AICPA

AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS

Technical Report

SERIES TWO

Considerations in the Development of Accounting Simulations

New Orleans, LA – April 2002

Manuscript for presentation at NCME

Richard N. DeVore, Ed.D
Simulations Development Leader, AICPA

May 2004 Number 13

Technical Report

SERIES TWO

Considerations in the Development of Accounting Simulations

New Orleans, LA – April 2002

Manuscript for presentation at NCME

Richard N. DeVore, Ed.D
Simulations Development Leader, AICPA

May 2004 Number 13

Introduction

In many quarters, measurement researchers and test publication companies are experimenting with simulations as a way to assess complex behaviors or skills that cannot be measured well through multiple choice questioning. In a few fields, the advent of computerized test delivery and the rapid advancement of computer technology have made possible the use of performance simulation testing in large-scale, high-stakes testing programs. Already licensing examinations in medicine, architecture, and other fields make use of simulations of practice in an effort to better assess candidate competencies and skills.

The AICPA plans to launch a new computer-delivered Uniform Certified Public Accountant examination that will use the capabilities of the computer to allow improved measurement of skills and knowledge through the use of simulations of accounting practice. This effort was initiated with attention to lessons learned in other large-scale high-stakes development efforts and with the intent to implement the program within tight budget and time constraints. In the development process, the AICPA has had to, or will have to, consider several issues that can derail efforts to include simulations testing. This paper discusses several of these issues and mitigation strategies or processes the AICPA has adopted to deal with them.

Required Candidate Preparation

In the computer-delivered examination, candidates must cope with navigation rules and controls imposed by the computer interface and testing network requirements. In addition, most simulations require candidates to employ tools with which they may be familiar, but which may have unfamiliar interfaces in the testing environment. Limitations of computerized delivery may also require the recording of answers in an unusual fashion. These conditions can be a source of error in test score, and candidates must be prepared in advance to deal with such idiosyncrasies in order to minimize their impact.

One way to reduce the impact of computer delivery is to utilize an interface with which most of the candidates are familiar and which is reasonably intuitive, even for those unfamiliar with it. In the computer world, this means using general controls common to both the Windows and MacIntosh interfaces and employing metaphors in the simulations with which all candidates are likely to be familiar. The AICPA approach is to use common controls such as copy and paste, drag and drop, and point and click for most

of the response types, and to utilize a tab style navigation system that resembles paper filing systems familiar to most accountants (and which is an increasingly common metaphor in both Windows and MacIntosh applications). Training in the use of this metaphor and these controls is easily accomplished through a tutorial that closely resembles an accounting simulation and gives candidates the opportunity to practice with and understand the interface before attempting to test with it.

In the accounting profession, the tools of the trade include the spreadsheet, the word processor, and authoritative standards and literature. Most of world uses commercial spreadsheets and word processors with common features familiar to candidates and CPAs. Since licensing issues prohibit distribution of the familiar products in candidate preparation products, the AICPA has chosen to license rights to programmable controls that can replicate the important functions and the look and feel of the commercial products for a one-time cost. In addition to overcoming the familiarity issue posed by the need for these tools, this solution permits the AICPA programmers to build into the products the features necessary to enable the capture of data during operational testing. This eliminates the need for a post-processing data extraction step that might have been required with the use of a commercial product with protected code.

CPAs use authoritative literature to access the standards to which accounting and auditing must adhere and to justify their conclusions and actions. Likewise, the taxation specialist must refer to IRS code, regulations, and judgments in order to perform competently. Most CPAs use commercially available searchable electronic databases for these purposes. The magnitude of these databases makes replication a practical impossibility, so the AICPA chose to approach this problem with a quid pro quo approach – use of the commercial product in a secure non-distributed format in trade for recognition of the product in the examination and preparatory materials. The vendor trades the use of the product for the potential of future sales and through AICPA-enabled access to the target population in which the vendor is interested.

In order to use the commercially available searchable database products on the operational test, candidates must all be familiar with their functionality. The solution that protects the interests of the database vendor is to provide for the candidate the search engine utilized by the commercial product, but to limit the database to a small subset of the that found in the real product. This will allow candidates to familiarize themselves with the search capabilities without requiring the distribution of the commercial product. Some cost may be involved in creating a limited-use, derivative product that can be used for candidate preparation and training, but these costs are anticipated to be relatively low.

Limited Measurement Opportunities

The primary purpose of most tests is to gather data that will permit inferences to be drawn about a candidate's abilities from his or her performance. Simulations often require the candidate to spend more time reading and digesting information as background for the tasks than that which is required for multiple-choice items. This can result in a low data-collection-to-time-spent ratio, and, in a timed test, it can lead to less measurement and lowered reliabilities.

To effectively use simulations for measurement in a high stakes test, the test developer must allow for a significant number of measurement opportunities in the time period allotted for the simulation. This objective calls for careful attention to the design of the simulations at the outset. Topics chosen to be addressed by the simulations must be complex enough require a number of meaningful sub steps requiring different knowledge and skills in the solution of a particular task. Finding these opportunities is not as difficult as it sounds, for many tasks require a significant amount of knowledge or skill to arrive at a result. Tests usually focus on the final outcome in such tasks and mastery of the subtasks is assumed for the successful candidate. Simulations provide an opportunity to examine candidate performance on the specific subtasks and to provide partial credit for imperfect candidate performance on the main task. Such measurement provides increased measurement per unit time, and potentially greater score accuracy.

Although simulations can be designed to provide an abundance of data about candidate performance in a particular area, the test developer must be careful that candidate performance on one task or subtask is not contingent on previous performance on another task. Such contingencies are unfair to candidates and may lead to spurious inferences about their abilities. The AICPA solution to this potential problem lies in the scoring methodology. First, tolerances must be built into scoring plans so that rounding errors and slight variances from intended results will not have a negative impact on a candidate's score. More importantly, when a response generated by the candidate as the answer to one item is used as the basis for the next item, subsequent scoring is algorithmic rather than absolute. In other words, each answer that relies on data generated in a previous item is scored as if the first item were correct, whether it is or not. AICPA scoring systems are being developed to allow for full credit for a candidate calculation irrespective of the number the candidate uses to start the calculation. Through this scoring methodology, measurement opportunities can be increased in a continuing simulation while providing independent data points from which inferences about candidate knowledge can reasonably and reliably be made.

Extended Time Requirement

This issue is a corollary of the previous one. Since most simulations require time for the candidate to read and digest data, and measurement standards require high reliability for high-stakes tests, the inclusion of simulations usually means that testing time must be extended. Longer testing time, in turn, can mean candidate fatigue. In the computer-delivered world, this also means extended seat time in the testing center and therefore higher cost to the candidate.

The solution to this problem lies again in maximizing the measurement opportunities related to a particular simulation and in condensing the information necessary for the entire simulation into a single presentation that can be easily read and digested. A format that allows for quick random access of data, as tasks are being completed, is desirable.

As in any assessment, the way to minimize reading load is to reduce or eliminate any extraneous information provided to the candidate. This can be done by ensuring that every piece of information in the stimulus for a simulation supports the solution to a problem, or provides intended distraction for the unprepared candidate. A second, equally important way to reduce time requirements is to provide the information in a format expected by and familiar to the candidate. In the world of accountancy, this means providing data in commonly encountered formats such as in spreadsheets, balance sheets, cash flow statements, and income statements. The AICPA simulations rely heavily on tabulated data for both of the reasons cited above – the format is familiar to the candidate and random access of the data is facilitated as the candidate undertakes the tasks.

One final way to minimize the time required in a simulation is to limit the amount of additional information related to a specific task provided to the candidate during the course of the simulation. In other words, the stimulus for the simulation should include all the information necessary to address each of the subsequent tasks. Adding additional information subsequent to the original offering causes to the candidate to reread and reevaluate the original information (and sometimes earlier tasks) in light of the new information, and that process can require valuable time, warranted or not.

Development costs

Well-crafted simulations require more effort than most multiple-choice items with respect to both creation and authoring for delivery and scoring. Development costs can limit the practicality of using simulations testing if means for efficiently mass producing the simulations cannot be found.

Good simulations have several inherent characteristics – they involve a series of realistic tasks that can be scored independently, they address several content and skill areas appropriate for the target population with little repetition, and they can be successfully completed by adequately prepared candidates in the time slot allowed for them. Clearly, the creation of simulations that simultaneously meet all these criteria takes a good deal of time on the part of both content and measurement experts. Development costs can quickly mount in a high-stakes effort that must have large numbers of simulations in order to limit security risks.

The AICPA has adopted a template model for development of simulations that has been successful in producing large numbers of parallel simulations that exhibit the characteristics noted above. Under such a model, much of the work necessary to ensure that the simulations will meet these criteria is done early in the program, before significant numbers of simulations are created. Prototypes or model simulations are created, carefully constructed, tested, and refined by committees of content experts and measurement consultants.

When the models are ready and approved by those concerned, they are converted to templates to be used to create additional simulations that resemble the original models. This is done by identifying the variables in the original simulation that, when changed, will create a completely different appearance and a completely different answer set. These variables are specified and given acceptable ranges in an attempt to control the difficulty of the simulations that result from changes. In the world of accounting, variables that change the appearance of simulations include business names, business categories, and reporting dates. Variables that change the answer sets include accounts, items of interest within accounts, dollar values, accounting treatments, and dates of acquisition or sale of items.

Once the models have been chosen and the templates constructed, item writers are given assignments to fit content to the templates rather than to create wholly new simulations. This makes their task a good deal easier than that of model creation and allows them to turn out a number of simulations in a relatively short period of time. Because the new simulations are inherently like the models, their chances of surviving critical content review are greatly increased, they possess the desired characteristics identified above, and their measurement properties can be estimated with reasonable confidence. In addition, because the item writers are guided through the process by using the templates and because the method requires relatively little time commitment on the part of content experts, it is possible to enlist item writers as volunteers, thereby reducing the item development costs.

The template methodology also serves to lower authoring costs to the program. Because new simulations closely resemble the models, it is possible for the test developer to author for presentation

and scoring by changing only superficial elements in the simulation. It is possible, in fact, to create close derivatives of extant simulations (those that differ only by names and numbers) by editing a list a variables and allowing software to perform a type of global search and replace in both presentation and scoring files.

Item Comparability

Since simulation difficulty depends to a significant extent on the context of the case or situation cited in the simulation, the concern arises of how to make testing experiences for different candidates comparable. Because of the time required for a candidate to respond to a simulation, simple pretesting plans may not be feasible.

Development of simulations for a profession requires the test developer to use the content expertise of members of the profession. Simply providing content specifications and examples of previous offerings to such experts engaged to write additional simulations will result simulations varying widely in requirements (both content and timing) and difficulty. To control this phenomenon and to maximize the probability that two candidates will have reasonably equivalent testing experiences, the AICPA will utilize limited pretesting and the template development methodology described above.

As indicated above, the template development model controls the variables that have an impact on item difficulty. Limitations imposed on the manipulation of these variable helps to maximize the probability that simulations created through this approach will have difficulty and discrimination characteristics similar to those of the model from which they are derived. Once the item parameters are established for the model simulations, those for the derivatives from these model can be attained with reduced pretesting by utilizing expected difficulty and discrimination as prior information in the calibration process. Through pretesting that will be undertaken prior to program launch and ongoing pretesting that will continue during the operational program, the AICPA expects to be able to gather sufficient data on the simulations to permit calibration of the items and preequating of the simulations. This, in turn, will permit the assembly of equivalent forms of the measures, even when the difficulty of the simulations varies.

Security Risks

By their very nature, simulations are more memorable than discrete multiple-choice items. To ensure adequate security for the bank of simulations and thus provide comparable testing experiences for all candidates, the number of simulations available for random administration or for repeating candidates must be large. Once again, the template model for creation of the simulations provides a solution.

In the plan for simulation production for the AICPA program, each simulation writer will be asked to write two simulations using two simulation templates and exemplary models. The writer does not have to conjure up a new simulation, but rather has to fit content to the extant model, adhering to the limitations on variables imposed by the template. Most writers find this task readily doable and can create a derivative of the original model in a few hours. Writers create situations that involve different initial circumstances with different companies and which involve different accounts. To the candidate, these problems look entirely different, but the tasks that the candidate must perform do not change and the content and skills tested remain the same.

When AICPA receives the writers' simulations, the content experts on staff create derivatives of this work by varying names, dates, numbers, and, therefore, the solution sets. These superficial changes provide us with double the number of new simulations at very little cost. In this fashion, the numbers of operational simulations grows rapidly and the probability of one candidate conveying useful information to a collaborating cheater decreases correspondingly.

The AICPA will also employ a blocking procedure for repeat candidates that will enhance security. If one thinks of the original model simulation as a first generation, the writers efforts based on a template as the second generation, and the derivatives of writers' simulations created by AICPA staff as the third generation of a simulation, the blocking will be done at the second generation level. Thus the probability of a returning candidate receiving a second simulation based on the same business, concerned with the same accounts, and having the same solution set as the first simulation he or she received is nil.

Conclusion

The issues in the paragraphs above represent several of the major concerns any testing program will need to consider when attempting to include simulations as part of the examination experience. The AICPA is dealing with these and other concerns by building on lessons learned in former development efforts in other fields. It is hoped that the presentation of these concerns, and of AICPA's proposed solutions, will stimulate further research and discussion that will help other organizations to surmount or avoid them in their development efforts.