



# **Unproctored Internet Testing at PreVisor**

*Jim Beaty, Ph.D.  
Chief Scientist, PreVisor*

**Table of Contents**

What is Unproctored Internet Testing? .....	p. 3
Why Use Unproctored Internet Testing?.....	p. 3
Risks Associated with UIT .....	p. 4
Does Unproctored Testing Work?.....	p. 4
Do UIT Scores Rise Over Time? .....	p. 5
Do Unproctored Tests Demonstrate Validity for Predicting Job Performance? .....	p. 6
Does UIT Reduce the Diversity of Applicant Pools?.....	p. 6
UIT Best Practices.....	p. 7
Ongoing PreVisor Unproctored Testing Research.....	p. 8
Summary .....	p. 8
References .....	p. 9

## What is Unproctored Internet Testing?

Unproctored Internet testing (UIT) is the administration of Internet-based tests to individuals outside a traditional test proctoring setting like a testing center. An unproctored Internet test could be completed by applicants in their homes or in libraries – literally anywhere they can access the Internet.

**There are at least three types of UIT in personnel selection: Open, Invitation and Supervised<sup>1</sup>.**

### Open

In the Open mode of UIT there is no human supervision of the testing process and the testing session is available to applicants with little or no registration required. An example of this mode would be having applicants access a testing session through a job board without requiring them to first register through an applicant tracking system (ATS).

*An important critique of the open mode is that it provides too much access to the testing process. Applicants can enter the testing process at-will and even practice taking the test until they pass. This approach results in unclean and dubious test data.*

### Invitation

UIT in Invitation mode has the potential to yield better test data. In the invitation mode, applicant information is captured and reviewed by either automated or human processes, or both. This review of applicant information is the basis for determining whether applicants are granted access to the testing session. The requirements to pass this step in the process can be as simple as completing the required fields on an application to as in-depth as meeting minimum qualifications, passing a phone screen with a recruiter or attending an on-site recruiting event. If applicants successfully complete this step they are invited to complete the testing session.

*Two important points about UIT invitation mode: First, there is no human supervision of the testing process, but applicants may interact with a person as part of the recruitment and screening process. Second, although invitation mode provides more information about applicants than open mode, the identity of the applicants who actually take the test is still unknown because the test is administered with no human supervision.*

Human supervision of the testing process is a key factor that distinguishes proctored from unproctored testing. A fully proctored testing session involves human

verification of applicant identity and human supervision of applicants completing the tests. The advent of online testing has created a testing scenario that involves human interaction, but is not fully proctored.

### Supervised

In the third mode of UIT, Supervised, applicant identity is verified by humans but then test takers complete online tests without being supervised. User-friendly testing applications with sample test sessions and timers built into the tests allow applicants to navigate through the entire testing process without the aid of a test proctor.

*Supervised testing typically takes place when a company partners with a vendor that provides identity verification and online/computer resources without proctoring the actual testing session.*

## Why Use Unproctored Internet Testing?

The three modes described in the previous section broadly capture the type of UIT that occurs today. These categories differ from proctored testing where trained human proctors monitor applicants, sometimes individually, through the entire testing process. One of the key points often made by HR leaders, line-managers and executives is that the costs associated with test proctors, office space for testing, computers and IT maintenance of a proctored testing program make proctored testing significantly more expensive than UIT. Our research has shown that this assumption is most likely true for high volume positions (Lahti & DeKoekkoek, 2006).

In addition to the cost implications for using UIT, Internet recruitment is becoming one of the standard methods by which companies must vie for applicants. The Internet has revolutionized how employers source and process applicants. Rather than a manageable number of applicants that HR personnel can sort through, literally hundreds or thousands of applications are now received for a few job openings. The point of UIT is to move testing from the end of the hiring process up into the recruitment phase to effectively deal with the influx of applicants (Shepherd, Drasgow, & Beaty, 2004). Validated testing at the recruitment stage has the potential to pay enormous dividends over notoriously less valid screening procedures (e.g., resume review).

Another benefit of unproctored testing is the accessibility of the hiring process through the Internet. UIT allows applicants to take significant steps in the

<sup>1</sup>This model is based on the four-mode model that can be found in the International Guidelines on Computer-Based and Internet Delivered Testing, published by the International Testing Commission (<http://www.intestcom.org>)

hiring process from any location that has a computer with Internet access. This feature has the potential to allow companies to attract high-quality “passive” job seekers. These are people who may be well-qualified for a job, but who are not actively looking for a different job. While on the Internet, though, they may respond to a job posting or ad. These applicants can quickly be qualified by the test and fast-tracked through the hiring process.

### Risks Associated with UIT

While the Internet allows for testing anywhere at anytime, administering tests without verifying applicants’ identities could lead to compromised test sessions. The potential for applicants to cheat unproctored tests is always present. Yet, despite the distinctions made between proctored and unproctored testing, it is important to note that in any testing program, proctored or unproctored, it is never known with 100% certainty that the person who takes a test is the person who shows up for the job on the first day of work. There are numerous examples of proctored testing programs being compromised. The difference between proctored and unproctored testing is that with UIT the certainty of applicant identity is lower. Depending on the UIT mode that is used, the certainty of applicant identity can be significantly lower with UIT.

Applicants who want to cheat on the test can employ a number of strategies to beat the test, including logging in to the test multiple times to practice or get the answers, colluding with another person while completing the test, or hiring a test proxy to take the test. Tests with objective right and wrong answers and tests of a specific body of knowledge are among the easiest targets of people who try to cheat the tests. These tests are the targets of the Internet facilitated phenomenon of “braindumps.” Applicants share actual test questions and answers with one another by going to a website and entering all of the questions and answers they can remember from a test they just completed. There are numerous documented examples of cell phones and digital cameras being used in the capturing of information to later be “dumped” on a website. In some cases the individuals capturing the information for a braindump are organized and funded by companies specializing in getting the questions and answers for these knowledge type tests or tests with objective right/wrong answers.

***It is important to note that these risks are not unique to UIT. What is important is to understand how the risks increase or decrease with UIT.***

Figure 1: Cheating Risk Scale

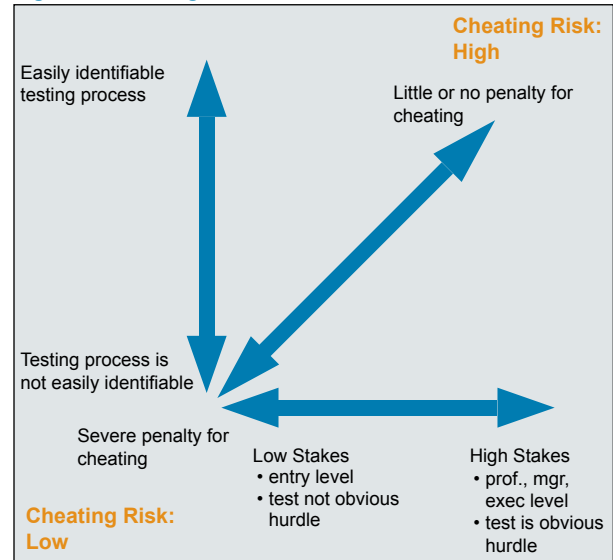


Figure 1 illustrates how some important themes we’ve seen in testing programs are likely to impact the risk of cheating (Beatty & Philo, 2005).

High stakes testing programs with an easily identifiable testing program (e.g., certification testing), and no clearly stated or truly aversive consequence for being caught cheating are key targets for cheaters.

Yet UIT does offer the potential for creating cheating deterrents. For example, dynamic test forms with the incorporation of new test items reduces the exposure of test items. This is a process that is facilitated by the use of online tests which can be more easily updated than static paper tests or tests that reside on a local computer rather than are delivered via the Internet.

### Does Unproctored Testing Work?

In spite of the risks associated with UIT, our research has shown that the UIT process works consistently and delivers ROI for our clients in pre-employment testing.

Three lines of PreVisor research have shown that:

- UIT scores, on average, are stable and do not typically rise over time (like you would expect if applicants were successfully cheating the test).
- In the face of any cheating that has gone undetected, unproctored testing sessions designed by PreVisor are typically valid and demonstrate ROI for our clients.
- In general, the diversity of applicant pools is maintained or increases when organizations switch to an Internet-based single point of entry into the hiring process.

At this point in discussing PreVisor UIT research it is important to say that the goal of our research is to determine how to make UIT work. We are not presenting only the research findings that cast our approach in the best light. Rather, the results of our research are determining our approach. We turn evidence-based findings into best practices. We believe UIT is a part of the present and future of pre-employment testing and want to provide the kinds of answers that help make UIT work in organizations.

The risks associated with UIT are real. The certification testing industry knows this pain all too well. But rather than conclude that UIT won't work without relevant data in hand, our goal is to understand the boundary conditions in which UIT will work and the tools and processes that PreVisor and our clients can use to make UIT work and generate ROI. A good first step in pursuing this research agenda is to monitor test scores over time.

### Do UIT Scores Rise Over Time?

A fundamental issue is whether UIT scores at a group level will be consistent or if cheating will raise the scores. Tracking average test scores over time is a basic but necessary first step in evaluating and establishing a baseline for UIT score stability.

A number of studies have revealed stability over time in PreVisor-built unproctored testing solutions. In an initial study on this topic we analyzed data from a client who was proctoring their leadership tests when possible, but out of necessity doing unproctored testing when they could not easily get the applicants to a testing center (Shepherd, Do, & Drasgow, 2003). Table 1 shows the results of the study.

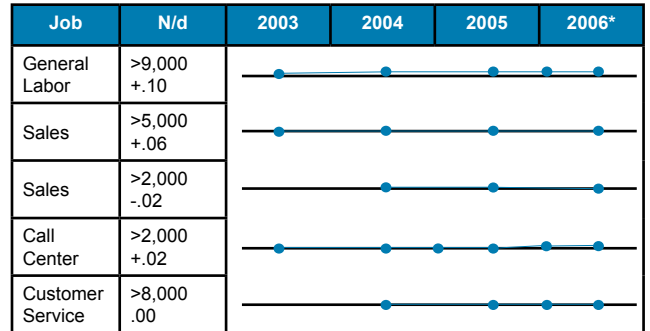
**Table 1: Proctored vs. Unproctored Test Scores**

Scale	Proctored Mean (n=1502)	Unproctored Mean (n=2628)	t	d
Conscientiousness (Non-Cognitive)	7.92	7.51	7.62	0.26
Leadership (Non-Cognitive)	9.68	9.25	5.37	0.18
Problem Solving (Cognitive)	6.84	7.16	-3.73	-0.13

Over a three month period the UIT scores for the non-cognitive content were actually lower than for the proctored content, and the UIT cognitive score differences were negligible in terms of the size of the score difference (a "d" value less than .30 is considered small in terms of score differences).

While these results were encouraging, we wanted to evaluate score stability across job levels, clients and over multiple years to more fully understand UIT score stability. The results of this research were dramatic: In more than 20,000 cases of testing, the majority of scores were highly stable over time (Beatty, Grauer & Davis, 2006). In the first graphic we see that scores are stable in a variety of entry-level jobs across multi-year testing programs.

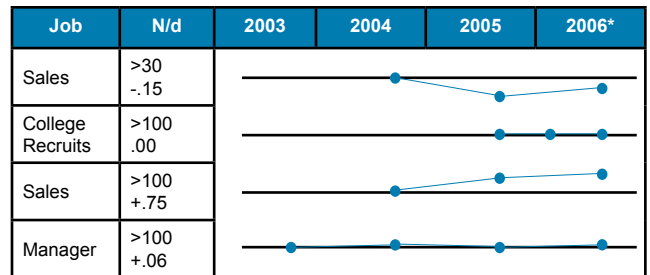
**Figure 2: UIT Norms for Entry Level Jobs**



\*January-February 2006

In Figure 3 a similar trend is seen for most of the UIT done with professional-level jobs.

**Figure 3: UIT Norms for Professional Level Jobs**



\*January-February 2006

In Figure 3, for three of the four clients doing professional level testing the UIT scores were stable or decreased slightly over time. In the fourth example there was a significant increase in scores over time. This client used the PreVisor recommended two-stage approach of unproctored testing with a follow-up proctored testing session for those who passed unproctored testing. What is interesting is that there is also a large score increase over this same time period for the proctored solution. One hypothesis we're investigating with this client is that their recruiting efforts have improved the quality of the applicant pool over time. In other words, average score increase over time does not automatically indicate that cheating has compromised the test. Score increase over time also does not necessarily indicate that UIT does not work and should be abandoned. Rather,

this example highlights the importance of one of PreVisor’s recommended best practices (reviewed in a later section of this paper): monitoring test scores over time to understand score change and adjust the testing process if necessary. So far our research has shown that stable UIT scores are the norm.

**Do Unproctored Tests Demonstrate Validity for Predicting Job Performance?**

Understanding score stability over time is the fundamental building block for our UIT research program. If scores are generally stable over time, the next logical question is, are those scores also valid for predicting job performance? Our research indicates that PreVisor-built empirically-keyed biodata, multiple choice and situational judgment pre-employment content delivered in an unproctored, on-line format is predictive of job performance. Two separate large scale studies are described below.

In the first study the participating organizations were in the financial services and telecommunications industries (Beaty & Philo, 2005). The jobs were entry-level call center customer service and collections representatives. The unproctored tests used in the studies were typically 30-50 biodata, multiple choice and situational judgment items that resulted in three to five scale scores. No cognitive ability measures were used. The results of four separate predictive validation studies are shown in Table 2.

**Table 2: UIT Predictive Validation Studies (2003-2004)**

Study	Metric Description	N	Observed r
1	Productivity	180	0.20
1	Dollars Collected	180	0.14
1	Supervisory Rating – Cust. Serv.	115	0.18
1	Supervisory Rating - Teamwork	115	0.24
1	Supervisory Rating – Sales	115	0.21
1	Supervisory Rating - Dependable	115	0.24
2	Average Call Handling Time	884	- 0.20
2	Composite Productivity Metric	884	0.17
3	Overall Performance Rating	75	0.26
4	Sales Metric	588	.24

More recently the results of six studies were aggregated across the retail, healthcare, finance and telecommunications industries. Table 3 shows the results for an individual business metric in each of the unproctored, predictive studies.

These short (8-20 minutes) PreVisor UIT solutions repeatedly demonstrate ROI for key business metrics. Depending on the quality of the business metrics

**Table 3: UIT Predictive Validation Studies (2005)**

Metric	N	Corrected r
Sales production	434	.17/.13
Retail sales metric	480	.17/.10
Call Center sales metric	662	.26/.22
Collections metric	232	.28/.16
Call Center - Call quality rating	255	.16/.14
120-Day Retention – Call Center	429	.18/.16

provided by clients, the validities range from the teens to the high .20s. These validities result in significant ROI for clients, particularly when used with high volume positions (Lahti & DeKoekkoek, 2006).

**Does UIT Reduce the Diversity of Applicant Pools?**

The concern about UIT impacting diversity stems from evidence of a digital divide (Payne & Weiss, 2006). That is, if minority groups are less likely to have computer/Internet access then will they be underrepresented in applicant pools? Payne and Weiss review several studies that show ethnic minorities, women and older workers, in comparison to majority groups, have either less access to the Internet in their homes or simply spend less time on the Internet. This digital divide is therefore a legitimate concern for the diversity of UIT programs.

In a recent large-scale study with five separate clients, applicant diversity (race, gender, age) was tracked before and after the implementation of UIT. In the two years after UIT implementation, race and gender diversity remained steady or increased (in other words, more minorities and females) for four of the five clients. Age diversity was tracked in two of the clients and in both cases there was a slight decrease in the average age of applicants (1-2 years).

There are too many factors at work in these results to conclude that UIT improves ethnic and gender diversity. But what the results do show is that when maintaining diversity programs that are in place before implementing UIT, race and gender diversity was by and large not negatively affected by the implementation of a UIT program. It is interesting to note that in the two clients where age was tracked there was a slight decrease in the age of applicants when moving to a UIT program. Though the age decline was not large (less than 2 years) and the findings are only from two organizations, it seems wise to monitor and potentially increase efforts to attract older applicants to the online hiring process.

## UIT Best Practices

Our best practice recommendations are driven by the results of our research and experience helping clients implement UIT solutions.

1. **Document the business case for implementing unproctored testing.** Like any testing related program, documentation is critically important to developing a world-class and defensible system. A potential outline for the documentation includes:
  - a. Why is testing needed for the target job(s)?
  - b. What, if any, testing is currently in place?
  - c. Why is UIT expected to be better than the current hiring process (consider using a basic utility model to quantify the savings)?
  - d. Describe the risks inherent in UIT, for example, that you never know with 100% certainty that applicants are who they say they are when completing the test
  - e. Describe the data that show that UIT works (use data and narrative from this white paper and other PreVisor ROI slides)
2. **Design your system with a single point of entry into the online hiring process.** Ideally this is through an applicant tracking system which collects significant identification information. Ideally your ATS is connected to a “one-use” url as well. That is, applicants can only complete the test one time when accessing a testing session link. PreVisor employs both single use links and a process called score re-use (recognizing applicants and using their past scores rather than allowing them to take the test again).
3. **Maintain current recruiting efforts.** Don't assume that because an application is posted on the Internet that there is no need to recruit from typical applicant pools. The research reviewed earlier suggests that diversity goals can still be met when the single point of entry into a hiring process is through the Internet AND recruiting efforts are maintained. The results from these early studies also suggest that a focus on encouraging older applicants to apply is warranted.
4. **Unproctored tests should be followed by a proctored assessment where the identity of the applicant can be confirmed.** Proctored testing does not ensure 100% identity certainty, but the probability is much higher that the person who takes the proctored test is the same person who reports for the first day of work.

We recommend organizations use the first, unproctored stage as a method for “screening out” candidates who are less likely to be successful on the job. The second stage is for “selecting in” the most qualified candidates. This two-stage model can be used to mitigate most of the risks associated with cheating. Cheating can be conceptualized as a two (pass-fail) by two (cheat-honest) matrix of candidates who completed the first unproctored step. The organization does not wish to proceed in the hiring process with either the cheating or honest failers. The cheating passers from the first stage are less likely to score well in the second proctored stage, making the honest passers from the screening stage more prevalent among the passers in the second proctored stage (Beaty and Shepherd, from Tippins, Beaty, Drasgow, Gibson, Pearlman, Segall and Shepherd, 2006).

The clear caveat about the two-stage approach is that stage one results should be confirmed by a second stage. We believe this approach is consistent with the APA Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2003), which counsel us to temper interpretations of data:

*“Psychologists provide opinions of the psychological characteristics of individuals only after they have conducted an examination of the individuals adequate to support their statements or conclusions. When, despite reasonable efforts, such an examination is not practical, psychologists document the efforts they made and the result of those efforts, clarify the probable impact of their limited information on the reliability and validity of their opinions, and appropriately limit the nature and extent of their conclusions or recommendations.”*

5. **Warn applicants not to cheat.** The unproctored test should begin with strong warnings to applicants not to cheat. In addition, the warnings should outline the potential impact of cheating on their test scores and steps that may be taken to verify answers to test questions. PreVisor can provide you with standard language we use in our UIT programs.
6. **Monitor test scores over time.** Despite data that show UIT scores in pre-employment testing are typically stable, UIT scores should be monitored over time for changes in score means and variability. This ongoing baseline evaluation helps identify threats to the UIT program.

## Ongoing PreVisor Unproctored Testing Research

Our current research is focusing on tools and processes for defending and detecting cheating. Although the cheating risks shown in Figure 1 make cheating a low risk in many pre-employment testing programs today, the future may be a different story. As applicants begin to understand more and more how the Internet is a part of the hiring process the cheating risk level is likely to increase. Our research on defending against cheaters and detecting when they have breached test security measures will prepare PreVisor and its clients for the future of pre-employment testing.

Another important area of current research is the use of cognitive tests in UIT programs. Most of the research described in this paper has focused on non-cognitive PreVisor solutions. Our research on the overall value of UIT programs has shown that when cognitive tests are not included in a solution the value of the testing program is reduced (Lahti & DeKoekkoek, 2006). However, the risk is that using test content with objective right/wrong answers makes the tests easier to cheat. That is, it is harder to determine the right answer for a test scored against job performance than, for example, a math question with only one right answer.

The first step in evaluating the feasibility of cognitive UIT has been to look at the stability of test scores in a small number instances where the data could be carefully reviewed. In a high stakes testing study using our recommended two-stage model for we found some test score increase for a little less than 10% of the applicants (Beaty, Fallon and Shepherd, 2002). As noted earlier though, these applicants tend to score lower in the proctored administration and were not selected based on the proctored test.

Our initial research with more low stakes testing has shown that at least in the short term (4 months or less) cognitive UIT scores have remained relatively stable. Tables 4 and 5 show the mean scores across time for cognitive UIT in two different client studies.

**Table 4: Cognitive UIT Norms Across Jobs – Client A**

Job	2005 to 2006 Change (d)
General Labor	+02
Customer Service	-.24
Manager	+13

**Table 5: Cognitive UIT Norms for an Entry Level Job – Client B**

2006	Mean	SD
Weeks 1-2	8.0	1.6
Weeks 3-4	8.3	1.5
Weeks 5-6	8.2	1.7
Weeks 7-8	8.1	1.7

## Summary

PreVisor research to date has revealed that UIT, in general, produces stable and valid scores without significant adverse effects on applicant diversity. We strongly recommend following our best practices in order to replicate the same findings in other organizations. Our research will continue to focus on identifying the best practices that make UIT a valuable and fair testing option for our clients and their job applicants.

## References

- APA Ethical Principles of Psychologists and Code of Conduct. American Psychological Association. (2003).
- Beaty, J., Grauer, E. & Davis, J. (2006). *Unproctored Internet testing: Important questions and empirical answers*. Presentation at the Annual Conference of the Society for Industrial/Organizational Psychology, Dallas, TX.
- Beaty, J. & Philo, J. (2005). *Ensuring test security and combating cheating in Internet-based testing*. Presentation at the Annual Conference of the Society for Industrial/Organizational Psychology, Los Angeles, CA.
- Beaty, J. C., Fallon, J. D., & Shepherd, W. J. (2002). *Proctored versus unproctored web-based administration of a cognitive ability test*. Paper presented at the Annual Conference of the Society for Industrial/Organizational Psychology. Toronto, Ontario, Canada.
- Lahti, K. & DeKoekkoek, P. (2006). *ROI for proctored versus unproctored assessment programs: Estimates from multiple utility models and identification of moderators*. Presentation at the Annual Conference of the Society for Industrial/Organizational Psychology, Dallas, TX.
- Shepherd, W.J., Drasgow, F., Beaty, J. (2004). New applications of computerized employment testing: Using human capital measurement for business decisions. *IHRIM Journal*, Sept/Oct, 40-46.
- Shepherd, W.J., Do, L., & Drasgow, F. (2003). *Assessing equivalence of online non-cognitive measure: Where research meets practice*. Paper presented at the Annual Conference of the Society for Industrial/Organizational Psychology, Orlando, FL.
- Tippins, N., Beaty, J., Drasgow, F., Gibson, W., Pearlman, K., Segall, D., Shepherd, W.J., &. (2006). Unproctored Internet testing in employment settings. *Personnel Psychology*, 59, 189-225.