Early in 2006, my colleagues and I published an article on unproctored Internet testing (UIT) in employment settings (Tippins et al., 2006). The purpose of that article was to identify the issues surrounding UIT and the ways in which those issues might be resolved. The panel of experts addressed a number of important questions about (a) the uses and applications of UIT and (b) the major issues and known problems associated with UIT, including test security, examinee identification, cheating, ethical use of tests, subgroup and cultural issues, standardization, and context effects. In addition, the panelists attempted to predict the future of UIT, highlight the research needed to facilitate UIT and provide advice to practitioners contemplating UIT. In the present article, “UIT” is used to refer to Internet-based testing completed by a candidate without a traditional human proctor. Thus, nontraditional forms of or alternatives to proctoring may be in place, such as quantitative analyses of response patterns, the use of video cameras, or follow-up testing with traditional proctoring.

At the end of the discussion, the panelists agreed on a number of things but did not reach agreement on others (see Table 1). Importantly, the panel could not agree that UIT was an acceptable practice. Some argued for it, others argued against it. Some claimed the practice of UIT was entirely ethical, others argued it was not. The reasons why UIT could not be supported included the beliefs that the identity of the actual test taker is not known; that some amount of cheating occurs on some UITs, particularly those that measure cognitive abilities; and that the effectiveness of deterrents such as warnings against cheating or threats of verification testing is not known. Another reason not to support the use of UIT was the panel’s confidence that overexposure of test items was a definite possibility with UIT and that familiarity with items would compromise the validity of the test at some point.

Because there was no consensus on the validity of inferences made on the basis of scores from UIT, the panel could not reach agreement on whether using a score from UIT to select individuals for further testing was an acceptable practice. Nevertheless, the panel agreed that UIT alone was never acceptable in high-stakes testing and recommended verification testing as the appropriate follow-up to UIT despite its costs and inherent inefficiencies. Acknowledging the possibility that the value of outside assistance differs depending on what the test taker is asked to do, the panel also emphasized the importance of considering the nature of the test in making the decision to use UIT and pointed out that noncognitive tests offered fewer opportunities for cheating.
In the 2006 article, UIT referred to Internet-based testing without proctoring. In the present article, UIT is used to refer to Internet-based testing completed by a candidate without a traditional human proctor. Thus, in the current conceptualization of UIT, non-traditional forms of or alternatives to proctoring may be in place, such as quantitative analyses of response patterns, the use of video cameras, or follow-up testing with traditional proctoring.

Since the 2006 article was written, the debate on UIT has continued. Exact figures on the number of unproctored tests given via the Internet are hard to obtain; however, papers presented at the Society for Industrial and Organizational Psychology’s (SIOP) annual conference and consultants’ literature suggest the usage of UIT is increasing. At the same time, test developers, both internal and external, have developed extensive item pools, and electronic means of monitoring test takers and sophisticated statistical means of detecting cheating have emerged.

Industrial and organizational (I–O) psychologists continue to wrestle with the appropriateness of UIT in employment settings and offer a range of opinions about UIT that seem to fall into five camps. Some argue that UIT is never acceptable because of questions regarding the identity of the test taker, the likelihood he or she cheated, and the validity inferences based on an unproctored test score. This group also poses questions about the ethics of UIT, given the unknown extent and forms of cheating that may have occurred.

The second camp believes that UIT is appropriate for some kinds of tests (e.g., non-cognitive tests) and/or for some purposes (e.g., personal development, practice testing) and attempts to define the parameters for acceptable usage. They argue that most kinds of cheating on noncognitive tests that can occur in an unproctored setting (e.g., social desirability distortions) can also take place in a proctored setting. Most of them concede that outside assistance is a possibility with noncognitive measures given in unproctored settings, but they believe that most test takers do not see a need for outside help and choose not to pursue that form of

---

Table 1. Consensus on Unproctored Internet Testing (UIT) in 2006

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The nature of the test (e.g., cognitive or noncognitive) plays a major role in the decision to use UIT because of the problems of candidate identification and unknown assistance.</td>
</tr>
<tr>
<td>2.</td>
<td>In high-stakes situations, UIT alone is never acceptable.</td>
</tr>
<tr>
<td>3.</td>
<td>Some form of verification testing is needed following high-stakes UIT; however, the benefit of administering the UIT in the first place is questioned by many.</td>
</tr>
<tr>
<td>4.</td>
<td>Industrial and organizational psychologists do not agree on the validity of inferences made on the basis of scores from UIT exams.</td>
</tr>
<tr>
<td>5.</td>
<td>Because of the open question on the validity of the UIT exam, the panel has not agreed that selecting a subgroup of those completing the UIT (i.e., high scorers) is an acceptable practice.</td>
</tr>
<tr>
<td>6.</td>
<td>Some cheating is likely when the stakes for the test taker are high.</td>
</tr>
<tr>
<td>7.</td>
<td>Neither the extent of cheating nor the effects of cheating deterrents (e.g., warnings, promise of a verification test) is known.</td>
</tr>
<tr>
<td>8.</td>
<td>The identity of the actual test taker in UIT is not known.</td>
</tr>
<tr>
<td>9.</td>
<td>In the absence of hard evidence, the effects of using UIT appear contradictory. On one hand, UIT opens up the application process to all with Internet access, and strong UIT scores may encourage some to complete the application and selection process. On the other, UIT may be limited to those with access to the Internet and comfort with computers and the Internet.</td>
</tr>
<tr>
<td>10.</td>
<td>Excessive exposure of test items is a possibility with any unproctored test, and Internet distribution increases the exposure. Familiarity with items may compromise validity and known solutions are imperfect.</td>
</tr>
<tr>
<td>11.</td>
<td>The panel is divided on the ethics of using UIT.</td>
</tr>
</tbody>
</table>

*aFrom Tippins et al. (2006).*
cheating. This group also points out situations in which there is no advantage to the test taker to cheat and encourages UIT at those times.

The third group of psychologists wants to prevent cheating before it occurs or stop it as soon as there is evidence that it is occurring. They present ways to prevent cheating and make the scores from a UIT equivalent to those from a proctored test. Some of these psychologists are focused on the effects of warnings against cheating, threats of retesting, or honor statements on test taker behavior. Others are exploring the use of technology to authenticate the identity of test takers, monitor test takers’ behavior during testing, and halt testing as soon as indications of cheating appear.

A fourth camp focuses on detecting cheating through verification testing or statistical means (e.g., multinomial logit, compromise item response theory (IRT), item functioning drift). The proponents of verification testing detect unusual differences in scores from unproctored and proctored testing sessions. The advocates of statistical detection of cheating employ statistical methods to identify test takers who are likely cheaters.

Finally, some are simply willing to accept UIT without extraordinary measures to prevent or detect cheating because of its utility. When comparing the benefits and the costs of UIT and proctored testing, it is cheaper to accept the costs of hiring a few cheaters than to spend the money to ensure accurate individual assessment, particularly when the validity of the test is minimally affected.

These camps of thought on UIT are not necessarily mutually exclusive. Some psychologists seem to adopt elements of several philosophical points of view. For example, almost everyone takes some steps to prevent cheating, and at the same time, a subset of these also pursues verification testing. Frequently, a psychologist will find a home in one camp for one kind of testing and in a second for another kind of testing. For example, one might believe that UIT is never acceptable for cognitive tests but is acceptable for noncognitive tests that depend on self-ratings. There are also a few psychologists whose beliefs and behavior are not aligned. They do not believe that UIT is entirely ethical because the validity and reliability cannot be fully described, but for a variety of reasons (e.g., no other realistic option, pressure from employer), they use UIT.

At the same time that I–O psychologists are debating the merits of UIT, many, including other UIT users outside of our profession, are looking for the profession to take a stand collectively on UIT and define best practice. Although some UIT users look to I–O psychologists to define what is acceptable, it merits noting that there are others who are unconcerned about our issues regarding ethics and validity when the economic case for UIT is so strong. When the value of a test is not recognized, the concern over practices that may diminish the value further is particularly weak.

The purpose of this section in Industrial and Organizational Psychology: Perspectives on Science and Practice is to explore Internet alternatives to traditional, proctored testing. This section is an outgrowth of sessions on the topic at the 2008 SIOP conference. This introductory article is followed by a set of commentaries based on these SIOP presentations. Many of the commentaries that follow will assess the current thinking on UIT, determine if there is any consensus on acceptable practices, identify new UIT issues as well as those that still face us, and present new ways to resolve the problems of UIT. Other commentaries will explore alternative ways of proctoring tests and discuss standards for testing via the Internet. This article sets the stage for those that follow by summarizing the important concerns I–O psychologists have about UIT.

Why UIT?

For many I–O psychologists and employers, the UIT train has left the station. They have embraced UIT as an efficient, cost-effective solution to the problem of testing large numbers of widely dispersed candidates. The question is not “Should we use UIT?” Rather, the question is “What is the best way to use
UIT?" By deploying selection programs through Internet-based test administration platforms, employers avoid myriad challenges to efficient staffing and minimize their costs. Job candidates do not have to travel to a testing location. Test proctors need not be hired, trained, and sent to testing locations. Testing equipment does not have to be purchased, distributed, or maintained. The practice of UIT also often results in easier, less costly maintenance of the testing system. By remote downloads from central servers, most online testing programs can be updated or corrected, cutoff scores can be adjusted, and experimental items inserted.

In addition to minimizing staffing and other on-site testing costs, there are other advantages of UIT. When a sophisticated test administration platform is used, Internet testing can provide a great deal of consistency in test administration, test scoring, and data recording, particularly when compared to the levels of accuracy and thoroughness that are found in some employment offices where human administrators are used. Extensive data about the testing event such as item-level responses, answer changes, response times, and so forth can be easily captured. Instructions are identical regardless of where the test is administered. Tests are accurately timed and scored, and all results are captured.

Although confirming data are lacking, many employers believe that an expeditious staffing process is a requirement if the best applicants are going to be hired. Because candidates do not have to wait for appointments, testing can occur as soon as an application is processed (and sometimes before an application is even made!). Thus, the employer with the fastest employment process is believed to attract the interest of well-qualified candidates before other employers. Many feel that UIT presents a positive image of a company that uses cutting-edge technology in its staffing systems.

UIT also opens the door to evaluating employed candidates who might not apply or test if they were required to go to an employment office during normal business hours or otherwise use their precious free time. UIT may also be attractive to candidates who are not sure they are qualified and are willing to spend the time taking a test at home but would be unwilling to go to an employment office.

Interestingly, some employers have investigated their traditional proctored testing and found that their assumptions about the advantages of proctoring are not true. Test proctors are sometimes unskilled, untrained, or uncommitted to their work. Consequently, they do not adequately monitor the behavior of test takers. For such employers, cheating opportunities and security issues that result from UIT may not be significantly greater than those associated with proctored testing.

In addition, technology has provided possible alternatives to human proctoring. Recent innovations have focused on electronic means of surveillance such as remote cameras and measurement of lag time between keystrokes. At this point in time, such tools have not been used sufficiently that we have fully evaluated and understood their strengths and weaknesses.

Why Not UIT?

Despite the obvious benefits of UIT, some nagging doubts remain. Only a few years ago, a primary concern was the stability of the hardware, software, and Internet connections necessary for UIT. The reliability of equipment and Internet connections, the consistency of experience across equipment, and the speed of processing were significant concerns. Today, many of the technical problems with UIT have been resolved, and the primary disadvantages of UIT stem from the many forms of cheating that are possible—assistance from others who have knowledge of the items before the test, assistance from others during the test, substitution of test takers, alteration of test scores in databases, and so forth.

Many I–O psychologists agree that some people cheat some of the time when taking a test under unproctored conditions; however, unless a proctored verification test is given, it is difficult to determine who cheated, when they cheated, how often they cheated,
how much they cheated, or what form their cheating took. Moreover, it must be noted that verification testing does not provide unequivocal evidence of cheating. Test scores may change because of a variety of reasons, such as regression to the mean, practice effects, anxiety reduction, and better health. In addition to the costs of multiple test administrations, employers must also factor in the costs of developing tests for verification testing. Most verification strategies require equivalent tests, adaptive testing, or large item pools whose item parameters have been identified based on item response theory. These efforts depend on large samples of test takers, significant investments in test item development, and sophisticated administration software.

An alternative to verification testing is statistical detection; however, most of these procedures assume very large databases of test scores from proctored and unproctored conditions. In addition, cheaters cannot always be identified until considerable time has elapsed.

Ideally, cheating detection occurs definitively on the first administration of a test to an individual and without regard to additional information about the test taker or the scores of other test takers. Moreover, because raising questions about someone’s integrity has serious consequences, the method of detecting cheating needs to be highly accurate and eliminate other causes of score changes, and the manner in which candidates are informed of their disqualification must be handled sensitively. In short, we need more information on the extent of cheating in general as well as effective methods for detecting and preventing cheating at the individual level.

I–O psychologists do agree that assistance on an item can distort the evaluation of the individual’s standing on the construct being measured, and familiarity with items on cognitive ability tests can affect the validity of the test. Thus, in at least some cases, UIT increases the opportunity for cheating, which leads to the likelihood of some cheating, which results in inaccurate assessment of the individual who cheated, which in turn diminishes the validity of the test.

Some employers also worry about the applicants’ perceptions of a company that relies on a staffing system that obviously has the potential for some forms of cheating. How many applicants want to work for an organization that appears to turn its head to cheating if not to tacitly endorse it? A few worry that UIT results in new hires who are less competent than their test scores would indicate and generally prone to cheating. The answer to the question regarding the extent to which the propensity to cheat on a test extends to the tendency to cheat on another work activity is unknown. Moreover, it is important to remember that proctored testing only screens out the low-skilled cheaters; highly skilled cheaters who meet the test standard may well be hired.

Another concern about UIT is the environment in which the candidate takes the test. A fundamental tenet of good testing practices is providing test takers with conditions that facilitate their performance on the test; yet, many test takers take their unproctored testing in environments full of distractions that are likely to affect their test performance. Although often dismissed as outside the employer’s responsibility when the candidate chooses the testing venue, the conditions under which the candidate takes a test can vary, and interruptions can profoundly affect performance. In tight labor markets, employers may well be concerned about finding enough people to meet their test standards and find that UIT limits the number of qualified applicants for the next stage of the employment process.

The security of test materials is often cited as a reason not to use UIT. Certainly, many selection systems employing UIT offer numerous opportunities for applicants to review test items before taking the test that counts. For example, applicants have been known to use multiple identities until they are comfortable with the test content. At the same time, computerized testing can enhance test security by eliminating paper copies of tests and scoring keys that are distributed in an organization for operational use.
Ethical, Professional, and Legal Guidance for UIT

An important question for all psychologists is the ethics of their practice; yet, psychologists are not in agreement on the ethics of UIT. Some psychologists believe strongly that UIT is not ethical, citing Section 9.09 of the American Psychological Association (APA) Ethical Principles of Psychologists and Code of Conduct (APA, 2002) as a proscription against UIT:

Psychologists who offer assessment or scoring services to other professionals accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use.

If one relies on reliability and validity evidence of a test administered under proctored conditions, the psychologist cannot accurately describe the reliability and validity of the inferences made under unproctored conditions. Some psychologists have established evidence of validity even when the test was administered under unproctored conditions. Perhaps a more salient concern is the relationship of an individual's test score to a normative database, which is not known unless the extent of cheating by the individual can be determined.

Other psychologists cite Principle 9.01 of the Ethics Code to support their use of UIT:

[Pr]sychologists 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.

These psychologists suggest that as long as they have clarified the effects of UIT, it is an acceptable practice.

Of course, the ethics of UIT are quite complex and raise many difficult questions. For example, if a test has been validated using scores from unproctored administrations to predict job performance and found to be valid, is it an ethical concern? Are we as I–O psychologists obligated ethically, legally, or otherwise to choose the test and administration protocol that produces the highest validity or merely tests and protocols with an acceptable validity? If the “highest validity” possible were a legitimate criterion for operational tests, many forms of testing in addition to unproctored testing would never be acceptable. Do we ever reasonably expect to get a completely accurate measure of a person’s true ability from one test score? Most of us would argue that a single measure is never completely reliable. Even when the reliability of a test is high, most of us would not be surprised to find scores of one or more individuals that are not very close to those individuals’ true scores. Complicating things further is the question of how to establish the validity of a score at the level of the individual rather than a test at the group level. Although we can predict performance at a group level with some accuracy, the prediction of performance of an individual is less stable.

For many, the primary ethical concern is that UIT allows cheating, and cheating is not moral. Therefore, UIT cannot be ethical. Yet, we are somewhat inconsistent in this belief. UIT is not ethical because cheating is possible; but other forms of tests such as personality testing and interviews also permit cheating in the form of socially desirable distortions, and many are not equally disturbed by the possibility of cheating. Some forms of cheating may be more acceptable than others. Putting one’s best foot forward when describing one’s past experiences or typical behavior is acceptable; accepting assistance is not. Alternatively, tolerance for different forms of cheating may be irrelevant. Psychologists may simply have a stronger obligation to detect cheating when it is possible.
Table 2. Open Questions Regarding Unproctored Internet Testing (UIT)

Extent of cheating
1) What kinds of cheating (e.g., outside assistance, substitution of test takers) take place, and what is the rate of each?
2) Does the rate of cheating vary by demographic subgroups?
3) Are there other moderators of cheating behavior (e.g., ability)?
4) How do the consequences of test results affect cheating behavior?
5) On what types of tests (e.g., cognitive tests) or test formats (e.g., multiple choice tests) is cheating most likely to occur?
6) What deterrents to cheating (e.g., warnings, threats of verification testing, honor statements) are most effective?
7) Does the kind of cheating or the rate of cheating change over the life span of a UIT program?

Impact on validity
8) What is the validity of inferences made from test results when the test was administered under unproctored conditions?
9) What is the acceptable tradeoff between decrements in validity and utility gains (e.g., cost savings associated with UIT, larger applicant pool)?

Impact on the organization
10) What effect does UIT have on each step of the hiring process?
   a. Does UIT affect the number and quality of applicants and test takers?
   b. What features of UIT (e.g., speed of hiring, accessibility of testing process) affect the number and quality of applicants and test takers?
   c. Does UIT affect the number of qualified applicants who progress to the next stages of hiring (e.g., interviews, background checks, drug tests)?
11) What effect does UIT have on the organization in terms of the capabilities of the new hires and their integrity?
   a. Is there a measurable decrease in abilities of new hires due to some percentage of the applicants hired having cheated and subsequently being hired?
   b. Is there a measurable difference in performance between those who test in a proctored setting and those who test in an unproctored setting?
   c. Are more questions about individual integrity raised in groups of people who took tests in an unproctored setting compared to those who took tests in a proctored setting?
12) What are the costs and benefits to UIT and the potential solutions?
   a. Is electronic surveillance cost effective?
   b. Is verification testing cost effective? How much do equivalent tests cost to develop? What is the cost of large item banks? How often must items be replaced?

Impact on the individual
13) How does UIT affect test performance? Does an individual do as well in an unproctored setting as he or she would in a proctored setting?
14) How does UIT affect individual applicant attitudes toward an employer? What are incumbents’ opinions regarding UIT?

Ethics and legal issues
15) Are there ethical concerns about UIT?

(continued)
In the United States, concerns about legal compliance in employee selection shape much of what I-O psychologists do. To date, UIT has not been directly evaluated legally in the United States, so there is little guidance. However, it is easy to imagine possible challenges ranging from the employer’s obligation to evaluate people under standardized conditions to the need to validate each test used in the verification process. Similarly, there is little professional guidance regarding UIT. The Principles for the Use and Validation of Employee Selection Procedures (SIOP, 2003), and the Standards for Educational and Psychological Tests do not directly address UIT. Instead, they provide some direction on appropriate test administration. Both have statements that at least give pause to the user of UIT and leave open questions regarding the expectations for effective test administration. For example, Standard 5.2 of the Standards states:

Modifications or disruptions of standardized test administration procedures or scoring should be documented.

Table 2. (continued)

16) Do the ethics of UIT depend on the type of test item? For example, would administration of a cognitive ability test be unethical in an unproctored setting whereas administration of biodata would not?

17) What are likely to be the legal challenges to UIT in the U.S. and other parts of the world?

Guidelines

18) How effective are the current guidelines?

19) What policies or guidelines need to be implemented?

20) What should SIOP/APA/EAWOP be doing?

21) Do the current ITC Guidelines represent good practice or should they be amended?

- If it should be amended, in the light of more recent research and experience, what changes should be made?

Technology

22) How can technology effectively monitor behavior in testing sessions?

23) Is electronic monitoring as effective in candidate identification and cheating detection as human proctoring?
when tests are administered in unproctored settings:

For moderate and high stakes assessment (e.g., job recruitment and selection), where individuals are permitted to take a test in controlled mode (i.e., at their convenience in nonsecure locations), those obtaining qualifying scores should be required to take a supervised test to confirm their scores. (p. 55)

Overall, there seems to be a need for clarification of the professional standards with respect to UIT.

Where Are We Now?

Although many of our clients have accepted UIT, many I–O psychologists remain wary of it because of questions about the effects of UIT and the standards for acceptable testing practices. Some psychologists believe UIT is an acceptable testing practice; some are waiting for answers before using UIT; others have all the information they need to reject UIT. Outside the field of I–O psychology, some expect the profession most closely associated with preemployment testing to define the standards for acceptable testing practice; others undoubtedly do not care what we think. A few actively disparage the value of testing in general and see the problems associated with UIT as trivial.

Table 2 highlights many of the open questions regarding UIT. The following commentaries in this section address one or more of these questions. The final integrative response to the commentaries will summarize what we know.

References


