



# **How Is the CPA Exam Scored?**

**Prepared by the  
American Institute of Certified Public Accountants**

**Questions pertaining to this decision paper should be directed to Carie Chester, Office Administrator, Exams Team, at [cchester@aicpa.org](mailto:cchester@aicpa.org).**

**February 13, 2006**

## **How is the CPA Exam Scored?**

### **Introduction**

This document is intended to provide a non-technical overview of the CPA Examination scoring process. It begins with a high-level overview of the Exam and how it is scored. Following the overview are a series of questions and answers that provide more detail about specific aspects of the scoring process. The purpose of this format is to provide a general understanding of the scoring process and offer a more in-depth explanation of specific aspects of the scoring process to readers.

The information included in this document is based on questions frequently asked by state board representatives and other stakeholders. We recognize that this document will not answer all questions about the scoring of the CPA Exam and that new questions will likely arise as this document is read and discussed. We welcome additional questions and anticipate updating the document periodically based on the feedback we receive.

### **Description of the CPA Examination**

The CPA Examination is comprised of four sections: Auditing and Attestation (AUD), Business Environment and Concepts (BEC), Financial Accounting and Reporting (FAR), and Regulation (REG). All four sections contain multiple-choice questions. AUD, FAR, and REG sections also contain simulations.

The multiple-choice questions within each test section are administered to candidates in three groups (called “testlets”). Each testlet contains “operational” and “pretest” questions. The operational questions are the ones used to generate the candidates’ scores. Pretest questions are not scored; instead, candidates’ responses to these questions are used to evaluate the questions’ psychometric performance. Pretest questions that are psychometrically acceptable become operational questions on future exams. This strategy for pretesting questions is common in the credentialing field.

The AUD, FAR, and REG multiple-choice testlets vary in difficulty—there are two levels that are labeled “medium difficult” and “difficult.” These are simply labels. Within the testlets, items often vary substantially in their difficulty levels, but across testlets, those labeled “difficult” contain more hard questions than testlets labeled “medium difficult.” Every candidate receives a medium difficult testlet first. Candidates who do well on the first testlet receive a difficult second testlet. Otherwise, they receive a second medium difficult testlet. The scoring procedures take the difficulty of the testlet into account so that candidates are scored fairly regardless of the difficulty of the testlets they take.

The AUD, FAR, and REG sections contain two additional testlets containing simulations. Simulations are complex case studies that allow candidates to demonstrate their

knowledge and skills by generating responses to questions rather than simply selecting a correct answer. Simulations require candidates to use spreadsheets, conduct research, and generate written communications. Each simulation requires numerous responses from candidates (up to several dozen in some cases).

Before appearing on the CPA Examination, both operational and pretest questions have passed through several extensive and rigorous subject matter reviews to ensure that they are technically correct, have a single best or correct answer, are current, and measure entry level content as specified in the Content Specification Outlines (CSOs). The CSOs specify the percentage of each section that should be devoted to each content area. The current CSOs were adopted in 2002 based on the results of a practice analysis and state board responses to an exposure draft of the recommended CSOs; CSO references were updated in 2005. Operational questions have also been statistically evaluated to ensure they meet the psychometric requirements of the CPA Exam.

#### Overview of the CPA Exam by Section

Section	Audit	BEC	FAR	REG
Time	4.5 Hours	2.5 Hours	4 Hours	3 Hours
Multiple-Choice Testlets	3 (Difficulty varies)	3 (Difficulty does not vary)	3 (Difficulty varies)	3 (Difficulty varies)
Operational Questions per Testlet	25	25	25	20
Pretest Questions per Testlet	5	5	5	4
Simulations	2	0	2	2

#### Score Scale and Passing Score

CPA Exam scores are reported on a scale that runs from 0 to 99. The total score is not a percent correct score. It is a combination of scores from the multiple-choice, simulation, and written communication portions of the exam<sup>1</sup>. (Written communication exercises are included in the simulation testlets, but are scored separately). Scores on the multiple-choice and simulation portions of the exam are calculated using formulas that take into account not only whether the question was answered correctly or incorrectly, but also statistical characteristics of the questions themselves.

A total score of 75 is required to pass each section. There are no minimum scores required on the different kinds of questions (multiple-choice, simulation, written communication) or on different content areas within each section to earn a passing score.

<sup>1</sup> Since BEC does not have simulations, the BEC score is based solely on multiple-choice questions.

## Weighting the Kinds of Questions and Calculating the Final Score

The multiple-choice questions count for 70% of the total score. Simulations are 20% and written communication is 10% of the total score. Through a multi-step process, a separate score ranging from 0 to 100 is calculated for each type of question (multiple-choice, simulation, and written communication). These scores are multiplied by the weights (.70, .20, and .10), summed, and transformed to the CPA Examination Score Scale that has the passing score set at 75. (See Question 9 for a more detailed description of this process.)

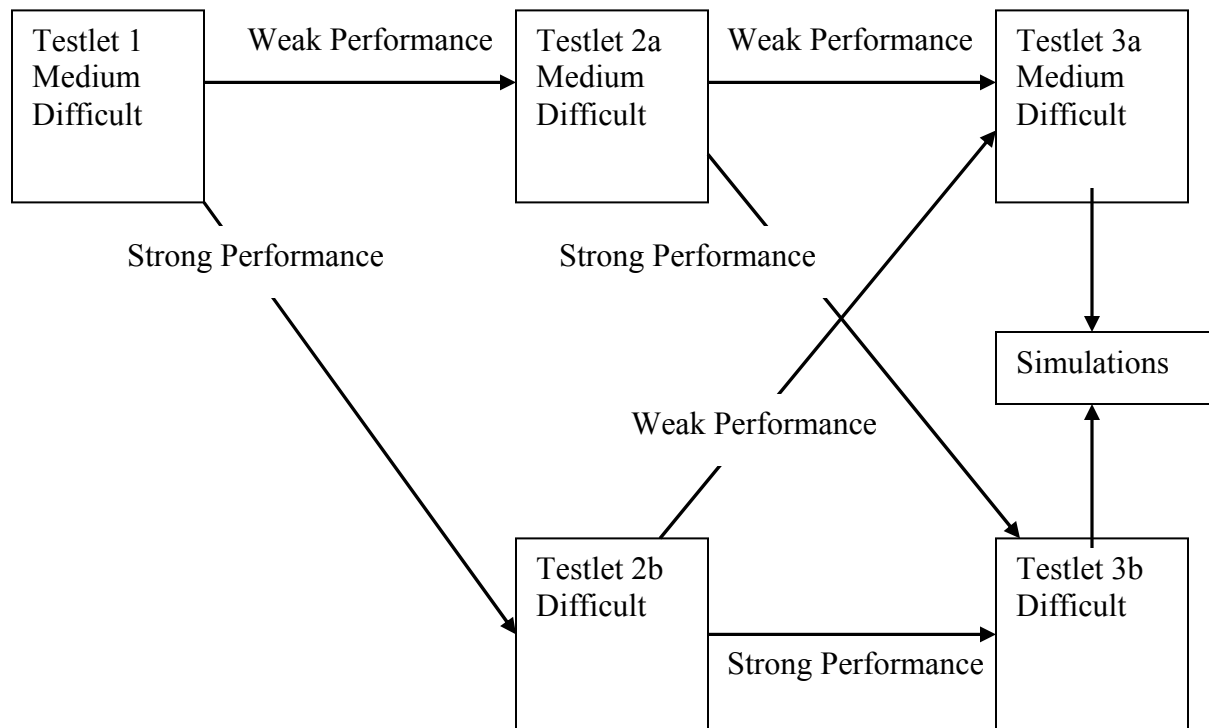
### CPA Exam Scoring: Questions and Answers

**1. If different candidates take different versions of the Exam are the scores comparable?** Yes. Scores from the different versions of the Exam are placed on a standard scale so they can be compared to each other. This process accounts for any differences in difficulty among the versions. All total scores are reported on the CPA Examination 0 to 99 scale<sup>2</sup>. The use of a standard reporting scale is common practice in the testing industry. You may be familiar, for example, with the 200 to 800 SAT scale or the 1 to 36 ACT scale.

**2. Are some versions harder than others?** Yes. Candidates take three multiple-choice testlets. The first testlet is always medium difficult. Candidates who do well on the first testlet will get a harder second testlet while those who do not do well on the first testlet will receive another medium difficult testlet. Similarly, the third testlet can be a medium or a difficult testlet and assignment of one of these to candidates is based upon performance on the two previously administered testlets. The diagram below shows how it works. Depending on their performance, candidates may receive (1) three medium difficult testlets, (2) two medium and one difficult testlets, or (3) one medium and two difficult testlets. This is called Multi-Stage Testing (MST) or Computer Assisted Sequential Testing (CAST).

---

<sup>2</sup> Scores from the current CPA Exam are NOT comparable to those from the previous, paper-based CPA Exam.



**3. Is Multi-Stage Testing fair? Why are you using it?** Yes, it is fair. Since the characteristics of the test questions are taken into account in the scoring, there is no advantage or disadvantage to being assigned testlets of different difficulty. The advantage of the design is that since candidates are seeing test questions that are matched to their proficiency levels, fewer questions are needed to obtain accurate estimates of candidate proficiency levels. The result is a better exam.

**4. How do you decide which questions are hard and which are moderate?** The difficulty of the test questions (and other statistics that are used to describe each test question) are determined through statistical analysis of candidate responses. Question difficulty is not a category (e.g. moderate or hard), but is a numeric value along a scale (e.g. 1.5). Testlets are classified as medium or difficult based on the average difficulty of the questions in them.

**5. Does that mean hard testlets can have easy questions and less difficult testlets can have hard ones?** Yes. All testlets have questions that range in difficulty. Hard testlets just have a greater proportion of difficult questions than moderate ones.

**6. What if a smart candidate does poorly on the first testlet? Can the candidate still pass?** Yes, the candidate can still pass the exam but the candidate will need to do better on the second and third testlets. We did an analysis of some candidates who did poorly on the first testlet, but whose performance improved on the second testlet. If we had

made a pass-fail decision about these candidates based solely on the multiple-choice questions about forty percent of them would have passed.

**7. Can a candidate get all medium difficult testlets and still pass?** Yes.

**8. Does this happen?** Yes, but not very often. It would require good but not excellent performance on the first two testlets, and then excellent performance on the last testlet.

**9. Can a candidate pass just by doing well on the multiple-choice questions?** No. The highest possible score a candidate could get from the multiple-choice questions is 70. A score of at least 75 is required to pass.

**10. What happens if candidates get a testlet that is too easy or too hard for them?**

They will probably receive the same total score they would have gotten if they had been assigned the right testlet. Every test score has some uncertainty or imprecision associated with it. The more questions someone answers and the more questions that are at the right level of difficulty for a particular candidate, the smaller the uncertainty (and greater the precision). The only shortcoming of receiving a testlet that is too easy or too hard is that the total score may not be quite as “precise” as the score obtained when the optimal testlets are administered to a candidate.

We did a study using generated data to model what would happen if people were assigned a completely inappropriate set of testlets. In the study, there were 1,000 “simulated candidates” at each of 5 proficiency levels ranging from very low to very high. Each simulated candidate was “administered” three moderate testlets (MMM) and a pass-fail decision was made. Each candidate was then administered one moderate and two hard testlets (MHH). The difference in the passing rates was negligible as shown in the following table.

Proficiency Group	Very Low	Low	Moderate	High	Very High
Difference in Pass Rates	0.1%	0.2%	1.4%	0.0%	0/0%

**11. Can I compute my score from the number of questions I answered correctly?**

No. A computer is required to score the CPA Examination accurately because the scoring takes into account the statistical characteristics of each question administered.

**12. What do you mean when you say “statistical characteristics”?** There are three statistics used to describe the questions: Difficulty – whether the question is easier or harder for candidates, Discrimination – how well the question differentiates between more able and less able candidates, and Guessing – the likelihood of candidates answering the question correctly just by guessing. The statistics for the multiple-choice questions are generated when the questions are administered as pretest questions and used

in the scoring when the questions are operational. The statistics for simulations are calculated the first time they are used. The formulas for generating the statistics and scoring the exam come from modern test theory, sometimes called “Item Response Theory.” Item response theory is being used or has been adopted by nearly all of the large licensing examination programs in this country and by many of the moderate-sized and smaller examination programs too.

### **13. What is Item Response Theory (IRT)?**

The term is simply the name given by Dr. Frederic Lord in 1980 to a class of psychometric models for exam development and analysis that report items and candidates on the same scale. When the psychometric model fits the data, as it does with our exams, it becomes straightforward to be very efficient in the design of exams, to obtain estimates of error in candidate scores, and to compare candidate scores easily and quickly when they are based on exams that differ in difficulty.

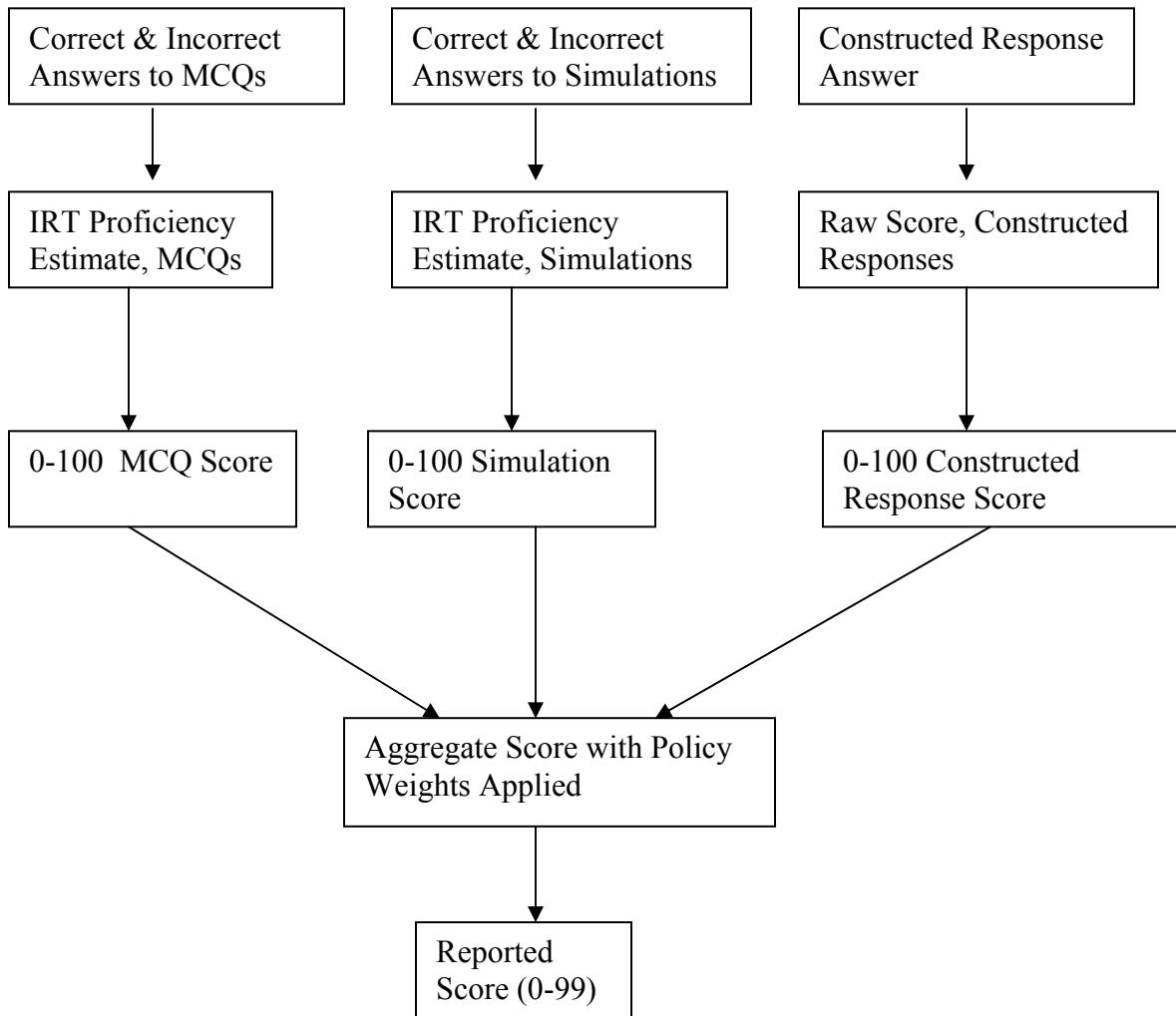
**14. Does this mean two candidates could answer the same number of questions correctly, but get different scores?** Yes, since candidate scores depend on the characteristics of the questions, not just how many they got right.

**15. In college some of my professors gave tests that had questions that were worth one point and others that were worth two points. If one student got five of the one-point questions right, he got five points. If another student answered five two-point questions right, he got 10 points. Is that what you're doing?** Conceptually, yes. But the professor assigned the weights based on judgment. In the CPA Exam, the weights are determined from candidate response data using item response theory.

**16. I understand that item response theory says the scores are comparable, but do you have any evidence that scores from tests that vary in difficulty can really be comparable?** Yes. It is possible to obtain an estimate of each candidate’s proficiency on each testlet. Although these estimates are not reliable or consistent enough to use for reporting scores, they do provide useful information. We compared proficiency estimates from each of the three multiple choice testlets for a sample of about 2,000 candidates. The proficiency estimates were comparable in 85 to 90% of the comparisons, regardless of whether the testlets were of the same or different difficulty. When the three testlets are combined, the reliability and accuracy of the exam increases considerably. The same finding has been reported by many testing agencies which is why item response theory models are so widely used.

**17. In general terms, what are the steps taken to produce the reported score?** For purposes of score reporting, each component (multiple-choice, constructed response, and the remainder of the simulations) is initially treated as a separate test. For the multiple-choice and simulation components item response theory is used to determine a proficiency estimate for each type of question. The multiple-choice estimate is then mapped to a score between 0 and 100 on a multiple-choice scale. Similarly, the simulation estimate is mapped to a score between 0 and 100 on a simulation scale. For

the constructed response questions, the grade assigned by a grader (0-4) is multiplied by 25 to put it on a 0 to 100 scale. The three scores are then combined with the policy weights (70% multiple-choice, 20% simulations, 10% written communication) to create an aggregate score. The final step involves mapping the aggregate score to the 0 to 99 scale used for score reporting. The process can be viewed in the schematic below.



**18. Who are the subject matter experts who do the question reviews?** They are CPAs with expertise in the area being tested who volunteer their time. There are also subject matter experts on the Exams staff at the AICPA.

**19. How was the passing score set?** Volunteer CPAs participated in a standard setting study. They reviewed test questions and how candidates performed on those questions in order to make judgments of what test performance was required to protect the public interest. The results of this study were used by the Board of Examiners as a guide when it

established the passing score. The passing score chosen by the BOE was then mapped to a score of 75 on the scale used to report scores to candidates.

**20. Is there more that I can read about the Exam?** Yes. There are many technical reports related to the psychometric characteristics of the Exam. All of these reports will be posted to the Psychometrics section of the Exam website, [www.cpa-exam.org](http://www.cpa-exam.org), soon. Other useful publications can already be found under the Research and Reports section of the site. In the interim, copies may be obtained by request from Carie Chester at [cchester@aicpa.org](mailto:cchester@aicpa.org). If you have additional questions about the scoring of the CPA Exam, please send them to Carie Chester at the e-mail listed above.