

# ALTERNATIVES FOR OFFICE AUTOMATION TESTING

## Purpose

This document is designed to provide SkillCheck customers, sales staff, distributors and other interested parties with a detailed understanding of alternatives available for automated testing on important skills such as usage of Microsoft Office desktop applications.

## Definitions

Important terms used in this document include:

<b>TERM</b>	<b>DEFINITION</b>
<b>Knowledge-Based Testing</b>	Tests consisting of traditional question types such as multiple-choice, fill-in-the-blank, true-false and matching.
<b>Performance-Based Testing</b>	For purposes of this document, “performance-based” is synonymous with <b>interactive simulation</b> , i.e. test items that simulate the behavior of a particular software product and ask the user to perform a specified function correctly within that simulation.
<b>Concurrent Testing</b>	Testing that utilizes the actual applications for delivering question items. As noted elsewhere in this document, concurrent testing can be delivered in a number of different ways.
<b>Validity</b>	Proof (often statistical) that a test accurately measures the skill or set of skills it is intended to gauge. Methods of assessing test validity include <b>content</b> , <b>construct</b> and <b>criterion</b> validation.
<b>Content Validity</b>	Proof, normally provided by subject-matter experts, that items in a test cover the most important and frequently-used knowledge, skills and abilities needed to accomplish the job being measured by the test. For example, a test on Microsoft Word can be said to be content valid if it can be proven that the subjects covered in the test (setting margins, printing, etc.) represent the most important and frequently used functions critical to the success of a Microsoft Word user.

TERM	DEFINITION
<b>Construct Validity</b>	Proof that the individual items in a test are accurate measurements of the subject being tested. For example, an item that asks an applicant to set margins within a performance-based simulation of Word or a concurrent test that uses the real Word program use a near-perfect construct to measure someone’s ability to perform a task (in this case, set margins). A multiple-choice question that asks: “Which menu contains an option to change margins?” is a less exact construct for measuring the same function.
<b>Criterion Validity</b>	Proof that the overall test accurately correlates with some other independent measure. For example, assessing a test taker’s on-the-job performance with Word one month after they have taken a Word test and seeing if test scores accurately predicted job performance is one type of criterion validation method known as a <b>predictive</b> study. Alternatively, a test can be administered to people whose job skill is already known and test scores can be correlated with an independent measure of those skills. This type of validation is known as a <b>concurrent</b> study. In both cases of criterion validation (predictive and concurrent) it is important that independent assessments be generated blind to test scores.

## **KNOWLEDGE-BASED TESTING**

### **Advantages of Knowledge-Based Testing**

Because of the popularity of knowledge-based testing in traditional education and professional certification, most test takers are intimate with the linear question types used in knowledge-based tests (multiple-choice, true-false, etc.). This reduces lack of understanding of test methodology as a factor that can unfairly affect test scores.

In addition, graphical computer environments such as Microsoft Windows, which allow one to mix text, graphics and multimedia elements on the screen, help to broaden the types of subjects that can be covered by linear items. For example, knowledge-based questions on Microsoft Word can ask a test taker to click on specified parts on the screen, rather than select from text-based multiple-choice options, to assess their understanding of application screen elements.

Also, knowledge-based questions are relatively easy to automate and many tools and authoring systems are available to both automate knowledge-based test questions and manage the administration of tests and the analysis of scores. The “one-screen-per-question” simplicity of automated linear questions also makes it easier to deploy tests containing these question types over the Internet.

### **Disadvantages of Knowledge-Based Testing**

While easy to automate, it is often difficult to come up with linear items that (1) accurately reflect the skill being measured and (2) provide appropriate “wrong” answers (called “distracters”) that are not so subtle that they cause confusion or so clearly wrong that the correct answer becomes obvious.

In the case of testing on computer applications such as Microsoft Word, this problem is particularly acute, especially since much of the employment, human resources and training industries have become used to using more extensive (and more construct valid) performance-based or concurrent testing to measure performance on the same subject matter. Given that interactive simulation has been the standard in the employment industry for the last ten years, knowledge-based testing on these types of applications are considered “old fashioned” by today’s standards and are not frequently used.

## PERFORMANCE-BASED TESTING

### **Advantages of Performance-Based Testing**

Most performance-based testing utilizes similar tools to those used to create computer-based training (CBT). Unlike CBT, which moves a user linearly through a course of study, performance-based testing places the user in a simulation of a computer application (or other job task) and asks the test taker to perform a function as if they were using the real software. Their ability to perform the function correctly (using any correct method the software allows or using a specified method) determines whether they get the item correct or incorrect.

The key advantages of performance-based testing include:

- (1) The ability to simulate (and thus test on) any computer application without requiring the application to have special programming “hooks” or, indeed, with out the real application being present on the testing workstation at all.
- (2) Full control over the testing environment that allows the performance-based test to accurately determine whether a test taker has performed the specified task correctly.
- (3) Since any feature of a software product can be simulated (even complicated features such as asking a test taker to use multiple Microsoft Office applications simultaneously), performance-based testing allows you to ask straightforward questions about any product feature being tested.

On the technical side, performance-based testing also has the advantage over concurrent testing in that most performance-based systems are self-contained programs. As such, they do not have to control other products in order to operate, thus eliminating considerable variation between different operating systems, product versions and languages.

## **Disadvantages of Performance Based Testing**

Two problem areas that have been solved with SkillCheck's Professional Plus product include (1) the quality of the simulation and (2) the inflexibility of traditional performance-based testing environment.

### **Quality of Simulation**

A performance-based item is only as good as the quality and depth of the simulation, and older products often did not go into enough depth to cover a feature or did not provide test takers with the ability to answer a question using any method the software allows. This has changed markedly over the last several years as new authoring systems allow simulation of more application features, and as test developers put more effort into creating near-perfect simulations of software functionality. The SkillCheck Professional product asks test takers to finish a complete software task in order to score an item as correct. In Microsoft Word, for example, there are over 75 different pathways one can take to open a document. A SkillCheck question covering opening Word documents allows every one of these methods.

### **Need for flexibility**

Concerning issues of flexibility, the difficulty of programming performance-based questions meant that most vendors of such products would only create single tests of 30-40 questions that were meant to be used in all testing situations. The SkillCheck Professional Plus product has eliminated that constraint by creating over one-hundred performance-based questions on each software product and providing tools that allow test developers to create tests that can meet the exact needs of their certification, applicant screening or training programs.

### **Remaining problems**

While software simulation has improved markedly over the last several years, there are still several elements to a program that have not been and may never be made part of a performance-based simulation. Access to the complete product's online help system while running a simulation is one example of a feature that is normally not part of a performance-based test.

Also, since performance-based simulations only program correct pathways, marking any step taken outside those pathways as wrong, performance-based tests can seem "unforgiving" of stray mouse clicks or keyboard strikes. With most performance-based products,

this problem is minimized by providing a test taker with an adequate tutorial explaining the nature of the test and providing them feedback as to whether they got the question correct or incorrect.

It should be noted that, despite its drawbacks, performance-based testing has become the industry standard in temporary employment, an industry with a long history of using this type of testing in a high volume, sophisticated manner.

## CONCURRENT TESTING

### **Advantages of Concurrent Testing**

The obvious advantage of concurrent testing over both knowledge-based and performance-based testing is that concurrent testing places the test taker in the real computer application and gives them access to all of the application's features, including online help and other resources.

In addition, a concurrent system can speed the development of multi-lingual testing since the delivery platform is the actual live software which may already be available in multiple languages.

That said, how a concurrent test examines a test takers knowledge of product features varies considerably between different concurrent products.

Products such as PSI's Tapdance ask a test taker to recreate a document that was created in an application such as Microsoft Word. When that work is complete, the file is "printed" to an encoded file and compared to a "master" to see if the final state of the document is correct.

Ziff Davis has introduced a product that asks knowledge-based questions which can only be answered by correct usage of the live application. For example, a fill-in-the-blank question sits on top of an Excel spreadsheet asking the test taker to provide the word or number in a specified cell. Only by correctly going to that cell can they answer the question (thus providing a test of the user's knowledge of spreadsheet navigation).

Other products have used CBT "hooks" available within applications such as Microsoft Word to analyze whether a test taker is performing specified tasks correctly.

**Disadvantages of Concurrent Testing**

Despite placing people in the real world environment of the real software, the testing method chosen within that environment can limit the types of questions being asked.

For example, checking the final state of a document (as in the Tapdance product) can be used to check basic formatting and editing features that lead to a result in the final document. However, they cannot be used to test features that do not leave such a result in the final file (features such as Go To and other navigation controls). In addition, checking the state of very complicated documents that include tables, graphics or compound document elements from different applications is increasingly complicated. Because of this, it is difficult to use this strategy to do much more than test proficiency and accuracy in certain types of applications (notably word processors). The system tends to break down for other types of products such as spreadsheets, presentation programs and databases.

The Ziff Davis strategy of asking knowledge-based questions about live, running applications carries the same baggage as any knowledge-based test: assessing the knowledge of a feature using an indirect question, rather than a direct requirement to use a specific feature correctly. For example, a fill-in-the-blank question that requires someone moves to a certain cell cannot analyze how a user has performed this navigation function to answer the question. If the question is testing general knowledge of spreadsheet navigation, this is not a problem. However, if one wants to test specific knowledge of Excel's Go To feature, such a question is inadequate.

Using application "hooks" to monitor a test taker's progress in performing a task would give a concurrent test the same benefits of a performance-based test with the added advantage of giving the test taker access to the full range of product features in order to answer a question. Most of the limitations in this area, however, have proven to be technical in nature. Reading the Windows "journal" of all activity taken by a test taker to determine if they did the right thing is theoretically possible, but if a user can perform any number of steps before answering the question (access help, try different commands to search through dialog boxes, etc.), the correct answer can get lost in the "noise" of other activity. In some cases, this can lead to an answer being marked as incorrect, even if the user has ultimately performed the indicated task correctly. (Such problems are compounded if the testing product does not provide the user any sort of feedback while they are answering questions.)

While strategies can be developed to overcome technical limitations of concurrent testing, or different concurrent methods can be combined to assess performance in a real software product, as of now, no concurrent

product has proven to be adept enough to accurately test as many features as can be tested using performance-based methods.

Even if these technical hurdles are overcome (as they could be with new versions of software such as the Microsoft Office applications which contain more sophisticated training “hooks”), concurrent testing will always require multiple applications to work together correctly in a variety of environments. While a self-contained, performance-based system needs to be tested to ensure it works with different operating systems and networks, once such a system has been tested for one simulation product, it will work for all such products.

Concurrent products, on the other hand, are dependent on:

- **Product** -- Word, Excel, PowerPoint, Access and other applications are likely to all provide “hooks” in a different manner.
- **Product Version** -- File formats and “hooks” used for concurrent assessment are likely to vary between different versions of the same product, and may not be available at all for certain versions of the software.
- **Operating System** – Since linkages between an assessment program and the live applications are likely to make use of system files and controls, these can vary considerably on different versions of the operating system
- **Network** – A concurrent product must maintain its communication with the live application in different network environments, thus combining all of the problems associated with running the live application on the network with its own network issues.
- **Language** – At least up through Windows 95 and Office 97, system commands and features of different language versions of Microsoft products (e.g. macro language, decimals vs. commas for numbers) vary between different language versions of the products. This adds one more layer of complexity to a concurrent product that may already require extensive testing in a wide variety of different environments.

At the very least, concurrent testing will likely require specific drivers or other software be developed for each different product (Word, Excel, PowerPoint, etc.). It is likely that similar variation will be required for different product versions, and possibly different language versions as well.

**Conclusion** Each of the three types of testing system for office automation products (knowledge-based, performance-based and concurrent) offer specific advantages and disadvantages and it is likely that a mix of solutions will co-exist on the market for some time to come.

The industry most intimately involved with this type of testing, temporary staffing services, has chosen to standardize around performance-based testing. This is because performance-based testing provides an excellent construct for testing the ability to use computer applications without the expense and technical overhead of maintaining a concurrent system.

Interestingly, current concurrent systems designed to measure user ability using the live software have had to take less direct approaches to assess software skills than comparable performance based systems. Even as new concurrent systems become available which overcome this hurdle, concurrent testing is likely to face a number of technical challenges that may limit its availability to specific versions of specific products.

**About  
SkillCheck**

SkillCheck, Inc. is a leading supplier of testing software for the temporary employment, staffing services, human resource and training industries. SkillCheck's Professional Plus product is used by the world's largest temporary services and is available in multiple languages (French, German, Spanish, Dutch, Italian and Norwegian).

SkillCheck's products combine the quality and construct-validity of state-of-the-art performance-based test questions with the flexibility to create tests from a database of over one-hundred questions per software product. This unique feature allows SkillCheck users to create tests that can exactly meet an organization's certification guidelines, applicant screening procedures and training curricula.

SkillCheck testing products are also available for skills such as accounting, computer programming and call center skills. Contact SkillCheck at 1-800-648-3166 for more information.