analysis will provide additional information about the need for simulations in BEC.

² Objective assessment does not necessarily mean multiple-choice. It means that human scoring is not required.

³ This topic was discussed in "A Perspective on the Current and Future Assessment of Written Communication Skills on the Uniform CPA Examination" presented to the BOE in 2004.

Introducing simulations in BEC by concentrating the writing samples in that section would preclude the introduction of other skills assessment in BEC, at least for the time being. Staff believes this is an acceptable trade-off given the customer satisfaction, measurement, and financial benefits of concentrating writing in the section. Staff would recommend re-opening consideration of simulations covering other skills in BEC in the future if objective assessment of communication skills is successfully introduced in the other sections and if stakeholders were amenable to have objective assessment replace the writing sample.

Assessing Additional Communication Skills

Communication skills such as speaking and listening are important. Depending on the results of the practice analysis, it may be appropriate to begin to assess communication skills other than written communication.

Testing Windows

The CPA Exam is administered two months each quarter. There is interest in having testing available year-round. While this would provide some modest increase in convenience for candidates, it could well have unintended negative consequences.

The removal of the dark month would remove one of the few remaining “deadlines” affecting candidate test taking behavior. It is well documented that people tend to schedule tests toward the end of each window. If there is no end, volume could well decline. Even if it doesn’t decline, spreading volume over three months instead of two will impact the speed with which pretest data can be analyzed and new questions inserted into operational use. Blocking files, which ensure that repeat candidates do not get the same panel during a subsequent test administration might need to be reconfigured. Since it is not possible to ensure that new content would reach all test centers on a given day, the back-end processing system would require modifications to allow for simultaneous administrations of multiple resource files.

Length of the Examination

When the BOE set the length of the CBT, it committed to reconsidering the time allotted to each section and the total length of the Exam when sufficient information was available to do so. The length was set at 14 hours conservatively to ensure that the Exam would exceed minimum standards of reliability. It does and our data show that the current Exam could be shortened and, while the reliability would decline, it would still meet minimum standards.

At this time, we do not know what changes will result from the practice analysis nor do we definitively know how long it will take candidates to respond to task-based simulations. Given these uncertainties, staff does not think a reduction in test length is warranted at this time. Staff does believe some modification in section length may be required if the written communication is concentrated in the BEC section.

Risk Factors Impacting Decisions

There are a number of factors to be considered when deciding on the features to be included in the evolution of the CPA Exam and on the timing of the introduction of those features. The more significant ones are listed in this section.

Ensuring stability in the system. CBT has a remarkable track record of exceedingly high quality. There have been no scoring errors involving multiple-choice questions. There are continuing reports of typographical errors. However, a complete review of a recent administration (07Q2) indicated that typos are rare and insignificant. In addition, there have been only three occasions in which groups of candidates were rescored and the scores resulted in changes in passing status. One was clearly candidate failure to respond as instructed, one was due to errors in scoring keys for “alternate” correct responses, and one was due to a software anomaly.

There was one incident in which tests were incorrectly administered to candidates and the candidates had to retest or wait several months for their scores. Other incidents have included errors in scheduling software and lengthy delays in the processing of candidate information.

These are examples of incidents that cause a disruption to the system. They are highly visible and create substantial inconvenience to candidates, state board staff, and state board members. Although they are infrequent and affect a very small percentage of candidates, their visibility and the effort required to respond to them is damaging to the perception of the CBT.

It is very important, therefore, that changes to the CPA Exam be done with the lowest possible risk to the stability of the system – both to the current system during development and to the operational system during and after deployment of the changes. It is also important that time be allocated to gathering usability reviews from candidates and others about new item formats, screen interfaces, etc.

Accepting multiple small risks or fewer larger risks. Each change carries risk. If changes are introduced sequentially, it will be easier to manage the risk associated with each individual change because it will be the only active risk. However, sequential change implies sequential risk so the period in which the risks pose an active threat to system stability will be longer. Sequential change also introduces the risk that a subsequent change may have unintended consequences that require a revision to a prior change. Introducing several changes concurrently raises the risk, but it is active for a shorter period of time.

Impact on legislation and rules. When significant change is introduced, it could affect state laws or rules (e.g., conditioning). Changing legislation or rules takes time and substantial effort. If multiple changes are required on multiple occasions, it increases the potential for delay or failure.

Informing candidates of changes. It has proven challenging to prepare candidates for change. Candidates often do not take seriously their responsibility to review program materials, tutorials, and practice tests. They rely more on word of mouth and information provided by review course providers. As a result, each time a change is introduced, candidate complaints increase and often there is a perception that there are errors in the exam. The amount of confusion caused will be less if changes are introduced sequentially, but, as with the risk exposure, overall the confusion will be of longer duration.

Each time change is introduced, it can require revisions to candidate materials. Depending on the change, revisions may be required to candidate bulletins, notices to schedule, the AICPA tutorial and practice test, and websites maintained by the AICPA, NASBA, state boards, and review course providers. Course materials used in preparatory reviews offered by review course providers and academic institutions also may require revision. The lead time for revisions to some of these publications may be as long as a year.

Potential impact on volume. Exam volume (both sections administered and number of candidates) experienced a substantial decline when CBT was introduced. Candidate volume has been steadily increasing, but has not yet fully returned to pre-CBT levels. Exam section volume has been rising steadily, but it is not clear whether the increases will continue.

To date, we have not seen significant fluctuations in volume when changes have been introduced (e.g., new software releases or increases in test fees), but the introduction of a substantial highly visible change could result in a volume decline.

Impact on supporting activities. Depending on the magnitude of the change, special studies may be required. The most obvious example is the potential need for new standard setting studies. Any change that results in a significant change to the structure or content of a section of the Exam will cause the results from the revised section to be incomparable to the results prior to the change. In some cases, the change may be small enough that a statistical adjustment can be applied to apply the old passing standard to the new test. If the change is large, a full new study may be required.

Full standard setting studies are expensive and time-consuming. If there are several changes that have the potential to require a full standard setting study, consideration should be given to implementing them concurrently.

Impact of decisions on current operations and systems. This paper focuses on enhancements to the Exam from an external perspective. Decisions about the enhancements described herein will require resources and modifications to computer systems and work processes. Not all of the required changes will be at the AICPA. State boards, NASBA, and Prometric may also have to make changes. The “invisible” impact of decisions needs to be considered in both choosing how to improve the Exam and timing the introduction of the changes.

Assumption

The project can be postponed if other priorities require the diversion of project resources. The project plan assumes stability in the system. If a destabilizing event occurs, the prudent course of action may be to postpone the project to address that event and to revise the project plan and timeline following resolution of the disruptive event. For example, we should carefully monitor the stance of the SEC with regard to accepting financial statements prepared under International Financial Reporting Standards (IFRS). Under the current model of harmonizing US and International standards, the impact of internationalization on the CPA Exam will be slow and gradual. Decisions that introduce rapid change with regard to international standards could require a redeployment of resources to modify Content Specification Outlines (CSOs) and development significant numbers of new test questions and simulations.