The Evolution of Psychological Testing: Embarking on the Age of Digital Assessment

Hadas Pade, Psy.D.

Chapter 17 in *The Changing Faces of Therapy: Innovative Perspectives on Clinical Issues and Assessment*
Ron Valle, Ph.D. and Jon Klimo, Ph.D. (Editors)

In a dramatically changing world, clinical psychology does not have the choice of standing still (Casper, 2004). Technology has become a fixture in our daily life, and psychological assessment, a cornerstone in the field of mental health, must also keep up with changing times. As a psychologist who regularly conducts testing, I am accustomed to incorporating advanced technology into my work on a regular basis. I have been utilizing computer administration, scoring, and interpretive report options for some time now, but more recently I was especially intrigued to read about the newest developments in psychological testing. There is a new digital format now available for a multitude of well-recognized and frequently utilized measures known as Q-interactive that brings psychological assessment into a new era.

The goals of this chapter are to inform fellow professionals in the field, address some central questions, and begin a discussion regarding the evolution of psychological testing and the potential impact this inevitable shift towards digital modalities of testing will have on the field of mental health. My perspective comes from my multiple roles as a psychologist, including conducting forensic-based and psycho-educational evaluations, working as an

assessment instructor at the graduate level, and serving as a clinical assessment supervisor.

This chapter reviews the increased role of computers in psychological testing and emphasizes the most recent innovations in detail. I attempt to briefly address a variety of issues that are likely to be influenced including examiner and examinee experience, individualization of assessments, results and related diagnoses, and treatment recommendations, efficiency, clinical training, and the overall assessment processes. Additional issues discussed are new ethical considerations as well as validity and reliability of familiar measures, now in a less familiar format. Finally, the future opportunities in psychological assessment that were not possible before are described.

Electronic and digital options in daily personal and professional lives are quickly becoming the preferred way of functioning for many. In fact, today's children and adolescents are considered *digital natives* as they were born into a technological world. Whether we like this shift or not, or have adapted to it in our own lives, the fact is that this is the new reality, with digital, electronic, online, and computerized format communications and interactions here to stay. The continued growth of the internet and its expanding commercial applications has also increased the potential for psychological assessment (Butcher, Perry, & Hahn, 2004). Computers are now being used for test administration, item scoring, transformation of scores, test interpretation, and data storage, as well as for rendering clinical judgments and report writing (Butcher, 2013; Garb, 2000a;

Snyder, 2000). In addition, complex data analysis necessary for many currently available measures necessitates a computer (Butcher, Keller, & Bacon, 1985). In fact, computers have become a powerful asset for psychological assessment and will likely continue to enhance and advance our work in the field (Butcher et al., 2004). However, before describing the remarkable technological advances in psychological testing over the past few decades and related impact, it is helpful to take a look back to the more humble beginning of psychological testing and what has occurred over the years that has led us in this direction.

A Brief History of Psychological Testing

Few would argue with the fact that "psychological testing has firmly established itself as a diagnostic procedure in clinical practice" (Hunt, 1946, p. 311), and that psychological assessment has an extensive history within the mental health field (McIntire & Miller, 2007). But, long before computers, psychological testing became increasingly recognized throughout the world from its initial inception, to standardization of administration procedures, to scoring and statistical analysis, and, of course, to the interpretation of results.

In the Beginning

While Alfred Binet is generally regarded as the first psychologist to construct a modern intelligence test, the term "mental test" was first introduced in 1890 by James McKeen Cattell (Geisinger, 2000). According to Boake (2002), however, tests assessing cognitive and perceptual abilities had been developed well before then. "The first widespread use of psychological and educational

assessment occurred in China more than 3,000 years ago. Measures of problem solving, visual spatial perception, divergent thinking, and creativity were used somewhat commonly" (Oakland, Poortinga, Schlegel, & Hambleton, 2001, p. 4). According to Oakland et al. (2001), it was not until the 20th century that several other countries duplicated these assessment practices and developed measures on a national scale. A variety of social and economic conditions during the industrial revolution within Europe and the United States gave rise to the increased need for such tests.

European Milestones in Psychological Testing

Oakland et al. (2001) point out that in Europe, German psychologists, including Fechner, Weber, and Wundt, established the scientific foundation for psychology, and in England, Galton and his colleagues conceptualized mental abilities in the context of evolution and stressed the importance of individual differences. The German experimental psychologist Wilhelm Wundt was a teacher to several well-known names in today's psychological assessment world, including Charles Spearman, Victor Henri, Emil Kraeplin, Edward Titchener, Stanley Hall, and James Mckeen Cattell (Geisinger, 2000) mentioned earlier. Wundt stated that "the administration of measures must be strictly controlled so that they can be interchangeable across individuals" (Geisinger, 2000, p.117).

In the early 1900s, following Wundt's thinking, Binet, Terman, and Thorndike addressed standardization in testing procedures providing the foundation for testing procedures utilized today. "The logic of testing follows

the paradigm of the experimental method so that the variance in the test scores is reflective of intra-individual differences rather than any differences in test administration" (Geisinger, 2000, p.117). In 1905, Binet and Simon published an intelligence scale developed for use with Paris school children that served as a model for intelligence tests developed in years to come (Boake, 2002). The Binet-Simon scale was widely used in North America and Europe and, following World War I, several revisions were made to further advance the scale better known today as the Stanford-Binet (Boake, 2002).

Psychological Testing in the United States

While schools provided the initial platform for cognitive and intelligence testing in France, it was the military and state of war that helped further develop testing throughout the world, and particularly the United States (Boake, 2002). In the United States, group tests of mental abilities during World War I were helpful in selecting soldiers. Such measures were found to be psychologically sound and led the way towards individual intelligence scales with significant advances in statistical analyses and understanding of results (Oakland et al., 2001). In the 1930's, intelligence testing saw a period of aggressive growth. "The entry of the United States into World War II (WWII) created the need for new individual intelligence tests for screening and assigning recruits" (Boake, 2002, p. 399). The Wechsler-Bellevue scale was developed, where technical advances at the time allowed for more sophisticated statistical analyses that soon made it a dominant adult individual intelligence scale (Boake, 2002). From a

historical perspective, the Wechsler-Bellevue Intelligence Scale represented a particular approach to cognitive assessment between the 1880s and World War I. The main contributions of the Wechsler's subtests included a variety of technical innovations such as standard deviation scores and combining verbal and performance tests into a single scale. These were just a few of the Wechsler-Bellevue scale features that have been replicated by intelligence tests developed later (Boake, 2002).

Further Advances in Psychological Testing

Aside from intelligence tests and other cognitive measures, a multitude of personality and psychopathology tests were also developed beginning in the 1920s. Weiner and Greene (2007) noted that "awareness of individual differences among people is almost as old as civilization itself . . ." (p. 3). The initial Rorschach inkblots were designed in Switzerland in 1921, and in the years that followed, various scoring systems were developed along with updated norms and interpretative guidelines. In the 1930s, another well-known projective measure, the Thematic Apperception Test (TAT), was introduced (Groth-Marnat, 2009). Among the most recognized personality self-reports is the Minnesota Multiphasic Personality Inventory (MMPI) which was developed in the 1940s in order to serve as an aid in assessing adult patients during routine psychiatric care, and to determine severity of symptoms as well as an estimate of change. Theodore Millon developed multiple scales starting in the early 1970s to

measure dimensions of personality and assist with the diagnostic process (Groth-Marnat, 2009).

Just as with cognitive measures, the era of WWII contributed to the increased use of personality and emotional functioning tests as well. The emergence of such measures not only helped define and expand the role of psychological evaluations in the mental health field, but some also led the way in terms of computerized assessment. Oakland et al. (2001) summarized the increased importance of testing and related issues in the early years:

> The ability of those who use tests to accurately describe behavior and to use this information for making important decisions became widely recognized following WWII. This knowledge spawned greater public confidence in and demand for test use, the development of additional tests, as well as various professionals prepared to use them. The public, national governments, and persons interested in clinical and institutional service began to demand that graduate programs prepare applied psychologists and that tests and other technology needed by them be developed. (p. 6)

Establishment of Ethical Guidelines

As psychological tests grew in popularity, the important need for the ethical use of such measures was recognized early and multiple national and international efforts have been made to address this issue. The International Test Commission (ITC) was established in 1976 to support the ethical development and use of psychological testing across the world (Coyne & Bartram, 2006a). In the United States, the American Psychological Association (APA) (2010) developed standards for the psychological use of tests, and the current APA guidelines under Standard 9, including recent 2010 amendments,

address a multitude of issues relevant to the practice of assessment. Such topics include the basis and use of assessment, informed consent and release of test data, test construction, interpreting of results, qualifications, and competency of practice, as well as test security (APA, 2010). The APA code also includes particular guidelines for clinicians using automated test scoring and interpretation services stating clearly that it is the psychologist's responsibility to properly utilize and apply the information obtained. Thus, along with clinicians in the field, such ethical guidelines accepted the growing trend of computerized assessments.

Transition into the Computer Age

Geisinger (2000) described the growth and impact of psychological testing that have been clearly documented for the past 100 years. Measures have been regularly expanded and updated in a multitude of ways to keep up with changing times. Multiple authors, however, have commented on the slow and limited progress made over the years with respect to psychological testing. Starting in the 1940s, Hunt (1946) noted that "anyone surveying the tremendous development of clinical psychology during the last 10 years, and the major importance that psychological testing assumes in such clinical practice, one cannot help but be struck by the small amount of progress we have made in developing our psychological tests as diagnostic instruments" (p. 311). Hofer and Green (1985) suggested that assessment was certainly ready for improvement, as it had been relatively stagnant for years, and that there is no

question that computers offer opportunities for test development that simply could not exist without the new technology. Boake (2002) stated that "...in the midst of accelerating scientific progress, advances from early intelligence tests to those of today have been relatively minor" (p. 383). Furthermore, considering the scientific progress in the last century, it is surprising how slow change has been in terms of the technology of individual intelligence testing. In fact, until recently there have been significantly less advances in psychological testing when compared to many other areas (Boake, 2002).

Such statements are accurate, especially considering the reliance on unchanged paper-and-pencil administration of most measures. However, in recent years the role of computers has increased significantly in the area of psychological testing, and computers have resulted in more sophisticated and advanced psychometric techniques that contributed to, what some might consider, a radical change in the way testing is done (Coyne & Bartram, 2006a). Not only are computerized tests used for administrative tasks and research purposes, but also for direct clinical applications (Simola & Holden, 1992). In fact, computers have continuously increased their role in clinical assessment, interviews, diagnosis, instruction, treatment interventions, clinical consultation, and psychiatric interviews (Groth-Marnat, 2009).

The Role of Computers in Psychological Assessment

Since the introduction of computers into the field of psychology, the question of whether to use a paper-and-pencil or computerized version of tests

has been asked in multiple research studies and settings in the field of mental health. The advances in computer technology in the 1980s presented many possibilities for the design, administration, and scoring of tests, which meant that computers became an integral part of the entire testing process, rather than simply providing a means for scoring tests, as was the case initially (Foxcroft & Davies, 2006). Smoline (2012) indicated that "computers bring the possibility to use tests with a very complex structure without the lack of test validity, to implement quick and deep mathematical processing of each response during the test as well as overall test result..." (p. 205). Modern software used for psychological assessments requires significant modifications, and developers created special descriptive languages to accommodate diverse test structures and items along with complex scoring formulas. This allowed for easy use for psychologists without special technical skills (Smoline, 2012).

Research initially focused on a variety of personality self-report measures that were the first to be transformed in this new digital age (Groth-Marnat, 2009). Multiple sources have indicated the enhanced efficiency of computer-based testing, and many more make a clear argument for the role and benefits of computers in the field of psychological testing and assessment (Burke & Normand, 1987; Butcher, Perry, & Atlis, 2000; Butcher, Perry, & Hahn, 2004; Foxcroft & Davies, 2006; Greene, 2011; Groth-Marnat, 2009). Butcher et al. (2004) stated that "computerized testing methods have long been regarded as a potentially powerful asset for providing psychological assessment services" (p.

331). In addition, computerized testing improves the reliability, standardization, and objectivity of test administration by administering items the same way each time (Butcher, 1987). It is difficult to deny that computer administration has multiple recognized benefits such as speed, reduced cost, and increased accuracy; thus, the popularity of this kind of administration has been consistently growing within multiple mental health settings (Hays & McCallum, 2005).

The first MMPI computer program appeared in the 1960s by John Pearson and Wendell Swenson, (Butcher et al., 2004). During that same decade, the first Rorschach interpretation program by John E. Exner became the precursor for modern computer programs developed in later years that are currently widely used (Butcher, 2013). Continued advances using computer technology have been most evident in the area of self-reports, with improved statistical analysis for scoring and interpretation purposes, as well as increased administration options. Assessments using structured interviews for purposes of diagnosis and treatment planning have also increased utilization of computers (Garb, 2007). Test publishers have developed several systems to make computerized measures readily available, for example, Pearson's Q-local and Q-global (2014, Psychological Assessment Resources' PARiConnect (2014), and the Multi-Health Systems (MHS) Online Assessment Center (2014). However, the question of which modality might produce more accurate and beneficial results, and whether traditional and electronic versions are comparable, continues to be addressed.

Paper-and-Pencil versus Computer Tests

Butcher et al. (1985) state that “objective tests are the most easily computer-adaptable personality assessment devices because both the test stimuli and response options are structured and standardized" (p. 805). A critical issue is the examination of the equivalence between traditional (i.e., paper-and-pencil) and computerized forms of the same instrument. According to Hays and McCallum (2005), different formats may impact results, as well as various factors related with each format. For example, anxiety or lack of comfort with computers may have a negative impact on a computer-administered self-report but not on a paper-and-pencil test. This is critical as the same individual may have different results on the same measure that may in turn influence recommendations for a treatment plan and possibly diagnostic considerations (Hays & McCallum, 2005).

In general, while there have been few inconsistencies in the results of some studies, more often than not, research has demonstrated that conventional and computerized instruments are essentially equivalent (Schulenberg & Yutrzenka, 2004). A variety of studies compared computer administration to paper-and-pencil administration of multiple well-recognized and utilized self-report measures including the MMPI-2 and MMPI-A. A meta-analysis of 14 studies concluded that administering the MMPI by computer instead of paper-and-pencil versions had little impact on scores (Butcher et al., 2004).

According to Hays and McCallum (2005), there are actually some suggested advantages for the MMPI-2 computer version compared with the paper-and-pencil test, including a preference for computer modality by many test-takers, increased scoring accuracy, cost efficiency, and greater objectivity. Disadvantages include some negative responses to the computer format, responses that are impacted by particular factors such as lower education level, certain psychiatric diagnoses, and the possibility of technical difficulties (Hays & McCallum, 2005). Clearly, such comparison studies are important since it cannot be assumed that such tests and methods are always equivalent.

Further research concerning computerized tests of spatial abilities and paper-and-pencil versions suggests that they are comparable, and that the computer modality might potentially be superior since the computer is able to be more precise when assessing the various factors that can affect these abilities (Butcher et al., 2004). Not all measures covering neuropsychological functioning, however, showed such results, and some caution is indicated when using computerized assessments in this area. In fact, some studies determined that computerized testing could not detect brain damage as well as clinical judgment (Butcher et al., 2004). Despite such data, it is likely that, as computerized measures continue to advance, any such limitations would be reduced. It is also important to consider how norms were determined, recognizing that most were based on paper-and-pencil methods and this has a potential impact on equivalency (Butcher et al., 2004). This concern will also

become less relevant with time as electronic testing options continue to increase and corresponding norms will be more readily available.

A variety of advantages and disadvantages of both testing modalities have been described throughout this section. An additional benefit of computer administration is that, unlike paper-and-pencil tests, computers can be more flexible, offering the option of adapting the test to suit the needs and preferences of the examinee, instead of administering all of the test items (Butcher et al., 2004). The issue of computerized adaptive testing is discussed later in this chapter. Overall, the decision to use one format or the other should be based on the specific needs, abilities, and circumstances of the examinee, the credibility of the equipment available, and the setting, to name a few, rather than simply assuming the definitive superiority of one over the other (Hays & McCallum, 2005). Such issues represent added ethical responsibilities of clinicians utilizing computerized assessments.

Ethical & Legal Issues

With great technological advancement come new and increased responsibilities for test developers, publishers, and the clinicians using testing. Hofer and Green (1985) state that "in any rapidly changing field, the requirements of good practice are uncertain simply because there is no precedent for a novel application" (p. 836). Aside from the direct and indirect clinical issues involved, the increased use of computer-based and internet-delivered testing has raised a number of ethical and legal issues (Foxcroft &

Davies, 2006). Coyne and Bartram (2006a) sum up the current situation in the field:

> The area of Internet-Based testing has seen rapid technological and scientific advances in recent years. Superior and more reliable hardware and software features, ability to test on a global stage, and the use of more advanced psychometric techniques have all contributed to a radical change in the way that testing is done. Although these advances have clear positives for the science of testing and for the practice of testing, [however,] they have moved at such a pace that issues of best practice and ethical testing have tended to be left behind. (p. 115)

Computer applications in the field of psychological test administration have raised significant ethical implications for clinicians, clients, test-takers, computerized test construction, and test administration. Lack of awareness of computer-related issues may undermine clinicians' ability to ethically perform computerized psychological assessments. Graduate training in computerized testing is limited, and clinicians need to be exposed to ethical concerns, potential judgment errors, and possible pitfalls in evaluating computer-generated reports. The APA's (2010) guidelines for computer-based testing and test interpretations address some of those concerns.

Additional relevant ethical factors when utilizing computers in psychological testing include technological, quality, control, and security issues (Coyne & Bartram, 2006a) leading to various other guidelines. These include the Guidelines for Computer-Based Testing and Internet-Delivered Testing of the ITC (2006) (Foxcroft & Davies, 2006), the Association of Test Publishers 2000 standards in their Guidelines for Computer-Based Testing (Weiner & Greene,

2007), the 1999 Standards for Educational and Psychological Testing (the joint standards authored by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education or the AERA Standards), the American Psychological Association's 2002 Ethical Principles of Psychologists and Code of Conduct, and the Association of State and Provincial Psychology Board's 2001 ASPPB Code of Conduct (Weiner & Greene, 2007). It should be noted that these standards vary in the degree of specificity with which they address issues relating to computerized assessment.

Computer scoring and interpretive programs are readily available for almost all clinicians including those who lack the necessary knowledge, skills, and competency to utilize them ethically and responsibly. Thus, this level of accessibility and ease of computer programs increases the potential for clinicians to misuse computer test interpretations (Schulenberg & Yutrzenka, 2004) and in some cases; such use may actually lead to potential harm for the client. In this context, Hofer and Green (1985) reviewed several sources of ethical guidelines relevant to computerized psychological testing and issues of competency. They explained that while psychologists do not necessarily need to know and understand all the computer algorithms that are associated with testing software, they do need to demonstrate a solid understanding of the underlying foundations of the test and relevant factors that might impact the computerized modality.

Adding to the risk of misuse of computer test interpretations is the fact that the validity for many computerized test interpretation packages has not been definitively established, and some may be biased (Garb, 2000b). Garb noted that some interpretive packages may produce reports that are more accurate for certain racial or ethnic groups. Overall, validity studies on test interpretation software are relatively rare for most programs (Snyder, 2000). Thus, Hofer and Green (1985) feel that test developers need to be held to high standards. Along with the many ethical considerations and complications, another well-recognized and potential challenge to consider when using computerized assessments is the lack of computer familiarity by the user, often with related anxiety.

Computer Anxiety

Even though computers are highly common in today's society, as computers were gaining notoriety in the world of psychological assessment, many individuals may have felt intimidated by such technology (McIntire & Miller, 2007). A test-taker's attitude toward technology may play a crucial role during computerized psychological test administration. Spinhoven, Labbe, and Rombouts (1993) found that clients with less computer familiarity and negative computer attitudes did not derive as much satisfaction and were not as relaxed as those with more positive attitudes and greater computer familiarity. Attitudes toward technology appear to have a significant potential for influencing interactions with computers, and are thus likely to affect testing results and thus

not necessarily provide an accurate representation of the individual's functioning. Fortunately, research suggests that most individuals overcome such technology-related anxiety fairly easily, and training, or simply continued exposure, helps that process (Hofer & Green, 1985).

Foxcroft and Davies (2006) point out that "although research generally indicates that test-takers with little or no computer familiarity can learn to effectively use a testing system, not enough research has been done on test-takers who are not familiar with technology in general" (p. 178). In addition, some studies found that reduced performance on a computerized version of a paper-and-pencil test was not necessarily related to computer anxiety or computer familiarity (Dimock, 1991). It is also possible that the novelty of the task along with format differences play a significant role (Dimock, 1991). An example of a format difference would be presenting one question at a time on the computer compared to simultaneously presenting multiple questions on paper. Thus, additional research in this area would be helpful. It might also be interesting to explore situations where a person experiences test anxiety but is actually comfortable and familiar with computers. Could taking a test on a computer potentially lessen test anxiety and enhance one's performance? Ultimately, clinicians who do not make adequate accommodations for persons who experience computer aversion risk inaccuracies in their interpretation of the test results.

Reflections on Current Issues

A strong relationship between computer use and psychological assessment has been established. Butcher et al. (2004) note that "ever since computers were first introduced and adapted into the field of assessment psychology in the 1950s, they have been a valuable aid for scoring, data processing, and even interpretation of test results" (p. 331). Given this connection, it is important to consider how computers help us in terms of diagnosing and better serving our clients in general (Caspar, 2004). Based on more recent reports, the use of computers has apparently been most common in the area of personality assessment with its use of various self-report measures. According to Groth-Marnat (2009), by 1990, 17% of practicing psychologists frequently used computer-generated narratives, with an additional 36% using them occasionally and a decade later the number of frequent users seemed to have doubled.

Computers initially aided with developing and better understanding the psychometric foundations of such tests via more sophisticated statistical analyses. Later on, computers were used to score such measures, and provide quicker and more accurate results reflecting an examinee's level of functioning in an easy-to-follow visual handout. As technology further progressed, computer programs were developed to provide interpretive descriptions of test results and such interpretive reports have become more and more popular with clinicians in various settings.

However, even after 50 years of development, many computer-based interpretive reports are still considered to be broad and generic descriptions, rather than integrated psychological reports (Butcher, 2013). In fact, Exner (2005) cautioned about using computer-based interpretive programs noting the lack of ability to think and integrate data, and a disregard for the complexity and uniqueness of each human being. Similarly, Butcher et al. (2004) stated that clinicians need to be cautious when using computerized psychological reports: "Clinicians should not blindly rely upon them or treat their conclusions as true or implicit. They are not intended to replace shrewd clinical observation and judgment or to allow clinicians to circumvent the process of thoroughly integrating all available data . . ." (p. 335).

It is important to note that many of the *computer-based assessments* discussed thus far may assign control of the testing process to a computer. However, *computer-assisted assessment* refers to an assessment procedure where the clinician maintains control over administrative procedures, and the digital format is used to aid in the process, rather than replacing it all together (Scheller, 2013). Computer-assisted assessment is a more recent term and it is utilized in reference to the newest development in testing, the Q-interactive system, that has transformed multiple psychological measures into digital formal and represents the next step in the evolution of psychological testing.

Digital Assessment: The Emergence of Q-interactive

Development of a digital version of several psychological measures addressing various areas of intelligence and cognitive functioning has been going on for nearly a decade (Cayton, Wahlstrom, & Daniel, 2012; Scheller, 2013). In 2012, Pearson, one of the largest psychological test publishers, completed the initial digital adaptation of one of the most widely used intelligence measures, the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV) via a system titled Q-interactive (Cayton et al., 2012). Q-interactive is a Pearson system for computer-assisted, individually administered tests. The Q-interactive system is designed to make assessments more convenient and accurate, to give the clinician easier access to a larger number of tests, and, eventually, to support new types of tests that cannot be administered or scored without computer assistance (Daniel, 2012a). The necessary resources for conducting such a computer-assisted assessment include a Q-interactive account, a computer with access to the internet, two tablets connected via Bluetooth, and the Q-interactive software (Cayton et al., 2012). Examiners can then follow the readily available directions to download the measures and subtests that they wish to administer in advance, and tasks are scored and stored in real time. There are several webinars (www.helloq.com) currently available that demonstrate the process and the steps necessary to conduct a digital assessment.

According to Cayton et al. (2012) and Scheller (2013), Pearson Assessment set out to adapt currently utilized measures so that previous research and norms

would still be relevant so digital measures would be considered interchangeable with the original format. A main goal was to make the assessment process simpler and more efficient by reducing the amount of time required for testing, including preparation, administration, and scoring. In addition, Q-interactive attempts to make administration and scoring easier and more convenient by utilizing computer software that also increases accuracy.

Scheller (2013) noted that such a digital format also reduces the amount of information the examiner often needs to memorize or refer back to during testing, including start points, discontinuation rules, necessary audio recordings, and timing tasks. Such information is more easily and readily accessible, as all of it is included in one location, a tablet screen, rather than multiple sources (e.g., protocol, administration manual, notepad, or stopwatch). Such developments are intended to allow the examiner to focus more on the examinee, rather than on the multiple and constant mundane tasks of testing, such as finding stimulus manipulatives and organizing materials. This new digital system has ensured that the automated scoring follows necessary rules, while retaining adaptability and flexibility, with the examiner maintaining control over the testing session with increased efficiency (Cayton et al., 2012). Some computerized self-reports are now available remotely, thereby possibly eliminating examiner participation and control over the testing session. The Q-interactive system, however, maintains control in the hands of the examiner, with human interaction being an integral part of the process (Scheller, 2013).

Lastly, multiple manipulative options for completing the testing session are readily available including clicking buttons, touching the screen, and using a stylus pen to take notes on the screen. Individual preferences and comfort levels are, thereby, acknowledged (Cayton et al., 2012). In a way, administration is becoming more customized for both examiner and examinee, while still following standardized administration.

According to the Q-interactive website (www.helloq.com/home), complete digital versions are available for multiple measures including, but not limited to, the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS–IV), the Wechsler Individual Achievement Test, Third Edition (WIAT-III), the Wechsler Intelligence Scale for Children, Fourth Edition (WISC–IV), the Wechsler Memory Scale, Fourth Edition (WMS-IV), and the California Verbal Learning Test, Second Edition (CVLT–II) along with selected subtests from multiple measures such as the Children's Memory Scale (CMS), the Delis-Kaplan Executive Function System (D–KEFS), and a Developmental Neuropsychological Assessment, Second Edition (NEPSY-II). Digital formats for additional measures are continuously being developed. For a complete list of all measures available see www.helloq.com.

To better understand this latest innovation in testing, it is helpful to discuss the process related to the development of digital assessment or computer-assisted testing, including common critiques and concerns often raised with respect to computerized testing, such as the validity and reliability of results, test security, and the protection of data. First, however, as discussed

earlier, the need for comparable data between paper-and-pencil and computerized tests is critical, and measures on Q-interactive are no exception regarding this issue.

Equivalency and Beta Testing

In order to ensure that the digital format of measures adapted was indeed interchangeable with traditional paper-and-pencil methods, two evaluation phases were conducted with each measure—equivalency testing, which was explained previously, and beta testing, which is the last stage of testing using the test at multiple sites for real world exposure (Cayton et al., 2012). Pearson conducted multiple studies to address the equivalency of results with all measures currently available in a digital format and detailed results are available in multiple technical reports available at www.helloq.com. Such equivalency studies were necessary in order to accurately utilize the existing norms along with maintaining the well-recognized reliability and validity of the paper-and-pencil measures with the newer digital format. In general, all the studies completed suggested a statistically acceptable level of equivalency (see Daniel, 2012a, 2012b, 2012c, 2012d, 2012e, 2012f).

To better understand such findings, a detailed description of the equivalency study with the WAIS-IV (Daniel, 2012a) is provided. The current digital version of the Wechsler scales maintained physical manipulatives, such as blocks for the Block Design subtest, and printed response booklets, such as the Processing Speed subtests. Daniel (2012a) explained that "digital administration

may affect test scores for multiple reasons including examinee interaction with the tablet, examiner interaction with the tablet, and global effects of the digital assessment environment" (p. 2). "If a task was not found to be equivalent across the two formats (i.e., digital effect), the cause of the digital effect was investigated" (p.3), and addressed accordingly. Such effects would suggest an improvement or advancement in assessment if they were due to increased accuracy in administration and/or scoring. However, if the opposite was the case, necessary modifications would need to be made with the Q-interactive system to remove the source of error. This process was completed via collection of data, as well as video recordings of the administration procedures.

The advantages and disadvantages of multiple design methods were considered with respect to equivalency studies including randomly equivalent-groups design, non-randomly equivalent group design, retest design, and dual capture design (Daniel, 2012a). The latter term referring to comparing digital and paper recording and scoring of the same video recorded administration. Two WAIS-IV studies were completed with non-clinical participants along with qualified and trained professionals who conducted the testing at four different sites. The first study, in the summer of 2011, utilized a non-randomly equivalent group design and was comprised of two sub-groups. Participants' age, gender, ethnicity, education, and region were considered. Each examinee took the entire WAIS-IV in its standard sequence, with half of the subtests administered in paper format and half in the digital format.

A multiple regression analysis was computed for each subtest based on the norm sample in order to predict each subtest score from demographics and scores on the subtests in the other half of the battery. Then, subtest scores obtained with digital administration were compared with scores predicted by the regression equation applied to scores on paper-administered subtests. Only three subtests showed statistically significant format effect sizes that exceeded the 0.2 criterion agreed upon at the beginning of the study. This 0.2 value is approximately equivalent to one-half of a scaled score point on the Wechsler subtest metric, thus suggesting minimal variance. Daniel (2012a) concludes: "the finding that 12 of the 15 subtests did not show format effect size reaching 0.2 supported the goal of this research, which was to demonstrate the equivalence of the Q-interactive administration method to the standard procedure" (p.13). Possible format effects for three subtests were thoroughly investigated and addressed. A second study was conducted later in 2011 to further explore and correct for the format effects identified. A similar process and demographic data were included. Image clarity and a minor timing issue accounting for format effect during the first study were no longer impactful and subtests did not exceed the effect size criterion.

While such studies and related results are promising, it is important to note that since studies use non-clinical samples, the potential effects when using digital interface with individuals of particular clinical conditions is not yet known (Daniel, 2012a). These studies noted the importance of examiner

practice and competence in reducing possible digital effects. Overall, the WAIS-IV equivalency studies along with those conducted for additional measures laid the foundation for future developments.

In addition to the equivalency studies described, WAIS-IV beta studies were conducted over a two-month period and included psychologists and psychometrists using the digital version at several different locations (Cayton et al., 2012). Following on-site training, the digital WAIS-IV was used at a forensic psychology practice in the Northeast, a large neuropsychology clinic in the Midwest, and an urban high school also in the Midwest. Continuous feedback was provided regarding the experience and different aspects of the digital assessment process. Such feedback indicated that the process worked well and, indeed, saved time. Some data currently available (Cayton et al., 2012) suggests that "...using Q-interactive saved 15-20% in total time required from beginning the assessment to completion of reporting the results" (p. 390). More recently, and based on a larger user base, data from a four-month public beta study revealed that clinicians using Q-interactive experienced a 30% time savings that could translate to a 35% savings in later cost when compared with using paper-and-pencil assessments (J. Ward, personal communication, October 2, 2013).

Pearson also utilized these beta studies to gain further insight into examiners' experiences with the digital format. Results indicated that, while it would likely take some time for examiners to become fully familiar and comfortable with the digital version, once that happened, they will be able to

observe the client with greater sensitivity, and to provide an overall more efficient testing session (Cayton, et. al., 2012). More specifically, such studies showed that, while learning the system and feeling ready to use it took anywhere from ½ a day to a day along with additional supervised practice, examiners might not fully master or gain the full benefits from the digital advantages until anywhere between 3-10 administrations (Cayton et al., 2012).

This length of time is fairly similar to what it generally takes for a beginning examiner to become familiar and confident with any test administration. However, for a veteran clinician, having to "start over" in such a way might be intimidating or frustrating, and some might prefer to stick with what they know and have mastered over years of practice. Considering that psychological assessment has taken an enduring step forward (Wagner, 2012), most if not all clinicians will eventually need to take the necessary steps to become more adept at using digital assessment methods. Such issues fall under ethical considerations along with Q-interactive test security discussed next.

Test Security

The Q-Interactive Data Security & Privacy Technical Whitepaper (2013) provides detailed information about test data security. Testing data is transferred from the tablet, using a secure connection, and stored in an encrypted Pearson database. All transfers to or from the web application to the tablet are automatically encrypted, using industry-best practices. Pearson security for data stored on Q-interactive complies with the standards established

in the Health Insurance Portability and Accountability Act (HIPAA) (1996) Security Regulations, and in accordance with the Health Information Technology for Economic and Clinical Health Act (HITECH Act) applicable to Business Associates (www.helloq.com). Administrative, physical, and technical safeguards protect personal data stored on Q-interactive. Access to the web-based Q-interactive Central and database is password protected. Assessment data is saved in an encrypted directory on the tablet device and is transmitted securely. Stored data is protected utilizing encryption, physical security, and administrative safeguards.

Furthermore, Pearson and its Q-interactive application comply with all applicable security and privacy regulations required by HIPAA and the HITECH Act as a Business Associate. Lastly, Q-interactive provides more comprehensive descriptions of the roles and expectations of Pearson and the user regarding its use. These documents are accessible on the Q-interactive website www.helloq.com and include a Business Associate Addendum, Privacy Policy, Subscription & License Agreement, and Terms & Conditions of Sales & Use.

Current Feedback and Reflections

While Q-interactive continues to gain greater awareness as clinicians take the plunge into digital assessment, there is still limited direct data from examiners about the system and the experience of using it. According to Ullman (2012), "clinicians using Q-interactive...have been impressed with the new technology and plan to permanently utilize Q-interactive in their assessments

clinic" (p.4) and "the decision to purchase Q-interactive was based on its advantages including improving administration accuracy and real time scoring" (p.4). Additional pros and cons of Q-interactive were provided by Wagner (2012) via an Illinois School Psychology Association School Psych Tech Report. Many pros were noted, including: lower initial cost to access a test versus purchasing a kit; the system uses familiar technology; fewer test administration materials; integrated process of administration, scoring, and interpretation; less emphasis on memorization of test rules; greater ability to attend to session observations; easier to record examinee input; saves time and increases efficiency, both of which offset purchase price; and improvement in training new administrators.

It is clear that many of the advantages listed meet the goals of Q-interactive. The cons discussed include higher front end-cost, possible networking connection difficulties, and some difficulty mastering handwriting on an electronic device. Wagner (2012) points out that "Q-interactive…is a model for a completely self-contained assessment system. Instead of a kit for each test, as well as protocols and manipulables, an entire library of assessments will be available on just two digital tablets" (p. 12). Whether one is impressed with such information or not, it does appear that Q-interactive makes selecting and creating custom assessment batteries for individual clients easier and less wasteful in terms of actual materials. According to Daniel (2013), a recent survey of 95 practitioners who have administered the WISC-IV using Q-interactive suggested that in most cases the

digital format either increased examinees' engagement and attention (especially with younger examinees) or had no effect. Additional data from this survey are available at www.helloq.com.

Thus far, it appears that the digital versions of assessment measures currently available have been positively received and Q-interactive is achieving its goals (Ullman, 2012; Cayton et. al., 2012; Wagner, 2012; Scheller, 2013). It is likely that many clinicians are unaware of or unsure about this new way of conducting assessments, and might not be as comfortable with a digital modality. It is also probable that additional issues and questions will need to be addressed as it gains more popularity and wider use. While direct feedback from the field is critical to the overall shift to digital assessment in the future, whether or not results of digital assessments are comparable to the traditional versions is a critical factor. Such studies have been conducted via Pearson Assessment demonstrating that no significant differences exist that would invalidate results and current norms, and that interpretative methods may be used in a valid and reliable manner with the digital format. Further studies by independent researchers would strengthen these findings significantly.

It is likely that at least some clinicians may not enthusiastically embrace this new digital modality for testing and what it currently offers. However, while both versions of tests are currently available and interchangeable, allowing for both to be utilized with equal effectiveness, the wave of the future in terms of psychological testing appears to be unavoidable. Cayton et al. (2012) predicted

that “Q-interactive will evolve rapidly” (p. 424). New technology and examiner feedback will dictate further progress and changes, and increased complexity and functioning of this digital system will be developed and implemented. It is also likely that additional testing publishing companies will develop similar systems to bring other psychological measures into a digital format.

Cayton et al. (2012) continue: “The true promise of Q-interactive and digital assessment in general lies in what can be provided in the future . . . as tests become digitally native, they will likely no longer rely on traditional norms, and more cutting-edge designs and scores will be developed” (p. 425). The Fifth and newest edition of the Wechsler Intelligence Scale for Children (WISC-V) marks the next step in the advancement of psychological testing. The WISC-V will be available via both traditional and digital formats, using equating where appropriate to enable both formats to rely on a common set of norms (J. Ward, personal communication, January 2, 2014).

Such developments are undoubtedly exciting, but such a shift may also be met with anxiety to some degree, since this shift represents a change or movement into relatively unknown arenas. Despite such mixed emotions and varying degrees of comfort with the new digitalized way of life, it would be important for all clinicians involved in psychological testing to become more familiar with these latest innovations in the field, follow their progress, and take at least some steps to adapt accordingly.

The Impact of Computers and Digital Assessments

It is clear that despite initially limited progress in psychological assessment, the field is quickly catching up in terms of using technological innovations. The roles of psychological assessments vary according to the changes that occur in the approaches to treatment and prevention methods for mental health disorders. Computerized assessment includes quite a variety of testing options and information available thus far suggesting multiple advantages for digital and computerized assessment, as well as some expected challenges. In this context, several additional issues are briefly explored below such as graduate training, research, and diversity, acknowledging the benefits and limitations for assessment modality on an individual basis whenever possible and appropriate. These issues are discussed in general, but, in some cases, a necessary distinction between computer-based and computer-assisted modalities is noted. It is important to note that each of these issues is worthy of a more in-depth discussion that is beyond the scope of this chapter.

Advantages and Disadvantages

Information about the use of computers and digital formats in testing continues to grow, but we still have much to learn about the benefits and limitations of such technological developments in the field of psychological testing. Casper (2004), among others, provides a detailed discussion of advantages and disadvantages of general use of computers in psychological assessment. It is important to remember that psychological tests cover a wide

range of measures and settings, and, thus, some benefits and limitations may be more relevant to certain types of assessments than to others. Overall, benefits include reduced assessment time, cost effectiveness, easier access and convenience for test-takers, access in other languages, increased neutrality and appeal, and adaptability. In particular, in the era of managed care, where cost effectiveness and increased efficiency reign supreme, Groth-Marnat (2009) recommended computerized assessment as a way to increase the financial efficacy of assessment.

Certain factors, some being more or less relevant to computer assisted assessment, may be considered as both advantages and challenges of digital assessment including cost, security, validity, reliability, and norms as well as varied comfort level and appeal. Limitations include necessary skills for examiner and examinee, cultural issues, access, engagement and responsiveness, lack of individualized context, and limited communication and rapport. The latter few issues are more reflective of computer based assessment. Another issue that needs to be considered when using technology, despite all of its advances, is technical difficulty (Burke & Normand, 1987). In addition, Butcher (2013) states that "the danger of misusing data applies to all psychological formats, but the risk seems particularly high when one considers the convenience of computerized outputs" (p.177). The possibility of human error in scoring is clearly recognized and computers are generally seen as superior in this regard. However, Butcher (2013) cited a number of sources suggesting that

computer-based reporting services, to some degree, may produce erroneous results. Weiner and Greene (2007) have suggested that only 60% of computer interpretations were clinically appropriate, and it has been clearly stated that computer-based reports should not be used to replace clinical judgment, but, rather, as adjuncts to clinical interpretations.

Collaborative Assessment

Clinical judgment often relies at least in part on observations and direct interactions between examiner and client. Behavioral observations can often help clarify the type and impact of various symptoms to better understand testing data, diagnosis, and client needs. Thus, the nature of this relationship and interaction between examiner and examinee are noted as critical to a productive and beneficial assessment process, especially in a collaborative or therapeutic assessment setting. Collaborative and therapeutic assessment, terms coined by Constance Fischer and Steven Finn, include an empirically supported procedure that involves the client as an active participant in the assessment process (Smith et al., 2011).

The impact of additional electronic devices on such interactions and observations is clearly relevant. Would the collaborative and therapeutic nature of the testing environment be improved or reduced with increased digital reliance? Remote computer-based testing options along with computerized clinical interviews eliminate such a component, and the increased convenience of such a modality comes at a cost of valuable clinical interaction. However,

developers of computer-assisted testing (i.e., Q-interactive) argue that increased use of digital modalities in testing would actually allow clinicians to increase their focus on the client (Scheller, 2013). While more research in this area is necessary, Smith et al. (2011) believe that web-based technology can actually help facilitate the use of collaborative assessment, and that even computer-based assessments appear to be useful for treatment providers.

Diversity Issues

Diversity issues relevant to digital assessment reflect national and international demographics and various individual differences. According to ITC international data, as reported by Oakland et al. (2001), "two-thirds of countries report a critical need for both group and individual tests of achievement, intelligence, vocational interests and aptitudes, social development, and personality..." (p. 24). This is especially true for developing countries. The question is whether or not such needs could be met with the help of computers and technology in assessments. The benefits of digital assessment have been described previously, and the large majority of individuals in North America have regular access to current technology. Foxcroft and Davies (2006), however, state that "guidelines on Computer-Based and Internet-Delivery Testing highlight the need to consider equality of access for all groups and the impact that inequality of access to computers and technology can have on test performance" (p. 177). With respect to computer-assisted assessment and Q-interactive, it is up to clinicians to provide the necessary technology.

Another important issue that relates to digital test formats is the potential effect on individuals with certain physical or learning disabilities. Will technological advances make testing more or less accessible for such individuals? A question rarely discussed in research is with respect to examiners who have a physical or learning disability. Would this new digital format allow for greater opportunities and accessibility for examiners with disabilities, or would the hindrances they currently face continue? Some studies have begun to address the first question with mixed perspectives. Burke and Normand (1987) suggest that computerized testing would be a potential benefit to individuals with visual, auditory, and physical limitations. Crook, Kay, and Larrabee (2009) suggest that most computerized measures have significant visual-perceptual demands, which can cause difficulty for patients with reduced visual acuity. The need to use a keypad, mouse, or similar instruments requiring manual manipulation in computer testing may also prove to be particularly demanding for certain clients with fine motor difficulties or for the elderly, especially when the tests are not measuring actual abilities in this area.

Computer administration can also be helpful for Non-English speakers by providing language options and deaf clients can read items on the screen. For those who cannot read or have visual impairment, questions can be read to them, and a multimedia administration could also be helpful with children (Garb, 2007). In addition, computerized testing capacities such as voice recognition and three-dimensional graphics might allow more opportunities in

terms of testing for those who have certain mental or physical challenges (McIntire & Miller 2007). Such factors seem relevant for both computer-based and assisted modalities with additional research in this area likely to come in the future with respect to Q-interactive.

Clearly, more research regarding such issues, along with the potential impact of specific mental health diagnoses, is necessary. Hofer and Green (1985) noted that "unfamiliarity with computers is probably most correlated with ethnicity, gender, age, and socioeconomic status" (p. 828). Mindful of this, Hays and McCallum (2005) suggest that "the decision to use one format or the other should be based on the specific needs and abilities of the examinee, the credibility of the equipment, and the unique characteristics of the setting, rather than the expectation that one format is 'inherently superior' to the other" (p. 612).

Attitudes toward Computerized Testing

Despite the possibility of some computer anxiety arising as acknowledged previously, several studies have showed good general acceptance for computer administration by clients (Casper 2004). Research specifically indicates that perception and the overall attitude of test-takers towards computers in testing has been mostly positive (Foxcroft & Davies, 2006), and client acceptance of computerized assessment appears high (Olson, 2001). In addition, while some claim that using computers may remove a personal component that is necessary to an effective interaction with a client, research suggests that clients tend to

react favorably to computerized testing or interviewing sessions (Burke & Normand, 1987).

In fact, it appears that some individuals provide more accurate information about themselves to a computer, especially regarding sensitive issues, than they would via a face-to-face interview or self-report, possibly feeling less embarrassed giving information to a computer than to a person (Burke & Normand, 1987). Forbey, Ben-Porath, and Gartland (2009) noted that test-takers' candidness towards computer-administered tests may increase because computers foster a sense of anonymity. Furthermore, while computer interviews were often ranked as less friendly, they were described as shorter and more relaxing. In addition, assessing the ongoing progress of interventions is highly valued, and computerized assessments represent an improvement over paper-and-pencil measures since they can provide both clinician and client with instant feedback (Smith et al., 2011). Smith et al. (2011) found that both practitioners and clients participating in a collaborative model of assessment experienced computerized assessments to be useful and user friendly.

Interestingly, according to Burke and Normand (1987) "although client acceptance of computer-based testing appears to be somewhat favorable, the acceptability by professionals has been less clear" (p.48). Despite the clear advantages of computerized psychological assessment, many psychologists still do not utilize such resources (Olson, 2001). Greene (2011) acknowledges that

"...there is an amazing reluctance to even consider electronic technology in the field of psychological assessment" (p. 200).

Olson (2001) discussed the modest rate of use of computerized resources in psychological assessment among psychologists and specified the limited increase in such use despite continued increase in availability. He described a study that found that in 1987, out of 227 surveyed APA members in professional practice, 26% used computers for administration, 41% for scoring, and 29% for interpretation, and that a decade later, the data was similar. Greene (2011) included data suggesting that only one-third of MMPI-2 tests were computer scored in 2007, and less than one-quarter were computer interpreted. Overall, the prevalence of computer-based assessments, whether administration, scoring, or interpretation, seemed to vary anywhere from approximately 10% to 70% of clinicians utilizing some form of electronic testing methods (Butcher, 2013). It is too early to tell the acceptance of Q-interactive. "Attitudes towards new technological developments are always characterized by radicalism stemming from various sources... therapists as well as trainers may be afraid that they might become superfluous" (Caspar, 2004, p.222).

While it is understandable that some of us may have this concern, as long as we do not become fully reliant on technological advances, computers will not be able to replace us. Instead, computers will significantly enhance our assessment capabilities. It is important to reiterate that computer-assisted assessment maintains the active role of the examiner. Even so, Hofer and Green

(1985) indicate that "psychologists must demonstrate their unique testing skill and judgment in reviewing computerized interpretations..." (p. 833). A lack of adequate training and experience with technology was noted as another significant factor for continued resistance to using computerized testing more consistently. This area is discussed below.

Graduate Level Teaching & Training

Coyne and Bartram (2006a) stated that the advances in technology related to psychological assessment require a re-examination of the training of those who use tests and that education and training in this area is lagging and there is an urgent need to catch-up. Olson (2001) gathered data from 251 APA accredited doctoral programs in psychology, with results showing that 25% of the programs did not utilize any computerized assessment. Of the 75% of programs that did use such technology, the median was only three computerized tests.

Olson (2001) noted that the limited exposure and training provided in graduate schools needs to be addressed given that they contribute to a sense of "technophobia" in some practitioners in the field, inhibiting them from utilizing computerized measures in their post-graduate experience. Coyne and Bartram (2006b) raised several important questions with respect to university education and training, and how specific computer-related issues should be addressed. Gabel (2013) addressed several practical considerations for trainers and instructors regarding incorporating new testing developments, specifically Q-

interactive, with students and clinicians in training. Gabel acknowledged that such information will take up additional time in class, and also indicated the importance of instructors displaying the necessary qualifications to teach such technology.

There is a clear need for universities and agencies to develop new policies for students using digital measures. Despite the added challenges to already demanding assessment courses, graduate students must have awareness of the future of psychological testing and be prepared, at least to a certain degree, to participate effectively in the field as it moves forward (Gabel, 2013). Developing measures to meet new needs or new conceptions of human characteristics is an ongoing goal in the field of assessment, and learning how to utilize such new measurements is a necessity (Coyne & Bartram, 2006a). Assessment courses would need to be revised accordingly, and early clinical training will also need to take such issues into consideration. While additional guidelines are needed in this area, information that is currently available helps establish the first steps in planning to incorporate the ongoing technological advances in psychological testing, thereby preparing graduate students and clinicians in training for what is to come in the field of psychological assessment. Both computer-based and computer-assisted modalities would need to be acknowledged accordingly.

A final issue to ponder is how assessment instructors and supervisors ensure that students learn and become fully familiar with testing procedures, including accurate administration and scoring, before they increase utilization of

computerized assessments? Such understanding is critical, as it has become far too easy for clinicians to rely on computer scoring and reports, thus potentially reducing their actual understanding of the psychometric foundations of measures they use as a basis for interpretations.

Treatment and Diagnosis

Butcher et al. (2004) point out that "there is evidence that computer-based test reports can be quite valuable for the purposes of treatment planning" (p. 334). Research has shown that by providing clinicians with increased information about the client's personality and symptoms, psychological assessment of clients in pretreatment planning can improve the effectiveness of psychotherapy interventions. Computers can further enhance the role of assessment in treatment by producing more comprehensive data quicker than ever before (Butcher et al., 2004).

Additional feedback via computerized assessment early in the treatment process can further contribute to positive treatment outcomes. Garb (2007) discussed some of the most widely used computer interviews and noted that almost any self-report measure available can be administered by a computer with acceptable equivalency results. Such measures provide relevant information at pretreatment regarding diagnostic considerations, as well as with respect to the progress of therapy. Structured interviews often provide clinicians with valuable clinical data for treatment and diagnosis, but can also be time-consuming, which can be directly and effectively addressed with the

increased efficiency of computers. There is a certain level of assurance when using computers that all items will be answered resulting in increased comprehensiveness, and computer reports are more legible and can be simply and quickly added to a clinical record (Bachman, 2003; Garb, 2007).

While there are clear advantages in using computers for the purpose of diagnosis and treatment planning, the impact of computerized assessment is not entirely positive. While some studies suggested that clients reveal more sensitive information to a computer during surveys than to a human investigator, others found inconsistent data (Garb, 2007). In addition, the inability to observe the body language of the client has been noted as a significant disadvantage. Behavioral observations, as noted previously, are often crucial in terms of the severity and intensity of one's symptoms, as well as important characteristics relevant to the therapeutic relationship. Garb (2007) further clarified that there is a possibility of false positives regarding both symptoms and diagnoses when using computerized interview assessment modalities alone. Despite this, Garb still maintains that the adoption of computer administration of interviews and rating scales "will likely lead to improved treatment for clients and an increase in accomplishment for clinicians" (p. 10).

When to Use Which Technology and With Whom

Whether referring to computer-based or computer-assisted assessment, Casper (2004) believes that "overall, the balance of advantages and disadvantages in the use of computers in clinical diagnostics certainly depends

on the individual case" (p. 230). Schulenberg and Yutrzenka (2004) reference multiple researchers to make several important recommendations when using computer-based testing and digital assessment in general. They encourage clinicians to give responders the opportunity to decline computer-assisted or computer-based testing, as well as to ensure confidentiality and security of testing data. These authors further feel that clinicians need to become critical consumers of computer-based interpretive programs. Finally, knowing the psychometrics of the computerized measure and monitoring user-qualifications are both essential.

In addition to the typical factors that need to be considered in any psychological evaluation (e.g., referral questions, test selection, time allotted, and cost), clinicians now also need to be fully familiar with issues stemming from new available technology. Snyder (2000) notes that the use of computerized interpretive reports is risky and clinicians may be liable if they include a computer-generated interpretation in a report without adequately evaluating its relevance. Essentially, it falls to the clinician to ensure that the computerized interpretation is an accurate reflection of the respondent being assessed, and not just a series of interpretations put forth in a software package by a testing company (Butcher, 2003). It is very possible that such responsibilities deter some clinicians from utilizing computerized assessments, even though clients seem to be more open to them than to the more traditional forms.

Adaptive Testing

One particularly “hot topic” that is relevant when discussing computerized assessments and utilization of current technology is adaptive testing. Weiss (1985) explains that “an adaptive test is one in which different sets of test questions (items) are administered to different individuals depending on each individual’s status on the trait being measured” (p. 774). This stands in contrast with conventional tests, usually paper and pencil, where all examinees are administered the same fixed set of items. Adaptive testing has also been referred to as tailored, response-contingent, individualized, branched, or sequential testing (Weiss, 1985). Weiss (1985) discusses the several advantages of adaptive testing that include increased efficiency, since fewer items are administered, and potentially higher validity and reliability due to greater precision in measuring certain traits. Adaptive testing may be especially favorable for individuals who have special needs or are "extreme" in one way or another (Casper 2004). Additional benefits include shorter test length (van der Linder, 2008) possibly leading to reduction in boredom or discouragement for the test-taker (Alkhadher, Clarke, & Anderson, 1998).

Computers allow for further advancement in adaptive testing by updating the score of the test taker each time a response is provided and automatically adapting the choice of the next item (van der Linden, 2008a). Therefore, computerized adaptive testing (CAT) shares the benefits of computer testing in general along with those associated with adaptive testing (Rudick, Yam, & Simms,

2013). CAT provides individually tailored tests with immediate scoring and feedback and the possibility of using innovative item types (van der Linden, 2008b). Rudick et al. (2013) discuss the process and advantages of CAT in personality testing and the many possibilities for the future.

Greene (2011) argued that it is not necessary to administer a fixed set of all items when using self-reports (such as the MMPI-2) with the computer technology that is currently available. Instead, Greene advocates for individualized administration utilizing available adaptive computer-testing systems. Overall, CAT has gained empirical support and there is great potential to further develop computerized adaptive testing with personality and other psychological measures (Forbey & Ben-Porath, 2007; Hol, Vorst & Mellenbergh, 2008; van der Linder, 2008b).

Impact on Research

Computer-based clinical assessments are already widely accepted and used in the fields of personality and neuropsychological functioning. There is ongoing growth in the areas of cognitive and intellectual testing as well. The progression of the psychotherapeutic process is another area where assessment plays an important role (Casper, 2004). Along with assessing for a variety of mental health and general functioning issues, psychological tests are also utilized to verify the effectiveness of interventions and new or ongoing concepts in the field (McIntire & Miller 2007). Therefore, the established efficiency of computers in assessment would also likely be a benefit in terms of data

collection, especially in terms of processing large quantities of information quickly and accurately. Burke and Normand (1987) noted the vast opportunities for research utilizing the unique capabilities of computers when developing and researching new measures of psychological constructs.

Future Directions in Psychological Testing

Butcher (2013) comments:

> Computers have been an integral part of psychological assessment since the 1950s although initially only in scoring and processing of research information. Over the past 60 years, their applications in mental health care and forensic settings have broadened, and computers have become important and necessary components to assessment (p.165).

While multiple authors noted some limitations and concerns with respect to computerized testing, they also acknowledged the many benefits and exciting possibilities for the future. For example, Crook et al. (2009) state that "current technologies can be employed to provide highly realistic simulations of the cognitive tasks that must be performed in everyday life, on which developmental change or the effects of neurological disease or trauma are first noted" (p. 96). These authors further suggest that tests could potentially allow test-takers to play a more interactive role in testing where they would perform a variety of tasks while their performance was measured.

McIntire and Miller (2007) noted that computers can replicate real-life situations utilizing capacities such as reacting to voice-activated responses and displaying three-dimensional graphics. Butcher (2013) emphasized that psychologists have not fully utilized "...the extensive powers of the computer in

presenting stimuli to test-takers...with the power, graphic capability, and flexibility of the computer, it is possible to develop more sophisticated, real world stimulus environments than are currently available in computer-administered methods" (p.175). Cayton et al. (2012) suggested that digital assessment options, such as the Q-interactive system, would allow for new constructs for understanding cognition, emotion, and behavior to be developed, and that clinicians will have new ways of understanding clients. Alkhadher et al. (1998) provided a reminder that one of the most important issues in human measurement is to minimize the effects of unwanted variables that may influence one's performance. Thus, the goal is to ensure that test scores reflect what they are intended to measure. It appears that computer-based testing and advances in digital assessment strive to do just that. Lastly, Casper (2004) indicates that "there is no denying the important role that technology will play in the future of clinical psychology, of which psychotherapy services, training, and research are an important part" (p. 237).

Conclusion

Despite some discouragement from the world of managed care, psychological assessment continues to be a valuable and often necessary factor in a multitude of settings (e.g., mental health, education, forensics, medical arenas, and the workplace), and with a range of ages and populations (Naglieri & Graham, 2003). Perhaps Oakland et al. (2001) described it best, stating that "test use is universal [and] tests are used in virtually every country with

newborns through the elderly" (p. 24). The widespread use of assessment continues with the capacity to measure one's functioning in a wide array of areas, including intelligence, cognitive abilities, academic achievement, interests and vocation, neuropsychology, personality characteristics, social-emotional factors, behavior, and psychopathology. In these contexts, assessment utilizes various types of measures including, but not limited to, interviews, behavioral assessment, self-reports, and performance-based tests. Finally, psychological testing is relevant for diagnostic purposes, placement, treatment plans, and treatment evaluation. Just as the goals of assessment have evolved over the years to some degree, the methods of assessment have as well.

Boake (2002) states that it is likely that further emerging technologies using computerized assessment will continue to offer vital advantages. "Eventually, new tests based on these technologies will replace the individual intelligence test as we know it. Then it will be the job of these new tests to carry on the tradition of mental testing established by the Binet-Simon and Wechsler-Bellevue scales" (p. 403). "There has been an enormous increase in the use of technical devices including computers, for potential consumers in the field of clinical psychology and psychotherapy" (Caspar, 2004, p. 222).

As a practicing clinician, the thought of more precise and possibly briefer assessments that may produce a similar level of results and insight into a client's functioning and needs is, indeed, appealing. At the same time, any significant change to how we conduct psychological assessment is frankly a bit intimidating.

The inevitable advancements to come in assessment are needed, and will require additional research to determine how such technological innovations can provide possible solutions to the various problems and concerns noted throughout this chapter and by various researchers in the field. As psychologists, we are not necessarily flexible with change, somewhat ironic considering our expectations for our clients. Thus, maybe with ongoing support and communication, we can not only accept, adjust, and move forward, but embrace and thrive in this increasingly specialized field of ours. Consider Geisinger's (2000) words:

> Psychological testing continues to evolve, with both the nature of the assessments that we use and the criteria against which we evaluate tests changing, at times rapidly. The psychological profession will need to adapt as well by developing measures to meet new needs or new conceptions of human characteristics, continually learning to use new measurements, and changing existing measures. Such changes necessitate continued excellence in professional training in graduate school and beyond (p.118).

Since the goals of psychological testing have changed since the first tests were developed, continued progress is necessary. Naglieri and Graham (2003) speculate: "Only time will tell if the next 100 years of assessment psychology will be more of the same or if innovative approaches will develop and be embraced" (p. 657). Despite some uncertainty and the continued need for research, it is clear that the role of computers and digital assessment will continue to advance.

The future of psychological testing is bright, and is now led by digital assessments and the expanded use of computers that will allow for adaptive

testing opportunities, potentially decreased cost, and increased efficiency. This could lead to more individualized and real life-based evaluations than ever before. Such future developments, however, should not lose the direct interactive exchange between the clinician and examinee, with the necessary clinical observations, insight, and judgment taking into consideration the context of each unique individual in terms of the conclusions made and the recommendations offered.

The importance of establishing and maintaining rapport with test-takers before, during, and after testing cannot be minimized (Butcher, 2013). It is also important to remember that computers do not make decisions, but, rather, follow complex and detailed instructions (Butcher, 2013). Furthermore, recognizing both the benefits and limitations of computer-based testing, Crook et al. (2009) did not think that digital assessments were likely to become an appropriate or complete substitute for traditional clinical diagnostic evaluations.

It is important to remember that it is the clinician who is still responsible for producing a comprehensive, individualized, and well-integrated report regardless of modalities and measures used. Digital format will certainly help with the ease and accuracy of administration and scoring, just as computerized reports will greatly assist in the analysis of scores and interpretation. Thus, technological options are invaluable resources to any clinician. However, the art of report writing, the end product of psychological assessment, remains the same, and, at least for now, in the hands of the

clinician. In fact, the roles and responsibilities of psychologists conducting testing are as critical, if not more so, in this new era of digital assessment than they were before.

References

Alkhadher, O., Clarke, D., & Anderson, N. (1998). Equivalence and predictive validity of paper-and-pencil and computerized adaptive formats of the Differential Aptitude Tests. *Journal of Occupational and Organizational Psychology, 71*, 205-217.

American Psychological Association. (2010). *American Psychological Association ethical principles of psychologists and code of conduct*. Available from http://www.apa.org/ethics/code2002.html

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24*(3), 383-405.

Bachman, J. (2003). The patient-computer interview: A neglected tool that can aid the clinician. *Mayo Clinic Proceedings, 78*, 67-78.

Bartram, D. (2006). The internationalization of testing and new models of test delivery on the internet. *International Journal of Testing, 6*(2), 121-131.

Burke, M. & Normand, J. (1987). Computerized psychological testing: Overview and critique. *Professional Psychology: Research and Practice*, *18*(4), 42-51.

Butcher, J. (1987). *Computerized psychological assessment: A practitioner's guide*. New York, NY: Basic Books.

Butcher, J. (2013). Computerized personality assessment. In J. Graham, J. Naglieri, & I. Weiner, (Eds.), *Handbook of psychology, volume 10: Assessment psychology* (pp. 165-191). Hoboken, NJ: John Wiley.

Butcher, J., Keller, L., & Bacon, S. (1985). Current developments and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology*, *53*(6), 803-815. doi:10.1037/0022-006X.53.6.803

Butcher, J., Perry, J., & Atlis, M. (2000). Validity and utility of computer-based test interpretation. *Psychological Assessment, 12*(1), 6-18.

Butcher, J., Perry, J., & Hahn, J. (2004). Computers in clinical assessment: Historical developments, present status, and future challenges. *Journal of Clinical Psychology, 60*, 331-345.

Casper, F. (2004). Technological developments and applications in clinical psychology and psychotherapy: Introduction. *Journal of Clinical Psychology, 60*(3), 221-238.

Cayton, T., Wahlstrom, D., & Daniel, M. (2012). The initial digital adaptation of the WAIS-IV. In E. Lichtenberger & A. Kaufman (Eds.), *Essentials of WAIS-IV assessment* (pp. 389-427). Hoboken, NJ: John Wiley.

Coyne, I. & Bartram, D. (2006a). Introduction to the special issue on the ITC guidelines on computer-based and internet-delivered testing. *International Journal of Testing, 6*(2), 115-119.

Coyne, I. & Bartram, D. (2006b). Design and development of the ITC guidelines on computer-based and internet-delivered testing. *International Journal of Testing, 6*(2), 133-142.

Crook, T., Kay, G., & Larrabee, G. (2009). Computer-based cognitive testing. In I. Grant & K. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric and neuromedical disorders* (pp. 84–100). New York, NY: Oxford University Press.

Daniel, M. (2012a). *Equivalence of Q-interactive - Administered cognitive tasks: WAIS- IV (Q-interactive technical report 1).* NCS Pearson. Retrieved from http://www.helloq.com

Daniel, M. (2012b). *Equivalence of Q-interactive - Administered cognitive tasks: WISC-IV (Q-interactive technical report 2).* NCS Pearson. Retrieved from http://www.helloq.com

Daniel, M. (2012c). *Equivalence of Q-interactive - Administered cognitive tasks: CVLT-II and selected D-KEFS subtests (Q-interactive technical report 3).* NCS Pearson. Retrieved from http://www.helloq.com

Daniel, M. (2012d). *Equivalence of Q-interactive - Paper administered cognitive tasks: NEPSY-II and CMS subtests (Q-interactive technical report 4).* NCS Pearson. Retrieved from http://www.helloq.com

Daniel, M. (2012e). *Equivalence of Q-interactive and paper scoring of academic tasks: WIAT-III (Q-interactive technical report 5).* NCS Pearson. Retrieved from http://www.helloq.com

Daniel, M. (2012f). *Equivalence of Q-interactive and paper administration of WMS-IV cognitive tasks (Q-interactive technical report 6).* NCS Pearson. Retrieved from http://www.helloq.com

Daniel, M. (2013). *User survey on Q-interactive examinee behavior.* NCS Pearson. Retrieved from http://www.helloq.com

Dimock, P. (1991). The effects of format differences and computer experience. *Measurement & Evaluation in Counseling & Development, 24*(3), 119.

Exner, J. (2005). Comment on "science and soul." *Rorschachiana, 27*(1), 3.

Forbey, J. & Ben-Porath, Y. (2007). Computerized adaptive personality testing: A review and illustration with the MMPI-2 computerized adaptive version. *Psychological Assessment, 19*(1), 14-24.

Forbey, J., Ben-Porath, Y., & Gartland, D. (2009). Validation of the MMPI-2 computerized adaptive version (MMPI-2-CA) in a correctional intake facility. *Psychological Services, 6*(4), 279-292.

Foxcroft, C. & Davies, C. (2006). Taking ownership of the ITC's guidelines on computer-based and internet-delivered testing. *International Journal of Testing, 6*(2), 173-180.

Gabel, A. (2013). Teaching assessment: Evolved - Session 3. Retrieved from http://www.pearsonassessment.com

Garb, H. (2000a). Introduction to the special edition on the use of computers for making judgments and decisions. *Psychological Assessment, 12*(1), 3-5.

Garb, H. (2000b). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that! *Psychological Assessment, 12*(1), 31-39.

Garb, H. (2007). Computer-administered interviews and rating scales. *Psychological Assessment, 19*(1), 4-13.

Geisinger, K. (2000). Psychological testing at the end of the millennium: A brief historical review. *Professional Psychology: Research and Practice, 31*(2), 117-118.

Greene, R. (2011). Some considerations for enhancing psychological assessment. *Journal of Personality Assessment, 93*(3), 198-203.

Groth-Marnat, G. (2009). *Handbook of psychological assessment.* New York, NY: John Wiley.

Hays, S. & McCallum, R. (2005). A comparison of the pencil-and-paper and computer-administered Minnesota Multiphasic Personality Inventory - Adolescent. *Psychology in the Schools, 42*(6), 605-613.

Health Insurance Portability and Accountability Act (1996). Security Regulations. Retrieved from www.helloq.com

Hofer, P. & Green, B. (1985). The challenge of competence and creativity in computerized psychological testing. *Journal of Consulting and Clinical Psychology*, *53*(6), 826-838. doi:10.1037/0022-006X.53.6.826

Hol, A., Vorst, H., & Mellenbergh, G. (2008). Computerized adaptive testing of personality traits. *Journal of Psychology, 216*(1), 12-21.

Hunt, W. (1946). The future of diagnostic testing in clinical psychology. *Journal of Clinical Psychology, 2*(4), 311-317.

International Test Commission (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing, 6*(2), 143-171.

McIntire, S. & Miller, L. (2007). *Foundations of psychological testing: A practical approach*. Thousand Oaks, CA: Sage Publications.

Multi-Health Systems Online Assessment Center (2014). Retrieved from www.mhs.com

Naglieri, J. & Graham, J. (2003). *Current status and future directions of assessment psychology* (pp. 645- 658). New York, NY: John Wiley.

Oakland, T., Poortinga, Y., Schlegel, J., & Hambleton, R. (2001). International test commission: Its history, current status, and future directions. *International Journal of Testing, 1*(1), 3-32.

Olson, K. (2001). Computerized psychological test usage in APA-accredited training programs. *Journal of Clinical Psychology*, *57*(6), 727-736. doi:10.1002/jclp.1045

Psychological Assessment Resources (2014). PARiConnect. Retrieved from www4.parinc.com

Q-interactive Data Security & Privacy Technical Whitepaper (2013). Retrieved from http://www.helloq.com/research.html

Q-local and Q-global (2014). Retrieved from www.pearsonclinical.com

Rudick, M., Yam, W., & Simms, L. (2013). Comparing countdown- and ITR-based approaches to computerized adaptive personality testing. *Psychological Assessment*, 25(3), 769-779.

Scheller, A. (2013). Q-interactive: Overview. Retrieved from http://www.pearsonassessment.com

Schulenberg, S. & Yutrzenka, B. (2004). Ethical issues in the use of computerized asessment. *Computers in Human Behavior*, *20*(4), 477-490. doi:10.1016/j.chb.2003.10.006

Simola, S. & Holden, R. (1992). Equivalence of computerized and standard administration of the Piers-Harris children's self concept scale. *Journal of Personality Assessment, 58*(2), 287-294.

Smith, R., Fagan, C., Wilson, N., Chen, J., Corona, M., Nguyen, H., Racz, S., & Shoda, Y. (2011). Internet based approaches to collaborative therapeutic assessment: New opportunities for professional psychologists. *Professional Psychology: Research and Practice, 42*(6), 494-504.

Smoline, D., (2012). Testing with the computer: State of the art, needs, and perspective. In A. Columbus (Ed.), *Advances in psychology research, volume 87*) (pp. 205-226). Hauppauge, NY: Nova Science Publishers.

Snyder, D., (2000). Computer-assisted judgment: Defining strengths and liabilities. *Psychological Assessment, 12*(1), 52-60.

Spinhoven, P., Labbe, M., & Rombouts, R. (1993). Feasibility of computerized psychological testing with psychiatric outpatients. *Journal of Clinical Psychology, 49*(3), 440-447.

Ullman, D. (2012). The future of psychological assessment. *Nebraska Psychological Association Newsletter.*

van der Linden, W. (2008a). Adaptive models of psychological testing. *Journal of Psychology*, *216*(1), 1-2.

van der Linden, W. (2008b). Some new developments in adaptive testing technology. *Journal of Psychology*, *216*(1), 3-11.

Wagner, K. (2012). Q-interactive and digital assessment. Illinois School Psychology Association.

Weiner, I. & Greene, R. (2007). *Handbook of personality assessment*. New York, NY: John Wiley.

Weiss, D. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, *53*(6), 774-789. doi:10.1037/0022-006X.53.6.774

Acknowledgements

The author would like to acknowledge Janet Ward, Mark Daniel, Jerry Michaels, Megan Carlos, and Ron Valle for their feedback and support of this chapter. The author is not affiliated with Pearson nor has she directly benefited in any way by writing this chapter.