

Psychological Testing on the Internet: New Problems, Old Issues

Jack A. Naglieri

George Mason University

Fritz Drasgow

University of Illinois

Mark Schmit

SHL USA, Inc.

Len Handler

University of Tennessee

Aurelio Prifitera

The Psychological Corporation

Amy Margolis

Brooklyn Learning Center

Roberto Velasquez

San Diego State University

Abstract

The past decade has witnessed a rapid expansion of the Internet. This revolutionary communication network has significantly changed the way people conduct business, communicate, and live. In this report we have focused on how the Internet influences the practice of psychology as it relates to testing and assessment. The report includes topics such as test security, how technical issues may compromise test validity and reliability, and hardware issues. Special attention is paid to ethical and legal issues, with particular emphasis on implications for people with disabling conditions and culturally and linguistically diverse persons. The report also covers issues specific to areas of practice such as neuropsychology, industrial-organizational, educational, and personality. Illustrative examples of Internet test use concretize the implications of this new medium of testing and its assessment limitations and potential. The most salient conclusion from this report is that the current psychometric standards, particularly those regarding test reliability and validity, apply even though the way in which the tests are developed and used may be quite different. Still, new methods made possible by emerging technologies will push the boundaries of existing psychometric theory and it is up to psychologists to test and expand the limits of psychometrics to keep pace with these innovations. The Internet provides a tremendous opportunity for testing but with opportunity there is a corresponding need for the ethical and professional use of test results. We encourage psychologists to think creatively about how their research and practice can be improved by Internet testing. Although there are many issues that await resolution, psychologists should look forward to this new medium with excitement and enthusiasm.

Introduction

At its spring and fall 2000 meetings, members of the Committee on Psychological Tests and Assessment (CPTA) discussed issues related to psychological testing and assessment on the Internet. They recognized that as psychological test instruments become more readily available via the Internet, issues arise concerning test reliability, validity, administration, item security, and test-taker confidentiality. Other issues that arise include access of tests by individuals who are not qualified to administer such tests and interpret the results; and tests that have been modified, changed, or translated without appropriate permission or validation. Moreover, it is possible that tests delivered on the Internet may be inappropriate for some groups, psychometrically flawed, or poorly translated into a variety of languages. Members of CPTA's parent boards -- the Board of Scientific Affairs (BSA), the Board of Professional Affairs (BPA), the Board of Educational Affairs (BEA), and the Board for the Advancement of Psychology in the Public Interest (BAPPI) -- reviewed and endorsed the idea of creating a task force to examine these issues. Specifically, BSA and BPA members agreed that it was important for APA to be a leader in discussing new and emerging technologies in psychological testing, assessment, and research, and in providing information about Internet-based testing and related issues. They noted that a jointly sponsored BSA/BPA task force on Internet-based testing would complement other APA groups focusing on related issues, including Internet-based research, the delivery of mental health services over the Internet (i.e., telehealth), and technology applications. Once the APA Board of Directors approved the concept, the Task Force on Psychological Testing on the Internet was formed with a broad mission of reviewing current practices on Internet-based psychological testing and determining psychometric, ethical, legal, and practical implications of this approach to testing. Throughout its work, the Task Force has been sensitive to the additional issues that arise from culturally and linguistically diverse populations.

As noted, the Task Force is a joint effort between science and practice. The BSA and the BPA appointed the Task Force co-chairs, who then selected the Task Force members from nominations solicited broadly across APA constituencies. Task Force members were chosen to reflect expertise across a broad range of testing areas (e.g., educational, school, employment, forensic, career/vocational, clinical, cross-cultural, neuropsychological), to be knowledgeable in Internet technology, and to represent the concerns of diverse groups that may be affected by testing.

The Task Force met twice during a six-month interval from winter 2001 through spring 2002 to discuss emerging issues and appropriate ways to respond. A primary objective of this Task Force was to prepare a report, which informs the profession of psychology about emerging issues and problems in Internet testing and actions psychologists can take to protect the integrity of testing and the consumer. An accompanying objective was to devise mechanisms for informing and educating the public about potential problems with Internet testing.

This report is organized into the following nine broad sections: a) Background and Context, b) New Problems Yet Old Issues, c) Technical Issues, d) Test Security, e) Issues for Special Populations, f) Types of Internet Testing, g) Illustrations and Examples, h) Ethical and Professional Issues, and i) Recommendations for the Future. Throughout these subsections, both practical and scientific issues are discussed with careful consideration of the consequences of decisions based on information obtained from Internet tests. The goal was not to provide a thorough summary of all Internet testing practices, but rather to describe broadly the current state of practice. It will become obvious to the reader, as it did to the committee, that many issues about Internet testing practices are similar to those faced by the profession in the past. This report also contains recommendations for the profession's response to current developments in Internet-based psychological testing.

Background and Context

Growth of the Internet

The past decade has witnessed a rapid expansion of the Internet. The foundations of the Internet began in the 1960's as part of the U.S. defense system development and the inception of new data-passing technologies. Since then it has grown from a university-based network to a worldwide network of interconnected computers accessible to people through many media in all civilized parts of the world (Abbate, 1999). The number of sites available on the Internet and the number of users grow larger each day. This relatively new and revolutionary communication network has significantly changed the way people conduct business, communicate with others, and live.

Over the past ten years, the number of Internet sites and users has grown from hundreds to millions. As of July 2001, there were over 125 million Internet sites (Internet Software Consortium, 2001). Approximately 143 million, or about 54%, of all US citizens have home access to the Internet (U.S. Department of Commerce, 2002). In fact, a recent report suggests that only 24% of Americans are truly offline, with no direct or indirect experience with the Internet (Pew Internet & American Life Project, 2003a). While access varies worldwide by country, it is clear that similar percentages hold, on average, across the globe and that more than 500 million people around the world have access (Nielsen//NetRatings, 2001). Further, close to 40 million US citizens have access to the Internet at work. At home, the average U.S. user spends about 26 hours a month on the Internet; while at work, the average user spends over 75 hours per month on the Internet (Nielsen//NetRatings, 2003). Clearly, the Internet is quickly becoming a medium as

pervasive as radio and television, with the capacity for infinitely more two-way communication.

The interactive nature of the Internet is evident in the substantial impact it has had on electronic commerce. Monthly online spending by US citizens is currently over 5 billion dollars and projections for 2002 Internet e-commerce were in the neighborhood of 72 billion dollars (Forrester Research, 2001). Internet transactions include both business-to-business and business-to-consumer sales. In fact, the business to consumer model has been exploited by many people because the Internet provides an easily accessible, low investment opportunity for any entrepreneur who wishes to develop a website and sell products and services. Clever website design and marketing practices can make small businesses look and feel like much larger organizations. Other, less interactive media, such as radio or television, have not provided the same revolutionary opportunities as the Internet.

There is no doubt that the Internet is quickly becoming one of the most pervasive communication and commerce media in the world (Abbate, 1999). It is also quite clear that usage and access will continue to grow for years to come. For example, while ethnic minority members in this country were sometimes among the last to utilize the Internet, their usage has grown significantly over the last five years (U.S. Department of Commerce, 2002). Individuals with disabling conditions are also connecting in growing numbers (U.S. Department of Commerce, 2002). Further advances in technology, such as broadband allowing quicker and larger transmissions of data, are likely to increase Internet activities. The impact of the Internet has been, and will continue to be, substantial for all people and organizations, including professional groups. The profession of psychology is among the groups that are beginning to explore opportunities and issues, both positive and negative, concerning the Internet (Barak, 1999). And so we should, as many individuals look to the Internet for psychological information. A recent report

suggested that 23% of Americans with Internet access have looked for information about a mental health issue such as depression or anxiety (Pew Internet & American Life Project, 2003b).

Growth of Internet Testing

Why has Internet testing created so much interest? Internet advocates stress better, faster, and cheaper services and products and Internet testing provides many good illustrations of this principle. For example, a new test with accompanying translations could be made available around the world almost instantly. Test publishers can download new tests to secure testing sites in a matter of moments, while other test developers can put tests on their web sites and make them available to anyone with an Internet connection. Updating a test is also much easier. For example, revising a paper-and-pencil test requires printing and distributing new test forms and answer keys and printing new or revised test manuals, an expensive process that may take several months or years. Revisions of a test that appears on the Internet can be downloaded to testing sites around the world in a few minutes at virtually no cost.

In many paper-and-pencil testing and assessment programs, examinees typically receive their scores and interpretive reports a month or two after they take a test. Their answer sheets must first be mailed to the test publisher, where they are scanned and scored, and perhaps interpreted. Then reports are created, printed, and mailed back to the examinees. In an Internet setting, responses are recorded in computer files as examinees answer each item. Software that computes test scores and generates interpretive reports can be run as soon as the last item is answered, with examinees receiving feedback within a few seconds of completing the test.

Internet testing is more scalable than paper-and-pencil testing. In the language of the Internet, "scalable" means that adding volume results in very little

additional cost. Therefore, over the course of a year, the number of times people who visit a website and respond to a test may increase, for example, from 5,000 per month to 10,000 but the test publisher does not incur the costs associated with printing, distributing, and scoring 5,000 additional paper-based tests. Of course, eventually an additional server may be required, but additional test administrations are much lower in cost in an Internet setting compared with paper-and-pencil administration. Additionally, because of the minimal costs involved, tests and assessments could be made available at no cost to respondents. For example, researchers may put tests and assessments on their web pages hoping that people will complete the assessment in order to receive a score report. In exchange, the researcher obtains the data provided by the respondents. Test publishers sometimes put free assessments on their web sites as a means of attracting potential customers.

Goals of Testing

Internet tests and assessments can be classified into three broad categories according to their goals. First, many Internet tests, instruments, and surveys are designed for personal development and growth, and may or may not be scientifically based. Measures such as these are usually designed for the layperson or public consumer. These instruments may be used to identify specific personality characteristics or traits (e.g., motivation for success; matchmaking), determine suitability for a particular type of job or trade, or facilitate psychotherapy (e.g., a personal rating of depression or anxiety). Secondly, many traditional psychodiagnostic measures like the MMPI, MMPI-2, and Beck inventories now appear on the Internet. These instruments are typically used to make important diagnostic and treatment decisions regarding individuals. In addition, there are web pages that are devoted to ways to respond so as to create a desired result, especially in forensic

settings where instruments like the MMPI-2 may be pivotal. Finally, cognitive ability tests, certification tests, and licensing exams can be administered via the Internet with the purpose of identifying the best candidates to be awarded some scarce resource (e.g., a job or admission to graduate school) or credential. Here the test or assessment is used to make an important decision about the examinee, usually related to access into a profession or area of study (e.g., medicine, psychology, etc.).

The goals of a test have important implications for how the measure should be administered. For example, a test that is designed for personal development or growth is less likely to be affected or contaminated by some response style. The person is likely to be candid and open, and not defensive or guarded. In these situations, there is little motivation for dissimulation on the part of the respondent. On the other hand, when a measure is used to make a decision about the examinee, the process is more likely to elicit a motivation to obtain a better score. If test takers are sophisticated, they will answer items in a way that they believe maximizes their positive results. For example, an individual may fake good on a personality inventory or cheat on a cognitive ability test. Much more care in test administration is needed in such situations.

Benefits of Internet testing

The benefits of Internet testing are speed, cost, and convenience. Testing over the Internet provides rapid communication of findings to clients, patients, researchers, and the public. It also allows researchers to collect data rapidly, conveniently, and at lower costs than in face-to-face research settings. Internet testing is cheaper and more efficient; it saves valuable time and provides results more rapidly and easily compared to face-to-face testing. Benefits of Internet testing also include sensitization and familiarization of testing to potential clients and the

presentation of test materials in a consistent, uniform manner. The more that potential clients become familiarized with these procedures, the more comfortable their approach to the tests can be, reducing spurious sensitization and situational effects.

Internet testing is also beneficial in that it allows patients in rural settings to be tested, where it would be difficult or impossible to travel to a testing center or to the office of a testing professional. Internet testing is of value to patients who lack transportation to such sites, or to those who cannot travel because of physical limitations. In addition, tests may be presented in a precise manner or in interesting and novel ways, so that the client's attention to the testing task is enhanced, compared with face-to-face administration.

Content and Quality of Internet Testing

In order to obtain some indication of the status of testing on the Internet, members of the Task Force used the keywords "psychological testing" and "Internet" to search several Internet databases. The result was that almost a million sites were found. Using the *Google* and *Yahoo* search engines yielded about 796,000 and 837,000 citations, respectively (in June 2003). It is difficult to estimate the exact number of actual web pages because some of the sites on both search engines were repetitions. To gain some idea of the content of all these web sites, a survey of the citations was conducted using the first 250 *Google* sites and the first 200 *Yahoo* sites, because there were fewer repetitions among the earlier citations. The following nine categories of sites were found:

1. *College and University Course Descriptions*: websites describing testing or assessment-related course and course content, included are lists of university-based courses in testing or assessment. Most of the information

seemed to be presented by credentialed individuals, but some of the websites were presented without such information.

2. *Individual Psychologists, Clinics and Medical Centers*: websites describing the various types of testing or assessment services they provide. These ads focus on the description of individual psychologist's credentials, or in the case of a clinic or a medical center, providers' credentials, plus a short description of the types of referral issues in which they specialize and the types of testing or assessments they do. In one case, a university-based department of psychiatry also offered online tests for depression, anxiety, sexual disorders, and personality disorders. The test taker could then obtain immediate feedback.
3. *Industrial/Organizational Firms*: websites describing the Internet-based services offered. Their advertisements range from essential descriptions of services noting the advantages of Internet testing to sites that actively promote their products. For example, one site announced, in large, bold letters, "Some people will say 'You Can NOT Predict Successful Performance'...Don't Believe It! That may have been true at one time...but NOT ANY MORE!" Some sites even quote their scores on accurate prediction (e.g., "Over 4 million assessments...90% accuracy!").
4. *Psychological Testing*: websites devoted to information sources concerning psychological testing of many kinds. Some of these websites provided links to other informational websites. A wide variety of sources was included, ranging from organizations that have some connection to testing and/or assessment, to sources of various tests available for purchase from various test publishers. Links to library information concerning tests and testing were also provided. Information was also available about the sources of

professional standards for testing and assessment, as well as sources of test reviews.

5. *Educational Tutorials*: websites describing, sometimes in detailed, college level language, important test and assessment related information, such as reliability, validity, test standardization, etc.
6. *Research*: websites of researchers who have collected or who are collecting data on the Internet, using a variety of psychological tests. Sometimes the research findings are included in great detail, and sometimes there is only a summary available. These websites typically contain information about the professional credentials of the researchers, often including a bibliography of their work. Sometimes the entire contents of their published professional articles were also included. Many of these sites invite the person who accesses the site to be a subject and to take the test the researcher is currently developing or validating. For example, one such site describes an ongoing project concerning the implementation of the International Personality Item Pool, designed for the Five Factor Personality Inventory. This site contains a description of the project to date, a description of the inventory and the scores obtained, ways in which the inventory may be used, and future plans for the research program. The researcher, Tom Buchanan, a lecturer at the University of Westminster, in England, described online his method of producing a 50 item inventory measuring the domain construction of the five factor model. He states:

The participants found the website through various "search engines" using the terms "online test" or "personality test." The participants completed the inventory on-line, and it was automatically scored. Individual feedback was then given, in which the participants were able to see descriptions of the constructs measured by the inventory

and how their scores compared (expressed in percentiles) with those of others who had taken the test.

Buchanan indicates that the test materials are free for anyone to use for non-commercial purposes (e.g., research or teaching) without permission. The approach to the use of the Internet described above is based on the concept of the "scientific collaboratory," defined as "a computer-supported system that allows scientists to work with each other, facilities, and data bases without regard to geographical location" (Finholt & Olson, 1997).

7. *Online Tests*: websites providing personality and intelligence tests, available to be taken by consumers. Some of these sites offer immediate feedback online and others offer feedback if the test-taker provides his or her email address. Sometimes there is no test fee, and sometimes there is a cost for each test taken. This type of website was the most frequent one found. The tests that were advertised for the most part had no associated reliability or validity data, and no standardization information available. Nevertheless, these limitations did not deter those who are responsible for producing these tests from offering detailed interpretations after the user completed the test. Often the website included screening tests described as anxiety screening, depression screening, mania screening, personality problem screening, or sexual dysfunction screening.
8. *Multi-purpose Sites*: websites providing links to related sites. For example, one site, called a "resource page," offered a variety of information, such as book and journal lists, information about practice guidelines, employment opportunities, resources for patients, support group links, plus many other related links to chat rooms, newsgroups, and conferences.
9. *Services*: websites constructed by companies that sell tests or test report services. While some of these sites merely list tests with brief descriptions,

others go into great detail, describing, in one case, the validity scales available for the test as well as a detailed list of other scales available. In several cases, enough information was available online that test security was possibly impaired.

Differences Between Testing and Psychological Assessment

The important distinction between testing and psychological assessment (Matarazzo, 1990) is particularly important for Internet testing sites. Internet site developers as well as many others unfortunately use the terms "testing" and "psychological assessment" synonymously when actually these terms mean quite different things. Testing refers to the administration, scoring and perhaps the interpretation of individual test scores by applying a descriptive meaning based on normative, nomothetic data. The focus here is on the individual test itself. Administering a test is typically a relatively simple process that can be conducted by psychologists and possibly testing technicians, sometimes with relatively little training, or even by a computer. While more than one test may be given, the emphasis in each case is the comparison of each individual test score with the scores of an appropriate normative group.

Conversely, in psychological assessment the emphasis is typically on the person being assessed and the referral question, rather than on specific test results. Typically, an array of tests is given with an emphasis on their integration, taking many factors other than normative findings into account. The results of the tests are integrated among themselves, in the context of additional available patient/client data, such as history, observations, referral source, and information from friends and/or relatives. The eventual goal of the assessment is to answer the referral question or questions. Tests are typically employed in the psychological assessment process, but much more information and often much more complexity is involved.

The integration and interpretation of data in the assessment process requires a high degree of skill, psychological sophistication, and education.

To illustrate the distinction between testing and psychological assessment, Handler and Meyer (1998) use a medical context in which the medical counterpart of psychological testing "is found when technicians or medical personnel obtain scores on such instruments as a blood pressure gauge or a thermometer, or data such as a blood chemistry panel, deep tendon reflexes, and so forth" (p. 5). On the other hand, the medical counterpart of a psychological assessment is the process by which the physician takes the information from these various tests "and places them in the context of a patient's symptomatic presentation and history to adequately understand the full scope of his or her condition" (p. 5). Method variance is a major problem when one is doing psychological testing, but may be less important in psychological assessment because the psychologist typically takes these interference factors into account in the interpretive process.

This distinction between testing and psychological assessment is important because most of what is available on the Internet is testing, not psychological assessment. Therefore, the issue of method variance is an important one; for example, the test results obtained on the Internet may be inaccurate because of the specific method employed in the testing and there is no psychologist available to assist in interpretation. While it is conceivable that it might someday be possible, the requirements for appropriate psychological assessment exceed current Internet capabilities.

The Practice of Psychology

Scalability and convenient availability to a broad audience of consumers make Internet testing attractive to practicing psychologists. Several new companies have been formed because Internet testing creates business opportunities; traditional test

publishers have also entered this playing field. Individual practicing psychologists can also utilize the Internet. For example, routine assessment activities can be conducted via the Internet, leaving psychologists to devote their time to interpretation and feedback. Wilson Learning Corporation explicitly considers what psychologists do best and what computers do best as they design an assessment system for a customer (Burroughs et al., 1999). Psychologists prepare for face-to-face meetings, conduct feedback sessions, and write final interpretive reports. Computers store scores, mechanically combine scores, and generate feedback report support information.

Psychometric Advantages

Computerized tests provide some psychometric advantages in comparison to paper-and-pencil assessments. In fact, considerable research has been conducted to document and demonstrate these advantages. A brief summary is provided here; more detail can be found in Sands, Waters, and McBride (1997) and Drasgow and Olson-Buchanan (1999).

An Internet test and assessment provides more accurate scoring compared with a traditional paper-and-pencil test. Optical scanning of paper test forms encounters difficulties with stray pencil marks, incomplete erasures, and insufficiently darkened answers. In computerized testing, an examinee enters a response, the response is displayed on-screen, and the examinee is provided an opportunity to change the answer. Suppose an examinee has selected "B" as his or her response. The computer monitor will then display a darkened circle next to option "B" and will allow the examinee to change the response or proceed to the next item. If the examinee goes to the next item, well-designed software will correctly record and score the "B" response. By eliminating optical scanning, a significant source of errors is removed.

Internet testing and assessment is especially well suited for the use of item response theory (IRT; Hambleton & Swaminathan, 1985; Hulin, Drasgow, & Parsons, 1983; Lord, 1980) . For example, computerized adaptive tests (CATs) that tailor difficulty to the ability level of each examinee can be efficiently delivered through this medium. In this process, IRT technology would be used to select which items are given so that they are of appropriate difficulty for each examinee. Internet assessment can also offer the potential for assessing abilities and skills not easily assessed by paper-and-pencil methods. For example, Vispoel (1999) developed a computerized assessment of musical aptitude; Ackerman, Evans, Park, Tamassia, and Turner (1999) created a dermatological test that allows examinees to pan in and out as they examine color images of skin disorders; and Drasgow, Olson-Buchanan and Moberg (1999) developed an assessment that uses video clips to assess respondents' conflict resolution skills. These and many other needs can be effectively met using computerized technology that can be delivered via the Internet.

New Problems Yet Old Issues

Test/Client Integrity

In the same way that we do not allow clients to take tests at home, given that they might not take them privately, Internet testing encounters this old problem with a new twist. When the goals of the test taker differ from the goals of the test user, it is important to confirm the identity of the person answering items. The simplest and most effective method is to require test takers to go to a secure test site and show a government issued photo ID such as a driver's license or passport. Of course, test administration at such sites is inconvenient and expensive. When a test or assessment is not administered at a secure test center, there are a number of ways to check a test taker's identity (e.g., "What is your mother's maiden name?").

Unfortunately, such methods can easily be circumvented; a more talented accomplice can sit with the supposed test taker and provide answers to items.

Segall (2001) suggested a clever means of confirming the validity of test takers' administered tests remotely via the Internet. He has proposed Internet administration of the lengthy Armed Services Vocational Aptitude Battery (ASVAB) enlistment test used by the U.S. military. Segall's idea is that individuals could take the ASVAB at their convenience in nonsecure locations. Individuals who obtain scores qualifying for enlistment would then travel to secure test centers, where they would be administered a much shorter confirmation test composed of highly discriminating items. A statistical procedure developed by Segall would then be used to check whether the test taker's original responses are consistent with the responses from the confirmation test. This method was found to be very effective at detecting cheating in a simulation study. A combination of informing examinees that a confirmation test will be administered as well as applying Segall's (2001) statistical analysis may also prove to be effective in discouraging cheating.

Technical Issues

Host/Server Hardware Considerations

Internet testing poses a number of technical considerations that are not an issue in conventional testing. Prominent among these considerations are characteristics of the hardware involved, which can be generally classified into Host and Client issues. On the Host side, Internet test developers should have many of the same concerns that any Internet delivered application might have. Primary among these concerns should be high availability of the host network and hardware, a high degree of fault tolerance through all components of the host configuration, data protection through frequent backup and secure storage procedures, and

sufficient network bandwidth and hardware capacity for the testing requirements. Many useful checklists are available for evaluating the quality of services offered by prospective host providers. Hardware should be Enterprise class servers with built in redundancy in order to minimize potential impacts of hardware failures. Most providers will recommend and lease the appropriate hardware as part of the service agreement. Most host provider plans limit the amount of data that will be transferred to and from a site within a monthly period. Clients are charged for data exchanges that exceed this maximum data transfer rate. Therefore, it is important to consider the amount of testing (inbound) and, if appropriate, reporting (outbound) traffic that will be conducted in a monthly period.

The amount of data storage capacity that will be required will depend on testing volume, the amount of data collected per test and the duration that the data need to remain accessible. While storage costs are relatively inexpensive, these costs can be reduced by periodically archiving aged data. Virtually all reputable providers offer some type of network monitoring; however, many providers offer active monitoring services for additional fees. These services are intended to identify and correct network or site problems immediately when and, to the extent possible, before they occur. Reputable providers should be able to provide a guarantee of reliability for their site and should be able to provide historical reliability data.

A major consideration for any Internet-delivered application is available network bandwidth. Host providers will guarantee a minimum amount of network bandwidth available to handle all traffic coming to a site for all applications operating on the site. Bandwidth limitations are more typically observed on the client side. While client side broadband is becoming more widely available due to decreased costs of high-speed services, conventional dial-up 56K connections still represent the vast majority of Internet traffic, particularly in the consumer and home markets. While Internet connectivity has significantly increased within the schools, much of

that connectivity is dial-up. Items that include high quality color graphics will be constrained by bandwidth available through conventional dial-up 56K connections. Bandwidth constraints and bottlenecks can be a significant threat to the standardized delivery of an assessment; the extent to which variability in network performance can be tolerated depends on the nature of the assessment and the degree to which variation in time between stimulus or item presentation affects examinee performance. Where strict control is necessary, a locally installed stimulus control applet should be used as discussed below.

Client Hardware Considerations

The required client hardware configuration will largely depend on the type of test and items being administered. In general, tests that display items with graphics, present timed stimuli, or collect response time data require greater consideration and specificity of the client hardware configuration. Graphics are memory intensive and often require workstations with additional Random Access Memory (RAM) and Video RAM (VRAM) to operate efficiently. The graphics and stimulus presentation characteristics of the test will constrain the minimum hardware necessary to present the items within the parameters defined for the test. Due to inherent constraints in an application's ability to control the timing of the delivery of Internet packets to a client, accurate presentation of timed stimuli and recording of response time data require a stimulus control application running locally on the client's workstation. Timing applications control item presentation, save examinees' responses at the client's workstation, and send response data back to the host application. Although such applications are generally small, the resource requirements of the application should be considered in the overall hardware requirements for the client. Peripherals will generally include a mouse and possibly a microphone if speech input is required or is an option.

The hardware characteristics of the client's display require special consideration where timed stimuli are presented. The first consideration is the display refresh rate, which indicates how many times per second the screen gets redrawn from top to bottom. When presentation of stimuli must be controlled to the millisecond, applications must take into consideration or attempt to control the refresh rate of the display. Furthermore, while most individuals cannot perceive any flicker on displays operating at refresh rates in excess of 72 Hz, a small percentage of individuals can detect flicker at rates as high as 85 Hz on conventional (non-LCD) displays. Because flicker is known to cause discomfort, it is likely to have some effect on performance on items that require visual decoding, particularly if performance is timed.

The standardized presentation of stimuli presents a particular challenge in the development of computer-administered assessments. While on the surface it may seem relatively easy to present exactly the same picture or graphic on any computer display, it is, in fact, almost impossible to do so without a recalibration of the display environment. All displays present the same colors at slightly different hues and, based on user preferences, have different contrast and brightness settings. Furthermore, the actual size of an object will vary depending on the resolution settings of the display. Text may be displayed differently depending on the fonts installed on the user's workstation. While such variations may not be of concern for some assessments, they can raise serious questions regarding the standardized administration of items with graphic stimuli. To address the issue of stimulus color, test developers should include a methodology and template that would allow the user to calibrate the display to a template standard. To control the standard size of a stimulus on different displays operating at different resolutions, the developer must either a) require the examinee to operate the test at a particular resolution or b) dynamically adjust the size of the stimulus based on the resolution setting. These

technical issues will require careful consideration in the Internet testing environment.

Test Security

Levels of security can range from highly secure and restrictive (e.g., high-stakes testing programs) to unsecured and permissive (low stakes testing). As might be expected, the greater the level of security, the higher the cost for implementing and maintaining an application. The level of security implemented for a given test or test site should be appropriately matched to the usage of the test. Secure test environments should use a 3-tier server model. Within this model, the test system is actually made up of 3 independent servers: an Internet server, a test application server, and a database server. It is imperative that the application server is solely dedicated to the test application. In order to maximize the security of client data, a separate data server should be maintained behind a secure firewall. This configuration reduces the possibility of unauthorized intrusions into client test data. If scoring and reporting services are required, it is recommended that these applications be placed on yet a fourth server in the middle tier with the application server in order to minimize processing bottlenecks that may affect the test application or data access. Regular and frequent backups of all collected data should be conducted and the provider should be able to give prospective customers a detailed disaster recovery plan. Redundancy allows a site to continue to operate even if one of its components completely fails. A reputable provider will have redundancy on all systems throughout its site including incoming and outgoing communications lines. As with any secure application, client and administrator password formats need to be robust (non-trivial) and actively maintained. Finally, server traffic should be actively and continuously monitored for intrusions.

On the client side, one of the most important security considerations is the prevention of unauthorized copying by the examinee or an observer and printing of test items. This can partially be achieved within a browser by disabling access to

menu selections such as cut, copy, paste, export, save, save as, print, print screen, etc. Hot keys and right mouse context menu selections should also be disabled. However, it is only possible to partially secure items by controlling browser functions. Even with such controls in place, it is still possible for more technically knowledgeable examinees to make use of operating system features and other applications to capture items from the screen. Therefore, where full client-side security is required, it is necessary to install a test security agent on the client's desktop, which completely prohibits an examinee from dropping out of the test application while it is in operation. Such an application prevents users from launching screen recorders, word processors, e-mail applications, and any other unrelated application that may be used to compromise the security of test items.

Issues for Special Populations

The delivery of psychological tests through the Internet provides the opportunity to meet the needs of a wide variety of individuals, in particular, important special populations including people with disabling conditions and culturally and linguistically diverse persons.

People with Disabling Conditions

A critical issue in determining appropriate accommodations for a person with a disability is demonstrating the clear relationship between the individual's deficit and the nature of the accommodation. The challenge of determining the type of accommodations required for Internet-based assessment arises in part because little is known about the unique aspects of testing in this format. Although many of the accommodations developed for paper-and-pencil testing can be used for Internet assessments, new issues will likely arise. As psychologists begin to make recommendations to institutions on behalf of individuals with disabilities or on behalf

of institutions attempting to design fair testing practices for groups of individuals with disabilities, it is important to consider new types of accommodations to address the unique problems inherent in Internet assessment.

Accommodations may be considered in terms of operating at the level of the individual or at the level of the group. At the individual level, adequate accommodations include alterations in

the testing environment. Although this is standard practice with paper-and-pencil testing, unique challenges may be encountered with Internet assessments because the computer may be permanently affixed in one position. Adjusting the height and placement of the table on which the computer sits is critical for an individual in a wheelchair. For some disabilities, accommodations require alterations to test administration itself rather than alterations to the environment. For example, a reader is often recommended for individuals who are sight-impaired or who have a specific reading disability (e.g., dyslexia). During Internet-based assessment, there is likely to be a new vocabulary to describe the spatial layout of the material and the actions taken by the reader (e.g., rather than stating, "I am filling in answer a on the scantron sheet," the reader might say, "I am clicking choice a on the answer screen"). These accommodations are not unique to testing over the Internet, but are unique to testing on a computer platform, the frequency of which will likely increase as Internet technology advances.

At the group level, a lack of equal access to technology may result in poorer test performance for some groups. Households with lower incomes have reduced access to computers and therefore the Internet. Assessment over the Internet, therefore, may be confounded by the novelty of the format. For example, during cognitive testing, individuals who are less familiar with computers will have a greater cognitive load due to divided attention than individuals who are familiar with computers. Further, a lack of familiarity with the security and privacy features of the

Internet may influence performance. In this sense, low access to computers may be viewed as a disability that requires accommodations to ensure fair testing.

Culturally and Linguistically Diverse Groups

Many culturally and linguistically diverse groups, including Latinos and African Americans, have been among the last to connect to the Internet because of economic and/or access issues. Yet, the number of people from these and other minority groups that have access to the Internet is increasing dramatically. For these groups, the Internet is proving to be a tool that connects them to their country of origin, resources in a particular language or dialect, and so forth.

Like the majority Euroamerican population, members of these groups have also begun to access the Internet for information related to mental health and psychology. For example, it is not unusual for a Spanish-speaking Latino to seek out information about a particular mental condition or even a psychological test instrument through the Internet. The person is now likely to find information in Spanish, usually from an Internet site in Latin America. Similarly, the person may seek information about a particular test that he or she is about to take that will be administered by a psychologist (e.g., for employment screening or child custody purposes).

There remain many unanswered questions regarding the psychological testing and assessment of these groups via the Internet. In many ways, these issues are similar to concerns related to test use with culturally diverse or minority groups (e.g., fair assessment). For example, it is unclear if it is necessary to have separate norms, including norms for minorities, for an instrument that is administered via the Internet versus administered in the traditional manner. Also, a review of various webpages indicated that many instruments are poorly translated, or have been modified for use with Latinos in the United States or Latin America, by Spanish-

speaking professionals usually outside of the United States. Additionally, older measures, such as the MMPI, can still be found in Spanish despite the appearance of recent translations of the MMPI-2 that are superior to the translations of the older MMPI. People may use old and outdated instruments in a manner that is inappropriate or problematic, which may then result in negative consequences for clients and the public.

Types of Internet Testing

Neuropsychological Testing

Neuropsychological testing over the Internet could simplify the process of administering a measure repeatedly over time to track treatment progress, disease process, or medication benefits (Heaton et al., 2001; Erlanger, Saliba, Barth, Almquist, Weberight, & Freeman, 2001) by reducing the costs of administration and storing the data. Some companies have taken advantage of this Internet feature in order to gather normative data based on multiple administrations of neuropsychological tests designed for the computer (e.g., Erlanger et al., 2001). A second advantage that the Internet offers to neuropsychological assessment is the ability to conduct group rather than individual assessments. For example, in the assessment of sports-related concussion, until recently, athletes had to be tested individually to gather baseline data prior to the season. This was extremely time intensive and expensive. Some products are now available for use over the Internet that allow entire teams of athletes to complete baseline cognitive function assessments simultaneously in a school's computer lab. The Internet provides easy centralization of data for future use if an athlete is injured. Further, by using the Internet, athletes can complete measures multiple times to determine when they should be permitted to return-to-play. Thus, the Internet allows for the centralization

of longitudinal recovery data that should further standardize return-to-play decision-making. Moreover, potential applications of this type of group-baseline administration may be a useful tool to collect longitudinal follow-up data of individuals, to study developmental disorders such as ADHD, to track medication benefits, or to study aging and identify markers of disease processes, such as degenerative dementias.

I-O Internet Testing Sites

Industrial and Organizational (I-O) Psychology is quickly embracing the Internet as business and industry demands for electronic data processing have become the standard (Harris & Dewar, 2001). The Internet has provided the opportunity for delivery of both traditional testing (i.e., a simple transformation of paper-and-pencil tests to Internet delivery platforms) and new and unique forms of testing. In addition, the potential for cost-efficiencies and timely interaction with test-takers that is provided by Internet applications is tremendous. Many of the I-O psychology consulting firms and organizational departments employing I-O psychologists have built test engines and other information technology platforms to meet these growing needs and challenges (Schmit, 2001a).

I-O psychologists use testing in a variety of organizational interventions and programs. Perhaps the largest use of testing in this field of psychology is in personnel selection programs, including both the hiring of external candidates and the promotion of internal candidates through a succession management or career-pathing process. When hiring externally (or internally for promotion or placement decisions), job candidates are administered tests which are used in conjunction with other information about the candidate (e.g., application, interview, references, or other performance information) to make a hiring decision. For internal job incumbents, testing is used in training, development, and coaching/mentoring

programs. Here testing is often used to establish contextual information about training needs, trainability, or characteristics that may enhance or detract from training success. Test information is also used to help people in the training process understand their own (and often others') inclinations, preferences, and styles for interacting with others or for engaging in other job-related tasks. Professional facilitators then help these individuals to understand how to use this information in their workplace interactions and tasks. For all of these programs in which testing is used by I-O psychologists, the Internet is quickly becoming the medium of choice for administration.

Traditional personnel selection testing programs have typically involved the administration of a battery of tests that may have included knowledge, skill, cognitive ability, personality, situational judgment, biodata, physical ability or other types of tests that have been shown to predict job performance or training success (Guion, 1998). Most often, all job candidates will be asked to come to a physical location (e.g., a company site, job fair, or contracting vendor site) to complete the battery of tests under the guidance and watch of a trained test administrator. This process can place heavy demands on time, space, cost, and human resources of an organization, particularly for organizations with high job growth and/or turnover burdens. The Internet is being used to reduce some of these resource demands associated with traditional personnel selection testing (Harris & Dewar, 2001).

Many large business and industry organizations have moved toward automation and cost-reduction in the hiring processes. To achieve this goal, various applicant tracking systems and hiring management systems are used that are either part of the internal information technology infrastructure or are provided by an applicant service provider (ASP) where the computer program and data reside on a third-party's computers and is accessible through the Internet. This latter form of data processing is quickly becoming the standard because the advantages, such as

instant upgrades and shared maintenance costs, often outweigh the disadvantages. In brief, these systems allow candidates to access company information, such as job openings and descriptions, and to apply for jobs, while providing company recruiters and managers access to candidate information and qualifications. The advent of these types of applications and internal pressures for cost-efficiency have led organizations to push all vendors associated with the hiring process (including vendors of psychological tests) to become more efficient in the delivery of their part of the process. Thus, vendors of psychological tests have adapted similar models and technology to meet client demands (Schmit, 2001a).

Assessment can be more efficiently conducted through Internet testing when some portion of the procedure is completed before the candidate is physically at the service delivery site. That is, if all candidates can complete a test on the Internet as part of the application process, this information can be used to screen in the candidates with the greatest probability of job success or screen out those who may be detrimental to the organization in some way (e.g., those who might engage in counter-productive work behaviors). This information, combined with other application information, can be a basis for paring down a large number of candidates to a smaller pool, which reduces the administrative burden on the organization. Thus, many vendors of psychological tests used in the personnel selection process have developed screening methodologies that can be used for this purpose. The most basic form of assessment involves questions that may simply ask the candidate about work preferences or availability. Other applications have been developed that use biodata, or work background items that have been shown to predict performance on a specific job (Mitchell, 2001). Situational judgment tests are often used in screening as well. This type of test includes items that present a work context and problem to which candidates must respond with the action they would take, usually

picking from a list of possible alternatives. Personality, cognitive ability, and knowledge tests are also used in some of these screening applications.

The tests used in the screening process are often somewhat different from the more traditional paper-and-pencil tests used in selection. Typically organizations demand that the tests used in these situations are short in length and present the candidate with a pleasant experience. That is, organizations want the test to be a recruiting tool at best, and, at worst, to provide an experience that will not turn the candidate toward an organization competing for the same talent. To achieve this objective, test developers have often shortened tests to provide the maximum validity (i.e., the ability to predict job performance) in the shortest possible time frame. Others have developed computer adaptive tests (CAT) that generally take less time to complete. As mentioned earlier, this type of test is based on IRT and test-takers are only presented with enough items from a large bank of possible test items in order to estimate their standing on a particular construct that is predictive of future job performance or training success (Drasgow & Hulin, 1990). The objective of all of these screening tests is only to provide an initial impression of the candidate's suitability. Further testing or gathering of other information will occur when the "short-list" of candidates is invited to visit the organization.

A concern about these types of Internet screening tests is that their unproctored nature allows for cheating. That is why these instruments should only be used to provide a preliminary impression until they can be confirmed by a secure proctored re-examination. Further, vendors of these screening devices are making efforts to rotate equivalent items, scales, testlets, or tests, in addition to randomizing items (or scales, testlets, etc.) to reduce the potential for cheating (Schmit, 2001a). The CAT methodology may also help to reduce cheating, but it requires a large number items calibrated by difficulty and the ability to discriminate among candidates. This information and large numbers of items can be used to

generate an almost infinite number of unique tests, if the item pool is sufficiently large. Development of an item pool of this nature takes a great deal of time and requires many test subjects. This method, however, cannot prevent cheating by examinees in unproctored settings where one or more helpers sit by the side of the examinee and provide assistance.

Internet testing is not limited to unproctored situations. Clearly, proctored versions of Internet testing can be conducted where facilities allow for such activity. There are several large third-party testing facilities in the U.S. that can be contracted by organizations who do not have the technological facilities to conduct this type of testing. This approach, however, is quite expensive and is often only used by organizations for high-level or professional/skilled jobs.

Computer delivered testing, whether over the Internet, through a local network, or on a personal computer, is leading to many innovative types of assessments (Drasgow & Olson-Buchanan, 1999; Olson-Buchanan, 2002). Currently, many larger organizations have access to high-speed Internet connections, which accommodate most of the same types of innovations available on local networks and personal computers. However, any kind of testing that is offered to home users must account for the wide variety of access speeds at which test-takers will connect on-line. Clearly, the emergence of broadband technologies and access points is alleviating this problem, but it will be some time before mass access to these technologies is achieved. Still there are many exciting possibilities emerging in computer testing for use by I-O psychologists, particularly in the development of face valid tools.

The face validity of testing materials has become a major point of focus among the organizational professionals involved in the adaptation of testing programs. As noted earlier, human resource leaders want tests that will serve both as measures of capabilities or potential and as recruiting tools. In other words, the

tests must give a realistic preview of what happens on the job. Video-based simulations are being developed that have this high fidelity quality (McBride, 1998; McHenry & Schmitt, 1994, Olson-Buchanan et al., 1998). Typically, these are high quality portrayals of work situations and problems that the test-taker is asked to respond to in a variety of formats, ranging from open ended-responses to multiple choice decision lists. The more elaborate tests include multiple scenarios that get played out following a specific response. These branching scenarios further measure the test taker's ability to solve problems that he or she may have created through a poor initial decision. Even virtual reality technologies are being explored in the development of these highly realistic tests. The face validity of these real-to-life scenario tests provides the kind of test human resource professionals are seeking.

Another type of test that is growing in popularity is game-based testing (Handler, 2001). This type of test appears on the surface to be a video game, but the decision points in the game are scored and validated using job performance as the predicted criterion. The objective behind this game-based form of testing is to build a test that is stimulating and enticing to the test-taker, while at the same time measuring job-related skills or potentialities. Some vendors of these test products suggest that the games can be used as recruiting tools, whereby potential job candidates are invited or lured to an enticing website where the game is played. They are encouraged to get other friends to play the game as well. From the players, the organization chooses to contact those who attain impressive scores that have been previously found to predict job performance or training success. This type of testing is just beginning to be explored and many psychometric issues require further study. There is clearly a need for future research on this emerging technology.

The technologies that are used for testing in the personnel selection process are also becoming widely used in other I-O interventions and programs that are

moving toward Internet administration. Testing is frequently a part of training and development programs. In this area, simulations, game-based testing, and other face valid testing methodologies find favor among users who want real-to-life methodologies for helping people understand their own and others' cognitions and behaviors. Performance and associated succession management programs often use behavioral data as part of the decision making process for compensation, for other rewards, and for promotions. In essence, this too constitutes testing. That is, in many of these programs I-O psychologists will work with an organization to identify the competencies associated with success on a given job. Then, behavioral measures of success are developed. These measures may include ratings from self, peers, subordinates, customers, and supervisors. These measures are typically referred to as 360-degree measures, because they obtain information from others in order to have many different perspectives from around the target's circle of influence. Other measures used in performance and succession management programs may include simulations of work, work samples, or knowledge tests used to assess the incumbents' skill levels. All of these are examples of tests. These tests, just like those used in employee selection programs, are being administered through the Internet in similar formats and methodologies.

Clearly, Internet testing is moving forward quickly in the field of I-O psychology and offers some unique advantages and challenges. Among the advantages are cost-efficiencies in time, money, space, and human resources; the ability to quickly aggregate and report data world-wide; and the access to a limitless medium for creative development of new testing methodologies. Among the challenges are the potential for cheating and other fraudulent behavior; the easy entry of many non-professionals selling tests that do not meet professional standards; the vast variability of points of entry to the Internet that can cause technological problems and introduce error into the testing process; high

development costs; and situations where not all potential users would have equal access to the Internet. In addition, many vendors are moving toward “black-box” approaches where methods and tools such as job analysis, recruitment systems, pre-screening tools (i.e., questionnaires, application blanks, etc.), selection tests, interviews, and hire decisions are integrated in a fully Internet-based or enterprise system (Corporate Executive Board, 2002) where the algorithms and system interface rules are unknown to users and quite possibly have low psychometric rigor. Still, the Internet delivery of testing will continue to grow in the field of I-O psychology. A recent survey suggests that close to one third of companies use some form of screening, while many more say they will use it in the near future (Wheeler, Foss, & Handler, 2001). Thus, it is the duty of professionals in the field of I-O psychology to both capitalize on the advantages and overcome the challenges to best serve these needs.

Educational Testing Sites

Internet assessment offers unique solutions to some problems inherent in large-scale educational testing. Elementary and secondary schools want to assess children as late as possible in the school year, and they want results of these tests to be available as soon as possible. Use of Internet testing can decrease the amount of time needed to score and report results of achievement testing. Scoring achievement tests with multiple-choice formats and essay responses currently requires human involvement in feeding scanning machines and in reading essays written by students. Optical scanning devices encounter many difficulties such as incomplete erasures, wrinkled response forms, and children omitting their names. All of these problems may be addressed through testing on the Internet. Scoring of multiple-choice format tests on the Internet eliminates the need to scan response forms; test-taker identity is automatically recorded as part of the protocol.

Computer tools for assessing the quality of an essay response (Landauer, Laham, & Foltz, 2002; Shermis & Burstein, 2002) or scoring written responses to math questions (Bennett, Steffen, Singley, Morley, & Jacquemin, 1997) greatly reduce scoring time. Finally, current research is investigating the utility of embedded assessment and the feasibility of assessing a student's ability to interface with computers and the Internet by requiring them to use the Internet to solve a problem (ETS research).

Personality and Psychodiagnostic Testing

Perhaps in part because of the availability of the Internet, but certainly concomitant with its use, there has been a great deal of interest in present-day society in self-exploration, with an emphasis on understanding the issues involved in personal emotional growth. In addition, we live in a society where instant and rapid answers are now often available. The confluence of these factors has resulted in the vigorous pursuit of self-help information, such as personality and temperament tests, intelligence tests, and the like. People have become quite curious about ways in which to understand aspects of their personality functioning and their relationships. Thus, the Internet, with the availability of an array of personality and relationship tests, along with the promise of rapid results provided in response to questions and issues, has become a popular public medium. People can complete these instruments privately, without the complications and the possible stigma involved in consulting a professional psychologist. Although, for the most part, the assessments that are available on the Internet have not been demonstrated to be sound psychometrically, they nevertheless have the aura of authenticity to the lay public, and in some cases have taken the place of "tests" in magazines, Sunday supplements, and self-help books. Undaunted by the complex validation and standardization problems of testing described in this document, a variety of Internet

sites have been introduced, primarily, it appears, by non-psychologists, to provide products for a hungry public.

In the survey of Internet testing described in a previous section, there were some tests, such as the Luscher Color Test, that provide feedback immediately. While the availability of such feedback may be helpful to some, it may be at least disconcerting, if not damaging, to others. For example, on the Luscher test, in which the test-taker is asked to choose, in order, three colors he or she particularly likes, six attempts at choosing various (quite different) combinations of colors all resulted in reports that described significant problems; none contained positive statements.

In addition to a number of general personality measures, there are hundreds of additional assessments that are said to measure specific areas of personality functioning. For example, there are measures of aspiration, power hunger, assertiveness-extroversion, procrastination, time management, leadership ability, perfectionism, anger disorders, type A personality, self-esteem, anxiety, coping skills, depression, jealousy, optimism, sensuality, emotional intelligence, optimism, and character-temperament, to name just a few. Most of these measures are advertised as personal growth instruments, in contrast to those instruments designed to diagnose psychopathology. Most of the personal growth measures take about 15 minutes to complete, the shortest being about 5 minutes and the longest, about 30 minutes. While some Internet assessments were clearly labeled as entertainment, others were presented much more ambiguously as possibly valid tests, whereas still others were promoted as valid. It is possible that some of the measures provided as entertainment or as possibly valid were done so in order to provide the respondent with an excuse should the results not be positive. Such labeling would also reduce a respondent's defensiveness and would make attempting the test safer.

Certainly many people seek affirmation and validation of their self-image; they like to think they know themselves. Others are probably curious about themselves and some are truly troubled. Therefore, the audience for these websites is probably somewhat heterogeneous. While invalid results would not be harmful to many people, some might be adversely affected. The feedback, which may be given in a rather blunt and insensitive manner, provides no opportunity for interaction with a trained professional or even additional clarification or further information. Perhaps the most disturbing website we found was a university based website that bluntly provided on-line diagnoses of "37 common mental disorders" based on the Diagnostic and Statistical Manual (DSM-IV-TR; American Psychiatric Association). On this website, immediate feedback was given once a respondent completed the assessment, presumably as a service. However, given the discussion concerning the APA Ethical Code and test feedback described later in this report, the value of such a procedure may be questioned.

Professional Clinical and Counseling Psychology

Few Internet-based measures are available in the area of clinical and/or counseling psychology at present. A few of the most popular self-report measures (e.g., MMPI-2, Millon) have been available in this format for many years. These measures can be taken on-line and electronically scored; a detailed interpretive report can be obtained within minutes after the administration is complete. These results are not directly available to the patient/client, but are typically interpreted by a trained and licensed psychologist. There has been little progress in developing on-line versions of other popular instruments.

One important problem in this area is that many important measures are poorly suited for Internet administration. While it is conceivable that various projective measures (e.g., the Thematic Apperception Test, the Rorschach Inkblot

Test, various projective drawing tests, the sentence completion test, etc.) could be adapted for Internet use, scoring would require the development of sophisticated tools analogous to Bennett et al.'s (1997) symbolic computation algorithm used to score open-ended math items or Landauer et al.'s (2002) Latent Semantic Analysis used to score essays. Moreover, no information about the patient/client-examiner interaction would be available. Thus, the integration and interpretation of diverse types of data to obtain a comprehensive understanding of the referral issue, the most important and complex task of the psychologist in psychological assessment, is not possible.

Futuristic techniques, such as virtual reality, might fire the imaginations of some. However, at this point in the development of the Internet, such applications seem to be far in the future. Several authors (see Barak & English, in press) visualize the future application of testing websites as the therapist's assistant. Therapists would refer their patients/clients to specific websites to take certain measures. The measures would be scored and the interpretations provided to the therapist, who would interpret them to the patient, or would integrate them with other findings as part of a psychological assessment. This procedure would save time for the therapist and eliminate the need for filing cabinets devoted to storing paper-and-pencil test forms.

A commonly held assumption is that patients/clients may experience an aversive reaction to interacting with an impersonal and possibly difficult to use computer assessment program. This belief would have been supported in the 1970's when research showed that computer based intelligence testing led to increased levels of anxiety with negative affective reactions to poorly designed computer testing procedures (Hedl, O'Neil, & Hansen, 1973). Current research indicates that computers do not necessarily "dehumanize" the assessment process and many individuals actually prefer computer-based assessments to standard methods

(Fowler, 1985). Studies have indicated that 80% of college students preferred taking the MMPI by computer and few students preferred the paper-and-pencil administration. Even phobic patients showed no apprehension when completing a full behavioral assessment by computer (Fowler, 1985). It is important to note that psychologists sometimes glean important information from watching a patient/client complete an assessment; these potentially important observations are unavailable when testing is conducted via the Internet.

Research with vocational education students found that students had positive reactions to computerized testing and preferred Internet testing over standard methods (Barak, 1999). Currently the use of the Addiction Severity Index-Multimedia version (ASI-MV) in a number of state and correctional systems and the Veterans Administration has resulted in uniformly positive patient/client response (Budman, 2000). Researchers are not aware of a single patient/client, regardless of prior computer use, who has reported difficulty in using the ASI-MV, including some individuals in their seventies (Budman, 2000). Research addressing older persons' reactions to computerized testing in comparison to traditional testing revealed there were no striking differences in participant reactions to the various methods (Ivnik, Malec, Tangalos, & Crook, 1996). Not only do individuals show a significant preference for computer-based assessments in lieu of human interviewers, but these favorable attitudes have held through exhaustive computerized assessments of four to ten hours (Nurius, 1990).

While there is ample evidence indicating that individuals may not react negatively to a computer assessment and can handle such administrations with ease regardless of prior experience, there is essentially no research addressing the client's reaction to receiving feedback from Internet assessments. Budman (2000), who strongly advocates computerized assessment, detailed the experience of a colleague who completed some checklists regarding physical health and was told he had a type

of hip cancer. Despite the fact that this individual knew his hip difficulties were a result of a sports injury, he still experienced a feeling of alarm. This type of outcome could raise the potential of liability risks of poorly designed or potentially damaging computer feedback with case scenarios of test takers receiving information that is inaccurate, inappropriately used, and psychologically damaging.

Illustrations and Examples

This section provides examples of Internet assessments. These examples are based on recent experience with actual patients and real applications of Internet testing. They illustrate the variability in the quality of Internet testing.

Am I Depressed?

Joe was a fifteen year-old male who had depressive symptoms and academic difficulty, obtaining mostly D's and F's in school. He also had a history of substance abuse, typically using marijuana, with some LSD and alcohol experimentation. As these symptoms had become problematic in the last two or three years, the therapist planned to complete a psychological examination to assess the need for possible in-patient treatment.

One day, Joe walked into his therapist's office and stated, "I don't need any testing," to which his therapist responded, "Why not?" Joe went on to explain that he and his girlfriend had taken several personality tests on the Internet, and both he and his girlfriend came out "normal." The therapist asked if Joe had told the truth on the tests, and he replied, "Sure, and they came out fine." The therapist explained that tests on the Internet might be different from tests used in the psychological assessment they would be completing together and that the former might not be accurate. Joe was not receptive stating, "Why would they put those tests on the Internet if they weren't the real thing?" The therapist then gave a brief explanation

of the poor validity and lack of standardization of the tests used in the Internet assessment compared with the tests he was going to use and then he described those tests. Joe replied, "That's bull; what does that stuff have to do with your personality? What! Telling stories using old pictures? There's nothing like that on the Internet."

Although Joe completed the tests as requested by his therapist, he replied minimally and the feedback session had little effect on him. As the therapist tried to point out some difficulties Joe was experiencing in life, it was difficult for him to digest this information in light of having received positive feedback from the Internet. The discrepancy between the Internet feedback and the psychological assessment results created some cognitive dissonance in Joe. In order to make a decision as to which psychological description to accept, it appeared easier and more palatable for Joe to identify with the Internet interpretation.

Am I Smart?

Richard was a 31 year-old white male in his second year of an experimental psychology master's program. He was having trouble in some of his classes, and he began to wonder if he was smart enough to be successful. Not knowing his IQ, Richard decided to try to find out by taking an intelligence test on the Internet. It was easy to find a test of intelligence on the Internet, and the test was short and was easy to complete. Richard was, however, quite disappointed when the results indicated he was in the Average intelligence range. Prior to receiving his test results, he said he liked the way the test offered feedback after each question and appreciated the fact that the test seemed to ask a variety of questions in order to assess general knowledge. Besides the fact that the experience was satisfactory, he acknowledged that he resorted to a few rationalizations and justifications to help him deal with his consequent disappointment.

After reviewing each question he answered incorrectly, Richard decided the test was simply too narrow to adequately paint a picture of someone's intellectual functioning. Although it took a bit of energy to reach resolution, he was able to comfortably decide the Internet assessment of his IQ level was simply inaccurate.

Is My Daughter Smart?

Lucy and Richard have a quiet 8 year-old daughter named Betsy who never performed very well in school but her teachers described her as a hard working child who is just very quiet and shy. One night Lucy and Richard saw a short news clip about the Internet and they thought it might be good to find out Betsy's IQ. They went to a site that had a look that instilled confidence and had Betsy complete the test. Lucy and Richard were surprised to see how hard the word analogies and math problems were but they were shocked and disappointed when the results were given. The Internet IQ test results read this way: "You had 2 of 20 correct. That is 10.00%. Hey a random number generator could do better than you! Of the testers your age, the average was 49.12%. Your standardized IQ based on this test would be 68.940." Although the results sounded precise and scientific, Betsy's parents decided to try another site.

Finding another Internet testing site was not difficult and taking this new test proved to be interesting. Betsy immediately responded "Look mom, these questions are different. They are pictures and shapes." It was obvious to her parents that the questions on this test were different from the first test and Betsy seemed to enjoy the variety of the questions. After all the items were answered, the "Ultimate IQ test" gave the result "Congratulations, Betsy! Your IQ score is 93." The report went on to say "what is even better" is that for \$14.95, a more in depth analysis of the results for mathematical, visual-spatial, linguistic, and logical scores could be obtained. In fact, it said, "We compared your answers with others who have taken

the test, and according to the types of questions you got correct, we can tell your Intellectual Type is a "**Word Warrior**." Additionally, "to find out more about how your brain processes information and where your intellectual strengths lie, buy your personalized, 15 page report for only \$14.95."

Lucy and Richard became curious about how they would score, so they also took some of these IQ tests. The results were startling - they got IQ scores ranging from 56 to 147 by responding randomly! Although it took a bit of energy to reach a resolution, it became clear that assessment of Betsy's IQ using the Internet was simply not dependable, and the assistance of a professional was necessary.

Am I Manic?

Bob, a 40-year old businessman, was in need of a psychological examination due to legal issues in relation to a manic episode. Prior to completing his psychological examination with a licensed clinical psychologist, Bob decided to do some research on the Internet. He took a DSM-IV type test, which indicated to him that he was suffering from a serious affective disorder.

He presented himself at his first appointment in a fearful panic. He was practically shaking as he exclaimed, "I think I'm crazy, I took this test on the Internet and I must be crazy!" The therapist spent two sessions with Bob, working on calming him and helping him to understand the feedback he received from the Internet. The therapist focused on redefining the test results with Bob to repair some of the personal damage he experienced as a result of the Internet testing. He had previously been in denial concerning his disorder, and such a blunt response regarding his personality threw him into a terrible and threatened state. The therapist sensitively framed the psychological assessment in a fashion to help Bob gain an in depth understanding of his psychological functioning, providing him with explanations and recommendations for treatment.

Do I have an Anxiety Disorder?

Peter was a 28 year-old single, white male who initially discussed concerns involving Generalized Anxiety Disorder (GAD). He stated his father located the Website describing the symptoms of GAD, and he was curious to learn if he had this disorder. Although he found the Internet site to be helpful, he was disappointed that the information did not encompass all of the symptoms he was experiencing. He was confident in the Internet responses related to his experience, but definitely felt that there was more going on. He was not distressed by his testing experience and was motivated to seek additional services at an outpatient mental health treatment facility. The intake interviewer was given the impression that the Internet testing was a beneficial experience for Peter. Not only did he acknowledge the helpfulness of the feedback, even though it was incomplete, but he also had a name for his experiences and a rudimentary understanding of what he was going through.

Am I Bipolar?

John, a 23 year-old unemployed white male, came to a university-based psychological clinic complaining of intrusive thoughts that something bad was going to occur, and with feelings of hopelessness and worthlessness. He described himself as highly irritable and argumentative, and complained of insomnia. He explained that his aunt found an Internet site to assess Bipolar Disorder. He completed a Bipolar checklist and received affirmation that he was experiencing Bipolar symptoms. While completing the intake interview, he did not appear distressed by the feedback and appeared to accept the information as a fact since the Internet assessment fit his symptoms accurately. John was motivated to seek treatment through the clinic and although his experience was confirmed and given a label, he was open minded to alternative suggestions.

Does Anyone Do Good Internet Testing?

Jean has recently graduated from college and has spent a great deal of time preparing to take the licensing exam for her profession. The exam was developed by the profession's major association, and a passing score is required in all fifty states in order to become a licensed practitioner. Development of the exam began with a comprehensive practice analysis. In the practice analysis, hundreds of professionals were asked about the knowledge and skills that are required of entry-level practitioners to protect the public. Initially, focus groups were conducted and later a stratified random sample of association members completed a questionnaire based on information gleaned from the focus groups. A test blueprint was developed from the practice analysis; every form of the test must satisfy the content specifications contained in the blueprint. Jean scheduled her exam at a secure, proctored test site during one of several testing windows throughout the year. Drawing from a large bank of pretested items, automated test assembly procedures were used to create a test form for Jean that satisfied the test blueprint. Jean's score was computed immediately following the completion of her test; all data were uploaded to the professional association's server in data encryption packets. The professional association communicated the results to Jean's state licensing board. Because she has met all educational requirements and passed the exam, Jean was licensed to practice in her state.

Summary

Except for the last one, these examples do not follow APA guidelines for the provision of assessment information to the client in a way that is responsible, helpful, and unlikely to cause harm. Individuals with varying degrees of intellectual resources and psychological strengths will have variable reactions to Internet

assessment interpretations. Although Internet sites have the potential to provide assessment in a therapeutic fashion and to refer the test-taker to additional resources as necessary, this is not being done consistently or appropriately. Clearly, additional research is needed to understand how Internet assessment feedback influences the test taker. Despite online direction to additional resources, Internet assessments are not appropriate for individuals who require a substantial holding environment. If Internet assessments are to be conducted in a manner beneficial to the individual, we need to improve our clinical understanding of how such assessments should be carried out, with guidelines to protect the interests, rights, and overall mental and emotional state of the test taker.

The final example illustrates the value of Internet testing. Careful test development procedures were used to construct the test blueprint, automated test assembly was used to create a test form for each candidate, the Internet allowed the test form to be administered at a time that was convenient for the candidate, and responses were quickly uploaded via the Internet to the professional association's server in encrypted packets.

Ethical and Professional Issues

Ethical issues abound for psychologists who use the Internet in their practice. Many of these issues are being addressed by specific APA committees (American Psychological Association, 1997). While all of the ethical issues surrounding the use of the Internet in the practice of psychology are important, we will confine our discussion to the issues specifically raised around Internet testing. Our discussion is framed by the current APA Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2003), specifically Section 9, Assessment, which covers most of the issues surrounding Internet testing.

The first ethical issue to be considered is the professional context in which the Internet testing takes place. The associated Ethical Principle is as follows:

9.01 Bases for Assessment.

(a) Psychologists base the opinions contained in their recommendations, reports, and diagnostic or evaluative statements, including forensic testimony, on information and techniques sufficient to substantiate their findings.

(b) Except as noted in 9.01c, 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.

(c) When psychologists conduct a record review or provide consultation or supervision and an individual examination is not warranted or necessary for the opinion, psychologists explain this and the sources of information on which they based their conclusions and recommendations (American Psychological Association, 2003).

The issue raised by Internet testing is how testing is placed into a professional context when conducted on the Internet. Many tests on the Internet are accompanied by little other than some broad statements about the use of the test. Further, test-takers may not read instructions or may ignore disclaimers more than in face-to-face situations (Barak & English, in press). Under these principles, test materials that are posted for self-administration and interpretation on the Internet should be accompanied by a statement to the test-taker that clearly defines the

bounds and limitations of the professional relationship with the client that can be achieved through this medium. This may seem a bit counter-intuitive given the impersonal nature of Internet communications. However, a potential client who is browsing the Internet for professional advice is seeking a trust relationship. Providing preliminary test materials for diagnostic or evaluative purposes therefore implies an offer to form this trust relationship. Thus, the limitations of the relationship that can be developed through an impersonal medium such as the Internet should be clearly described in an opening statement to the test-taker. In addition, test providers may need to make available contact information (e.g., e-mail address, phone number) for those who do not completely understand directions or the purpose of the test. Finally, the limits of the feedback provided to the test-taker following the test should be clearly described both before the test and preceding feedback. This description should clearly describe the potential limitations of conclusions and recommendations that can be made as a result of a very limited and potentially non-personal Internet approach.

The contact with test-takers using the Internet is indirect and there are not sufficient means to confirm that they have understood the instructions and statements about the intended use of the test or how conclusions have been reached or recommendations have been (Barak, 1999). Steps must be taken to explain to test-takers how test information will or will not be used in conjunction with other information to form conclusions and recommendations. Given that testing is only one of many types of diagnostic or evaluative methods available to psychologists, a caveat should be presented with Internet tests that clearly states to the test-taker the limitations of diagnostic, judgment, or predictive information that can be obtained through a single method.

The fitness of the test taker is important as well. How can a test taker and provider ensure a match between the test taker need and the test to be

administered? Psychologists need to develop methods for overcoming these limitations and study the effects of alternative instruction sets and methods for assessing candidate fit with the test being offered. The indirect nature of Internet testing must be overcome with advances in test-taker screening that build both a trust relationship and a profile of client fitness for such testing.

The next area of ethical consideration involves the appropriate use of Internet testing/assessment. The Ethical Principle states:

9.02 Use of Assessments

(a) Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques.

(b) Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation.

(c) Psychologists use assessment methods that are appropriate to an individual's language preference and competence, unless the use of an alternative language is relevant to the assessment issues (American Psychological Association, 2003).

Internet testing, in many cases, has been simply a process of putting paper-and-pencil or computerized tests onto a new medium. However, while research has explored the equivalence of some forms of computerized and paper-and-pencil tests (e.g., Mead & Drasgow, 1993) very little research has been conducted on the equivalence of Internet testing with these other formats. This may call into question the evidence for the usefulness of these tools. Further, tests that may have been developed and researched in a proctored setting are now often being used in an

unproctored context that is facilitated by the Internet and its widening accessibility. This approach calls into question the proper application of the techniques. The effects of both the medium and the context require additional research to ensure appropriate use of tests and assessment on the Internet.

As noted earlier, an advantage to using the Internet to deliver tests is that it may provide greater accessibility and reach than an approach that requires an individual to be at a certain place, at a particular time. This advantage can also create a challenge. Wider access may cause a difference in the populations for which the test was developed versus the ultimate population that has access. For example, a pre-employment test may be specifically developed and researched for a management population. Under more traditional conditions, applicants for these management positions may be required to test at a specific location where a significant effort and commitment is involved. However, the Internet may provide easy access to a different population where a non-qualified candidate could decide that he or she might just take the test on the off chance that he or she might gain entry to an otherwise inaccessible position.

Language can also become a great challenge given the nature of the Internet and its "world wide web." The pervasiveness of the Internet and easy worldwide access create new ethical challenges for psychologists. Any test posted on the Internet immediately becomes available to people around the world. As such, many will experience the test in a second language unless the test provider has made special provisions to provide multi-lingual versions. It will be up providers of Internet tests to ensure that test-takers are aware of the implications and problems associated with use of a tool in a language that they may not fully comprehend. For those who do provide multi-lingual versions of a test, it is incumbent upon them to show psychometric equivalence (e.g., reliability and validity) of these tests as they are used in different populations versus the one in which they were developed.

Normative issues are also a related concern for Internet test delivery. With good intentions, a test may be placed on a website by a psychologist in the U.S., but someone in China may have access to it and complete the test. Feedback may be based only on U.S. norms. An inadvertent, but inappropriate, use of norms is the result. This is clearly an area of great potential for the inappropriate use of tests and associated norms. Psychologists will need to make substantial efforts to collect demographic information prior to testing and to provide feedback only to individuals in groups for which normative data are available.

The next area of ethical consideration involves informed consent. The Ethical Principle states:

9.03 Informed Consent in Assessments

(a) Psychologists obtain informed consent for assessments, evaluations, or diagnostic services, as described in Standard 3.10, Informed Consent, except when (1) testing is mandated by law or governmental regulations; (2) informed consent is implied because testing is conducted as a routine educational, institutional, or organizational activity (e.g., when participants voluntarily agree to assessment when applying for a job); or (3) one purpose of the testing is to evaluate decisional capacity. Informed consent includes an explanation of the nature and purpose of the assessment, fees, involvement of third parties, and limits of confidentiality and sufficient opportunity for the client/patient to ask questions and receive answers.

(b) Psychologists inform persons with questionable capacity to consent or for whom testing is mandated by law or governmental regulations about the nature and purpose of the proposed assessment services, using language that is reasonably understandable to the person being assessed.

(c) Psychologists using the services of an interpreter obtain informed consent from the client/patient to use that interpreter, ensure that confidentiality of

test results and test security are maintained, and include in their recommendations, reports, and diagnostic or evaluative statements, including forensic testimony, discussion of any limitations on the data obtained (American Psychological Association, 2003).

Gaining true informed consent through electronic means is likely to create unique challenges for psychologists. As noted earlier, the impersonal and standardized nature of Internet testing programs are not likely to fit all individuals the same. In other words, it may be very difficult to provide true informed consent to all individuals completing tests through the Internet. In many cases, it will not be known whether the person completing the test is capable of giving informed consent or whether permission is required from a legally authorized person. Take, for example, a pre-teen who poses as an individual over 18 years old or a patient under the legal guardianship of another who gives consent in order to gain access to testing. Psychologists who wish to use testing on the Internet, other than for excepted practices, will need to find ways to deal with this thorny problem of how to authenticate informed consent over the Internet.

The next area of ethical consideration involves the appropriate release of test data. The Ethical Principle states:

9.04 Release of Test Data

(a) The term test data refers to raw and scaled scores, client/patient responses to test questions or stimuli, and psychologists' notes and recordings concerning client/patient statements and behavior during an examination. Those portions of test materials that include client/patient responses are included in the definition of test data. Pursuant to a client/patient release, psychologists provide test data to the client/patient or other persons identified in the release. Psychologists may refrain from releasing test data to protect a client/patient or others from substantial harm

or misuse or misrepresentation of the data or the test, recognizing that in many instances release of confidential information under these circumstances is regulated by law. (See also Standard 9.11, Maintaining Test Security.)

(b) In the absence of a client/patient release, psychologists provide test data only as required by law or court order (American Psychological Association, 2003).

Psychological test data reveal very personal details about human characteristics, behaviors, preferences, and capabilities. This type of data is not only valuable to psychologists, but is also valuable to sales, marketing, political, and other groups who may or may not have the individual's best interest in mind when deciding how to use these data. Therefore, it is imperative that measures be taken to provide secure sites for the collection of psychological test data on the Internet. Without secure sites, test data could be intercepted, corrupted, or changed by unscrupulous data thieves and hackers. There are three major principles of Internet security and psychologists using the Internet for testing should take proactive steps in each of these areas to protect test-takers (Howard, Paridaens, & Gramm, 2001). The first principle is confidentiality, which deals with keeping information from being viewed by unintended readers. Encryption technology is designed to provide confidentiality by scrambling data so that only the appropriate senders and receivers can read the data. The second principle is integrity, which is concerned with keeping information from being altered. Message digests are fingerprints that do not allow the changing of information or at least can detect when information has been changed. The final principle is authentication, which relates to identifying the origins of the data. Digital signatures can provide the authentication through a system of keys that are used both in the sending and receiving of messages to identify the sender as authentic. Given the value and highly sensitive nature of psychological

test data, psychologists should use technologies in each of these areas to secure data.

The rise of complex data privacy laws is another issue that must be addressed by psychologists using Internet testing. Data privacy laws vary, by state, national, and international boundaries. The complexities make this topic too broad to discuss in this article, but psychologists who practice within and across these boundaries must become familiar with the relevant legal considerations before using Internet testing.

Similarly, the practice of psychology is governed by different laws and bodies across state, national, and international lines. The use of the Internet in the practice of psychology is not limited by these same boundaries. Thus, psychologists must consider their legal standing when practicing across these boundaries. Licensing issues need to be fully understood by psychologists before undertaking Internet testing.

The sharing of data and reports is infinitely easier as a result of Internet-based access to databases. People anywhere in the world can access databases anywhere else at the click of a button, given the correct security clearance. This ease of access provides more opportunity than in the past for both the intentional and the inadvertent release of data to unqualified individuals. Safeguards must be put in place by psychologists, in conjunction with information technologists, to avoid the release of data to those who are unqualified to use it. While there are many technological safeguards available to protect data, the psychologist's responsibility goes further. It extends to building lines of communication, to the use of training materials, and other means of safeguarding data that involve the human element that is more likely to result in security breaches compared with the failure of technology.

The next ethical issue of note in the use of tests on the Internet has to do with test development efforts. The Principle reads:

9.05 Test Construction.

Psychologists who develop tests and other assessment techniques use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias, and recommendations for use (American Psychological Association, 2003).

Traditional test construction techniques are appropriate for administration over the Internet in proctored test environments. However, as noted earlier, tests developed in a paper-and-pencil format and researched in monitored and controlled situations cannot be assumed to provide equivalent measurement when administered over the Internet in unmonitored and uncontrolled situations. Therefore, additional studies of test equivalence and norming should be conducted over the Internet with subjects completing the test under conditions that represent those that the intended target population will experience (Epstein, Klinkenberg, Wiley, & McKinley, 2001).

Advances in computer and Internet technologies have provided the medium for many new advances in testing. Video-based simulations, virtual reality, computer adaptive testing, precision measurement of physiological responses and characteristics are examples of technological advances that are shaping the new testing landscape (Olson-Buchanan, 2002). Integrated systems approaches are also beginning to appear. For example, in the area of pre-employment testing, methods and tools such as job analysis, recruitment systems, pre-screening tools (i.e., questionnaires, application blanks, etc.), selection tests, interviews, and hiring decisions are being integrated in a fully Internet-based or enterprise system (Corporate Executive Board, 2002). Each of these advances challenges the existing

psychometric knowledge base (Drasgow & Olson-Buchanan, 1999; Funke & Schuler, 1998). Therefore, it is incumbent upon psychologists to understand the bounds of current psychometric methods and to establish, research, and report on new methods that support emerging technological advances. It would be unethical to develop new measurement tools that cannot be held to existing psychometric standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) without providing arguments and evidence for new or revised standards. Advances in testing spurred by the Internet should be encouraged, but associated advances in psychometric theory may be a mandatory part of this advancement (i.e., in cases where existing models are inappropriate).

The interpretation of Internet test results poses some unique ethical considerations. The relevant Principle is:

9.06 Interpreting Assessment Results

When interpreting assessment results, including automated interpretations, psychologists take into account the purpose of the assessment as well as the various test factors, test-taking abilities, and other characteristics of the person being assessed, such as situational, personal, linguistic, and cultural differences, that might affect psychologists' judgments or reduce the accuracy of their interpretations. They indicate any significant limitations of their interpretations. (American Psychological Association, 2003).

Internet testing will often be conducted in unproctored and in variable environments. Test-takers will likely be in unstandardized settings (e.g., home, library, school) and psychologists will have little or no way of knowing exactly what conditions might exist that could influence or limit interpretations. This problem may be alleviated to some extent by the use of instructions to test-takers, but it is likely that this will only reduce a small amount of irrelevant variability in scores. Further,

when tests are completed in unmonitored situations, there is currently no way to guarantee the true identity of the test-taker (Schmit, 2001b). Thus, psychologists will need to weigh carefully the importance they place on tests administered over the Internet. Confirmatory methods, administration of equivalent forms, or gathering of data through additional methods will almost always be necessary before making anything other than preliminary evaluations, diagnostic, or predictive decisions.

Gaining an understanding the test taking abilities and specific personal characteristic of the test taker poses an even greater challenge given the impersonal approach that characterizes most Internet-based testing and assessment. For example, in pre-employment testing situations, a provider may have no way of know whether an applicant has a particular disability that might affect the test results and invalidate the possible interpretation of those results. Similarly, a test may be posted in English for use in counseling, but the test-taker speaks English only as a second language. Unless the test taker is asked about this condition, the interpretation of results will likely be flawed. The point is that psychologist using Internet testing and assessment must make provisions for understanding the unique needs of test-takers that may ultimately affect the interpretation of results.

In addition, test-takers must be given information that clearly identifies the purpose of the test so that they can determine whether the test is appropriate for their situation. However, this may not be as easy as providing a purpose statement. The test taker will need help understanding if the test is a fit for his or her situation. There will likely be a need for pre-screening the test-taker to help him or her understand if the test or assessment is right for his or her situation.

The next set of ethical issues to be considered with regard to Internet testing involves the use of Internet tests by unqualified persons. The Principle is:

9.07 Assessment by Unqualified Persons.

Psychologists do not promote the use of psychological assessment techniques by unqualified persons, except when such use is conducted for training purposes with appropriate supervision (American Psychological Association, 2003).

The Internet has made it very easy for anyone to publish any kind of material into the public domain. This freedom has led many to assume that anything published on the Internet is in the public domain and can be copied and used by anyone who chooses to do so. These and other Internet crimes are raising significant challenges for many professions (Reno, 2000). Whole tests, scales, and test items posted on the Internet can be copied and used by unqualified people. It is the responsibility of psychological test publishers and authors to keep their works under tight control and to report copyright violations. Most do this well with customers who use appropriate channels to gain access to the materials. However, publishers and authors must scan the web for whole and partial elements of tests that require professional training for administration or interpretation. Partial tests are likely to be the most difficult to identify, yet they may be the most damaging, as the original psychometric properties are likely denigrated. Consistent with Principle 9.11 (cited below) publishers must also protect their copyrights on test materials. It is the duty of the psychology profession to protect the public from unscrupulous vendors who exploit the Internet with tests of others or, worse yet, with bad renditions of the original test.

Principle 9.07 is written in a way that may suggest that psychologists take reactive steps rather than proactive steps in the protection of the profession. However, others have taken more proactive measures to protect the public. For example, a mental health consumer advocacy and education program has stepped up a process for checking credentials of on-line counselors (Ainsworth, 2002). Online therapists can register with this organization and have their credentials (e.g.,

education, experience, background) checked. Therapists who pass this check are issued a special icon for posting on their website. Clients can go to the advocacy group's website to verify the authenticity of the therapist. A similar program could be established by a consortium of test publishers who have or plan to have their psychological test products administered on the Internet.

The next Principle to be considered deals with outdated test materials. It reads as follows:

9.08 Obsolete Tests and Outdated Test Results.

(a) Psychologists do not base their assessment or intervention decisions or recommendations on data or test results that are outdated for the current purpose.

(b) Similarly, psychologists do not base such decisions or recommendations on tests and measures that are obsolete and not useful for the current purpose (American Psychological Association, 2003).

The Internet is full of obsolete and outdated information. Consumers often have difficulty sorting out the current from the outdated pages available on the Internet. Consistent with the discussion of the previous principle, when partial or whole replication of test materials is made through uninformed or fraudulent acts, these test materials are likely to become obsolete or outdated, because the original publisher updates the materials. Further, it is quite easy for web publishers to forget about published pages on the Internet that may be updated in different places, yet the old materials remain available to the public. Finally, psychologists who do not closely watch the literature and other materials from test publishers may inadvertently use outdated materials online. Others may resist change and intentionally use outdated materials. As previously noted, test publishers and authors must carefully monitor the Internet for obsolete and outdated materials and take both proactive and reactive steps to curb and eliminate these practices.

Third-party vendors of Internet tests and associated services also have a set of ethical issues to consider. The relevant Principle is:

9.09 Test Scoring and Interpretation Services.

- (a) 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.
- (b) Psychologists select scoring and interpretation services (including automated services) on the basis of evidence of the validity of the program and procedures as well as on other appropriate considerations.
- (c) Psychologists retain responsibility for the appropriate application, interpretation, and use of assessment instruments, whether they score and interpret such tests themselves or use automated or other services (American Psychological Association, 2003).

The Internet is full of psychological, para-psychological, and pop-psychology tests, as described in earlier sections. Psychologists must find ways to differentiate themselves from the mass of alternatives that do not meet professional standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Providing the information described in this Principle is the first step in overcoming such confusion.

Psychologists who provide tools to other trained professionals should go beyond the simple provision of providing basic psychometric information to potential users. Steps could be taken to show the equivalence of Internet testing with traditional forms of the measure (Epstein, Klinkenberg, Wiley, & McKinley, 2001). Efforts should also be made to provide consultation and training to test users regarding the challenges faced in using tests on the Internet (Barak & English, in press). The training should be specific to tests and populations who will take the tests.

Professional vendors of psychological tests to be used on the Internet may be able to overcome some of the noise of Internet marketing by becoming professional Internet test consultants. Producers of pop-psychology tests should be made to issue more detailed disclaimers, or warnings, describing their tests as entertainment and not as true tests, just as tobacco manufacturers must issue store warnings on cigarette packages.

The advent of technological breakthroughs and the ease of conducting a professional practice that results from these innovations can occasionally blind adaptors to the fundamental qualities that comprise quality tools. Psychologists must learn to discriminate among efficient delivery tools, flashy format, face-valid content, and psychometric quality. All of these qualities may be important to a psychologist in choosing a vendor, but the foundations of psychometrics are still necessary conditions that should be the first hurdle in a multi-hurdle decision process. Technology advances should not be considered in a vacuum when choosing an Internet test.

Perhaps one of the greatest challenges of Internet testing will involve the explanation of results to test-takers. The Principle dealing with this issue is:

9.10 Explaining Assessment Results

Regardless of whether the scoring and interpretation are done by psychologists, by employees or assistants, or by automated or other outside services, psychologists take reasonable steps to ensure that explanations of results are given to the individual or designated representative unless the nature of the relationship precludes provision of an explanation of results (such as in some organizational consulting, preemployment or security screenings, and forensic evaluations), and this fact has been clearly explained to the person being assessed in advance (American Psychological Association, 2003).

Providing feedback to test-takers over the Internet is a topic of concern to many psychologists. There are at least three major ethical issues to consider. First, there are limited ways to understand the conditions under which the test taker completed the test. Did the individual complete the test or did someone else help or do it for him or her? Under what environmental conditions was the test taken? These and many other questions should be answered in order to provide accurate feedback. Second, it is very difficult to provide feedback, particularly negative feedback, to a test-taker without knowing the person's emotional and mental state. The wrong type of feedback could exacerbate the individual's condition. Third, it is difficult to provide test-takers with immediate emotional support in cases where the feedback has traumatic effects on an individual. It is also difficult to know the extent of these reactions in the first place. Given these severe limitations and many other possibilities, psychologists should rarely provide feedback over the Internet. When they do provide feedback, processes for resolving these ethical issues either in "real-time" or within a reasonable time period should be established. Feedback should generally be limited and should include directions for seeking additional information and help through other means. Ultimately, feedback should rely on multiple methods of evaluation to provide assessment results consistent with professional best practice.

Another rather charged area of ethical concern is the maintenance of test security when tests are delivered over the Internet. The Principle covering this set of issues is:

9.11. Maintaining Test Security

The term test materials refers to manuals, instruments, protocols, and test questions or stimuli and does not include test data as defined in Standard 9.04, Release of Test Data. Psychologists make reasonable efforts to maintain the integrity and security of test materials and other assessment techniques

consistent with law and contractual obligations, and in a manner that permits adherence to this Ethics Code (American Psychological Association, 2003).

As the music industry can attest, the ease of posting material on the Internet has led to widespread violations of copyright laws. Many psychological tests and assessments are copyrighted because much effort was expended in their development. These instruments constitute much of test publishers' intellectual capital and must be safeguarded. It is unethical and illegal for unauthorized parties to distribute or use such copyrighted materials. In fact, a quick search of ebay on any given day will produce quick access to many copyrighted and sensitive test materials. For example, a quick search on June 18, 2003 produced the opportunity to bid on an MMPI Manual for Administration and Scoring together with unused testing materials, as well as the opportunity to purchase Rorschach Psychodiagnostic Plates.

In this section, we have made an attempt to raise issues, offer guidance, and delineate some of the ethical issues surrounding Internet testing. While we have touched on many issues, this is certainly not a comprehensive list. We have shown, however, that the current APA Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2003) provide strong guidance for Internet testing. We believe most issues can be resolved by studying these principles and making conservative interpretations that protect both clients and the general public.

Recommendations for the Future

This examination of the issue of testing on the Internet leads to several conclusions. First, and perhaps most important, is that the current psychometric standards, including test reliability and validity, apply even though the way in which the tests are developed and delivered may be quite different. Unfortunately,

because there are many more tests that are now available via the Internet, there is much variability in the quality of these tests. The extent to which there is documented evidence of the reliability and validity of these tests is also quite variable because many Internet tests do not seem to meet standards established by the profession. This puts consumers in the unfortunate position to have the responsibility of evaluating the quality of the information they receive, often with little knowledge and skill to do so. One conclusion is obvious: Internet testing should be subjected to the same defensible standards for assessment tools (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) as paper-and-pencil tests when their results are used to make important decisions. Still, new methods and combinations of methods that are made possible by emerging technologies will push the boundaries of existing psychometric theory and it is up to psychologists to test and expand the limits of psychometrics to keep pace with these innovations.

The Internet provides a tremendous opportunity for testing, and with that opportunity comes a corresponding need for the ethical and professional use of these tests and a responsibility to expand our science to test the usefulness of these interventions. Despite the flash and sparkle of Internet testing, critical questions of the validity of the inferences made from test scores must be demonstrated. This is a fundamental issue of test validity that must be weighed in relation to the ease of availability, cost, and convenience of Internet testing. All these advantages become irrelevant if scores are used in ways that are not supported by evidence of validity. The *Standards for Educational and Psychological Testing* provides extensive information about what is needed to justify a particular use of a test. Internet test developers and test users should carefully read the *Standards* and ensure that their tests are used in appropriate ways.

While the Internet has considerable potential as a means of testing, assessment will require the integration of information obtained via this medium with other relevant information. For this reason, what is typically available on the Internet is testing in contrast to psychological assessment. The test results obtained on the Internet may be inaccurate for a variety of reasons and therefore there must be a professional available to verify the validity of the information and assist in interpretation. Although it is conceivable that future Internet testing methods may approach a psychological assessment, the requirements for appropriate psychological assessment exceed current Internet capabilities. Practitioners must, therefore, be mindful of this distinction and utilize the Internet for its strength and augment it with their assessment skills.

Tests can be placed on the Internet in a manner that suggests authority and conveys confidence, although many of these tests may have little to no documentation of reliability and validity, test-takers often ignore disclaimers that might appear, and self-administered tests can yield inaccurate interpretations. What is needed is considerably more accountability of the Internet site authors so that the user receives the same kind of protections obtained in traditional assessment sessions. Similarly, test developers and publishing companies that enter into Internet testing programs should ensure that Internet tests are held to the same psychometric standards as traditional tests. This would include, for example, documentation summarizing standardization samples, reliability, and validity as well as additional evidence such as equivalence of tests delivered on Internet and paper, uniformity of stimulus quality on different displays, and so forth, to ensure high quality test administration.

There are tremendous opportunities provided by Internet testing. This report has described many of them, and other innovations await discovery. The importance of this new method of testing and assessment is clear, as is the need for formal

guidelines for Internet-based tests and the many ways in which psychologists may use this environment for a variety of applications. We encourage psychologists to think creatively about how their research and practice can be improved by Internet testing. Times have changed as the Internet has brought testing out of the secure environment controlled by a licensed professional psychologist or psychometrician. As testing becomes more accessible, it is important to realize that the principles of good testing still apply and the ethical standards for psychologists are still fundamental. Balancing widespread accessibility with good practice presents a critical challenge to psychologists for the new millennium. There are many issues that await resolution. Over the years to come, much research and critical thinking will be required to address these issues. We believe that psychologists should look forward to this work with excitement and enthusiasm.

References

- Abbate, J. (1999). *Inventing the Internet*. MIT Press: Cambridge, Massachusetts.
- Ackerman, T. A., Evans, J., Park, K. S., Tamassia, C., & Turner, R. (1999). Computer assessment using visual stimuli: A test of dermatological skin disorders. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment*. Mahwah, NJ: Erlbaum. (pp. 137-150).
- Ainsworth, M. (2002). *Credentials check*. Retrieved January 11, 2002 from <http://www.metanoia.org/imhs/identity.htm>
- American Psychological Association (2003). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.
- American Psychological Association (1997). *Services by telephone, teleconferencing, and Internet: A statement by the Ethics Committee of the American Psychological Association*. Retrieved January 11, 2002 from <http://www.apa.org/ethics/stmnt01.html>.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Barak, A. (1999). Psychological applications on the Internet: A discipline on the threshold of a new millennium. *Applied & Preventive Psychology, 8*, 231-246.
- Barak, A., & English, N. (in press). Prospects and limitations of psychological testing on the Internet. *Journal of Technology in Human Services, 18*.
- Bennett, R. E., Steffen, M., Singley, M. K., Morley, M., & Jacquemin, D. (1997). Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computer-adaptive tests. *Journal of Educational Measurement, 34*, 162-176.

Budman, S. H. (2000). Behavioral health care dot-com and beyond: Computer-mediated communications in mental health and substance abuse treatment. *American Psychologist*, *55*, 1290-1300.

Burroughs, W. A., Murray, J., Wesley, S. S., Medina, D. R., Penn, S. L., Gordon, S. R., & Catello, M. (1999). Easing the implementation of behavioral testing through computerization. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment*. Mahwah, NJ: Erlbaum. (pp. 221-247)

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 577-636). Palo Alto, CA: Consulting Psychologists Press.

Drasgow, F., & Olson-Buchanan J. B. (Eds.). (1999). *Innovations in computerized assessment*. Mahwah, NJ: Erlbaum.

Drasgow, F., Olson-Buchanan, J. B., & Moberg, P. J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment*. Mahwah, NJ: Erlbaum. (pp. 177-196)

Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior*, *17*, 339-346.

Erlanger, D. M., Saliba, E., Barth, J., Almquist, J., Weberight, W., Freeman, J. (2001). Monitoring resolution of post-concussion symptoms in athletes: Preliminary results of a web-based neuropsychological test protocol. *Journal of Athletic Training*, *36*, 280-287.

Finholt, T.A., & Olson, G.M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, *8*, 1-9.

Forrester Research (2001, November). *eCommerce will prevail through the economic downturn in 2002, according to a new consumer survey from Forrester*

Research. Retrieved January 11, 2002 from

<http://www.forrester.com/ER/Press/Release/0,1769,651,00.html>

Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology, 53*, 748-759.

Funke, U. & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection & Assessment, 6*, 115-123.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: LEA.

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Nijhoff.

Handler, C. (2001, November). *Simulations: A look into the future of scientific screening*. Retrieved January 11, 2002 from

<http://www.ereexchange.com/articles/db/49F4A7DBDBFB4F9CACCAE159799ACF6E.asp>

Handler, L., & Meyer, G. (1998). In L. Handler & M. Hilsenroth (Eds.), *Teaching and learning personality assessment*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Harris, M. & Dewar, K. (2001, April). *Understanding and using web-based recruiting and screening tools: Key criteria, current trends, and future directions*. Pre-conference workshop conducted at the 16th Annual Conference of the Society for Industrial-Organizational Psychology, San Diego, CA.

Heaton, R. K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D., Grant, I. (2001). Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology, 16*, 75-91.

Hedl, J., O'Neill, H., & Hansen, D. (1973). Affective reactions towards computer based testing. *Journal of Consulting and Clinical Psychology, 40*, 217-222.

Howard, B., Paridaens, O., & Gramm, B. (2001). Information security: Threats and protection mechanisms. *Alcatel Telecommunications Review*, 2nd Quarter, 117-121.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.

Internet Software Consortium. (2001, July). *Internet domain survey*. Retrieved January 11, 2002 from <http://www.isc.org/ds/WWW-200107/index.html>

Ivnik, R. J., Malec, J. F., Tangalos, E. G., & Crook, T. H. (1996). Older persons' reactions to computerized testing versus traditional testing by psychometrists. *Clinical Neuropsychologist*, 10, 149-151.

Landauer, T. K., Laham, D., & Foltz, P. W. (2002). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M. D. Shermis & J. Burstein (Eds.) *Automated essay scoring*. Mahwah, NJ: Erlbaum.

Lloyd, M. G. (1996). Have you had a long-distance therapeutic relationship? You will. *Ethics and Behavior*, 6, 169-172.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Matarazzo, J. D. (1990). Psychological assessment versus psychology testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist*, 45, 999-1017.

McBride, J. R. (1998). Innovations in computer-based testing. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 23–39). Mahwah, NJ: Erlbaum.

McHenry, J. J., & Schmitt, N. (1994). Multimedia testing. In M. G. Rumsey and C. B. Walker (Eds.). *Personnel selection and classification*. (pp. 193–232). Hillsdale, NJ: Erlbaum.

Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449-458.

Mitchell, T. W. (2001, April). *Using the Internet to advance biodata for selection Programs*. Pre-conference workshop conducted at the 16th Annual Conference of the Society for Industrial-Organizational Psychology, San Diego, CA.

Nielsen//NetRatings (2001). *Hot off the net*. Retrieved January 11, 2002 from http://www.nielsen-netratings.com/hot_off_the_net.jsp

Nielsen//NetRatings (2003). *Monthly usage statistics*. Retrieved May 1, 2003 from http://www.nielsen-netratings.com/news.jsp?section=dat_to

Nurius, P. S. (1990). Computer literacy in automated assessment: Challenges and future directions. *Computers in Human Services*, 6, 283-297.

Olson-Buchanan, J. B. (2002). Computer-based advances in assessment. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations* (pp. 44-87) San Francisco, CA: Jossey-Bass.

Olson-Buchanan, J.B., Drasgow, F., Moberg, P.J., Mead, A.D., Keenan, P.A., & Donovan, M. (1998). Conflict resolution skills assessment: A model-based, multi-media approach. *Personnel Psychology*, 51, 1-24.

Pew Internet & American Life Project (2003a). *The ever-shifting internet population*. Washington D.C.: Author.

Pew Internet & American Life Project (2003b). *Internet activities*. Retrieved May 1, 2003 from http://www.pewinternet.org/reports/chart.asp?img=Internet_A8.htm.

Reno, J. (2000, September). Statement by the Attorney General. Symposium of the Americas: *Protecting intellectual property in the digital age*. Retrieved January 11, 2002 from <http://www.usdoj.gov/archive/ag/speeches/2000/91200agintellectualprop.htm>.

Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.). (1997). *Computerized adaptive testing: From inquiry to operation*. Washington, DC: American Psychological Association.

Segall, D. O. (2001). ASVAB testing via the Internet paper. Unpublished manuscript. Segall, D. O. (1997). Equating the CAT-ASVAB. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 181-198). Washington, DC: American Psychological Association.

Schmit, M. J. (2001a, April). Assessments on the Internet. In Michael Harris (Chair), *The Internet and I-O Psychology: Applications and issues*. Symposium conducted at the 16th Annual Conference of the Society for Industrial-Organizational Psychology, San Diego, CA.

Schmit, M. J. (2001b, September). Use of Psychological Measures for Online Recruitment and Pre-Employment Selection. In L. Frumkin, (Chair), *Internet-Based Assessment: State of the Art in Testing*. Symposium conducted at the 109th Annual Conference of the American Psychological Association, San Francisco, CA.

Shermis, M. D. & Burstein, J. (Eds.) *Automated essay scoring*. Mahwah, NJ: Erlbaum.

U.S. Department of Commerce (February, 2002). *A nation online: How Americans are expanding their use of the Internet*. Washington, D.C.: Author.

Vispoel, W. P. (1999). Creating computerized adaptive tests of music aptitude: Problems, solutions, and future directions. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment*. Mahwah, NJ: Erlbaum. (pp. 151-176).

Wheeler, K. B., Foss, E. E., & Handler, C. A. (2001). *Screening and assessment: Best practices*. Fremont, CA: Global Learning Resources, Inc.