Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology†

Russell M. Bauer[1,7], **Grant L. Iverson**[2,8], **Alison N. Cernich**[3,7], **Laurence M. Binder**[4,8], **Ronald M. Ruff**[5,8], and **Richard I. Naugle**[6,7]

[1]University of Florida, Gainesville, FL

[2]University of British Columbia, Vancouver, BC, Canada

[3]VA Maryland Health Care System, Baltimore, MD

[4]Oregon Health Sciences University, Portland, OR

[5]San Francisco Clinical Neurosciences, San Francisco, CA

[6]Cleveland Clinic Foundation, Cleveland, OH

[7]American Academy of Clinical Neuropsychology

[8]National Academy of Neuropsychology

Abstract

This joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology sets forth our position on appropriate standards and conventions for computerized neuropsychological assessment devices (CNADs). In this paper, we first define CNADs and distinguish them from examiner-administered neuropsychological instruments. We then set forth position statements on eight key issues relevant to the development and use of CNADs in the healthcare setting. These statements address (a) device marketing and performance claims made by developers of CNADs; (b) issues involved in appropriate end-users for administration and interpretation of CNADs; (c) technical (hardware/software/firmware) issues; (d) privacy, data security, identity verification, and testing environment; (e) psychometric development issues, especially reliability and validity; (f) cultural, experiential, and disability factors affecting examinee interaction with CNADs; (g) use of computerized testing and reporting

Address correspondence to: Russell M. Bauer, Ph.D., ABPP, Department of Clinical & Health Psychology, PO Box 100165 HSC, Gainesville, FL 32610-0165, USA. rbauer@phhp.ufl.edu.

services; and (h) the need for checks on response validity and effort in the CNAD environment. This paper is intended to provide guidance for test developers and users of CNADs that will promote accurate and appropriate use of computerized tests in a way that maximizes clinical utility and minimizes risks of misuse. The positions taken in this paper are put forth with an eye toward balancing the need to make validated CNADs accessible to otherwise underserved patients with the need to ensure that such tests are developed and utilized competently, appropriately, and with due concern for patient welfare and quality of care.

Keywords

Computerized testing; neurocognition; neuropsychological test validity

INTRODUCTION

The use of computerized neuropsychological assessment devices (CNADs) is receiving increasing attention in clinical practice, research, and clinical trials. There are several potential advantages of computerized testing including: (a) the capacity to test a large number of individuals quickly; (b) ready availability of assessment services without advance notice; (c) the ability to measure performance on time-sensitive tasks, such as reaction time, more precisely; (d) potentially reduced assessment times through the use of adaptive testing protocols; (e) reduced costs relating to test administration and scoring; (f) ease of administering measures in different languages; (g) automated data exporting for research purposes; (h) increased accessibility to patients in areas or settings in which professional neuropsychological services are scarce; and (i) the ability to integrate and automate interpretive algorithms such as decision rules for determining impairment or statistically reliable change.

CNADs range from stand-alone computer-administered versions of established examiner-administered tests (e.g., Wisconsin Card Sorting Test) to fully web-integrated testing stations designed for general (e.g., cognitive screening) or specific (e.g., concussion evaluation and management) applications. This has been a highly active research and development area, and new tests and findings are being released continuously (Crook, Kay, & Larrabee, 2009). Researchers have used computerized neuropsychological testing with numerous clinical groups across the lifespan. Examples include children with attention-deficit hyperactivity disorder (Bolfer et al., 2010; Chamberlain et al., 2011; Gualtieri & Johnson, 2006; Polderman, van Dongen, & Boomsma, 2011) or depression (Brooks, Iverson, Sherman, & Roberge, 2010); adults with psychiatric illnesses, such as depression or bipolar disorder (Iverson, Brooks, Langenecker, & Young, 2011; Sweeney, Kmiec, & Kupfer, 2000); and adolescents and young adults who sustain sport-related concussions (Bleiberg, Garmoe, Halpern, Reeves, & Nadler, 1997; Bleiberg et al., 2004; Broglio, Ferrara, Macciocchi, Baumbartner, & Elliott, 2007; Cernich, Reeves, Sun, & Bleiberg, 2007; Collie, Makdissi, Maruff, Bennell, & McCrory, 2006; Collins, Lovell, Iverson, Ide, & Maroon, 2006; Gualtieri & Johnson, 2008; Iverson, Brooks, Collins, & Lovell, 2006; Iverson, Brooks, Lovell, & Collins, 2006; Peterson, Stull, Collins, & Wang, 2009; Van Kampen, Lovell, Pardini, Collins, & Fu, 2006). CNADs have also been applied to adult epilepsy (Moore, McAuley, Long, & Bornstein, 2002), cardiovascular surgery (Raymond, Hinton-Bayre, Radel, Ray, & Marsh, 2006), neurocognitive problems encountered by active duty military service members and veterans (Anger et al., 1999; Marx et al., 2009; McLay, Spira, & Reeves, 2010; Retzlaff, Callister, & King, 1999; Vasterling et al., 2006), and mild cognitive impairment in older adults (Doniger et al., 2006; Dwolatzky et al., 2004; Gualtieri & Johnson, 2005; Tornatore, Hill, Laboff, & McGann, 2005; Wild, Howieson, Webbe, Seelye, & Kaye, 2008) or dementia (Doniger et al., 2005; Dorion et al., 2002; Wouters, de

Koning et al., 2009; Wouters, Zwinderman, van Gool, Schmand, & Lindeboom, 2009). Computerized tests, sometimes administered as part of a predominantly examiner-administered battery, are also used to identify poor effort within the context of a comprehensive neuropsychological evaluation (Green, Rohling, Lees-Haley, & Allen, 2001; Slick et al., 2003). The potential application of CNADs to other medical and neuropsychiatric conditions seems limited only by available knowledge and recognition of neurocognitive symptoms seen in these disorders. For this reason, clinical application of CNADs is expected to increase in the coming years.

Computerized neuropsychological assessment is currently being used in many mainstream applications to which examiner-administered neuropsychological assessment has been historically applied. This paper describes the position of the American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) with regards to key issues in the development, dissemination, and implementation of computerized neuropsychological tests in clinical practice.

NATURE AND DEFINITION OF COMPUTERIZED NEUROPSYCHOLOGICAL ASSESSMENT DEVICES

We define a "computerized neuropsychological testing device" as any instrument that utilizes a computer, digital tablet, handheld device, or other digital interface instead of a human examiner to administer, score, or interpret tests of brain function and related factors relevant to questions of neurologic health and illness. Although it is tempting to consider CNADs as directly comparable to examiner-administered tests, there are significant differences between the two approaches. First, it is important to recognize that even when a traditional examiner-administered test is programmed for computer administration, it *becomes a new and different test*. One obvious difference is in the patient interface. In examiner-centered approaches, the patient interacts with a person who presents stimuli, records verbal, motor, or written responses, and makes note of key behavioral observations. For a CNAD, examinees interact with a computer or tablet testing station through one or more alternative input devices (e.g., keyboard, voice, mouse, or touch screen), in some cases without supervision or observation by a test administrator. Also, some CNADs utilize an "adaptive" assessment approach derived from Item Response Theory (Reise & Waller, 2009; Thomas, 2010) wherein the program adjusts task difficulty or stimulus presentation as a function of task success or failure on the part of the examinee. This does not typically occur in examiner-centered approaches.

Second, whereas most examiner-administered tests require documentation of test user qualifications on purchase, some CNADs are advertised and marketed to end-users with no expertise in neuropsychological assessment or knowledge of psychometric principles. Many contain proprietary algorithms for calculating summary scores or indices from performance data, and some provide the end-user with boilerplate report language derived from the examinee's performance that is intended as an automated form of interpretation based solely on test metrics. Third, the responsible interpretation and reporting of results of CNADs requires an understanding of test utility and accuracy when installed and used in the local clinical setting, which in turn requires familiarity with many technical details regarding their psychometric properties and normative standards. How the installed program interacts with the user's unique software and hardware configuration may affect important parameters including timing accuracy, screen resolution or refresh rate, or the sensitivity of input devices. These and other issues discussed in this paper lead to the conclusion that CNADs are qualitatively and technically different from examiner-administered instruments and require best practices for their competent and safe use.

KEY ISSUES AND POSITION STATEMENTS ON CNAD

(1) Device Marketing and Performance Claims

Position Statement—It is our position that CNADs[1] are subject to, and should meet, the same standards for the development and use of educational, psychological, and neuropsychological tests (American Psychological Association, 1999) as are applied to examiner-administered tests. This position echoes similar statements made over 20 years ago by Kramer (1987) and Matarazzo (1985, 1986). In addition, CNADs likely qualify, and thus will eventually be regulated as, "medical devices" according to prevailing definitions in Federal law. As regulated devices, developers of computerized neuropsychological assessment tools will likely need to provide additional documentation that meets specific labeling standards particular to medical device regulation.

Discussion—Section 201(h) of the Federal Food, Drug & Cosmetic Act (FD&C; 21 U.S.C. 301) defines a *medical device* as "an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is. . .intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals . . ." This definition would appear to include CNADs. As the field of computerized assessment evolves, it is reasonable to assume that such tools will come under the regulatory authority of the Food and Drug Administration (FDA), the agency that, under federal law, regulates all drugs and medical devices.

When considering the use of a CNAD in a particular setting, end users need to know the answers to questions such as (a) What does the device claim to do?, (b) What does it actually do?, (c) How does it do what it claims to do?, and (d) Is it safe and effective? Such answers are critical in making informed decisions about the reliability, validity, and clinical utility of CNADs in any particular setting. The claims made by the developer are critical to the ability of the end user to evaluate the product. Some products may claim to be stand alone diagnostic tests for specific conditions whereas others may purport increased technical accuracy over examiner-administered neuropsychological assessment. The marketing claims and claims made in the documentation for the measure should be evaluated in light of the data presented and the technical information included in the manual, similar to the evaluation of examiner-administered neuropsychological measures. In addition, understanding the mechanism by which the measure provides diagnostic information or normative comparisons is critical for users of CNADs. Even if the algorithm is proprietary, the methodology must be understandable and transparent so that the user can evaluate the validity of the claim made.

Because some CNADs are intended for end-users who have no neuropsychological expertise, claims may be difficult to evaluate by the potential user. When users consider installing a CNAD in their practice or research program, key information about the device, its intended application, and its expected performance is needed. In this regard, it is important to distinguish among (a) "marketing" (the healthcare professional to whom the device is targeted), (b) "labeling" (the information about the device that is provided on packaging or accompanying inserts), (3) "use" (the intended application for the device), and (4) "documentation" (information that accompanies the device, including installation instructions, normative data, or information about device utility). Readers who consult

---

[1]The authors suggest, in the event that specific position statements are quoted in secondary sources, that the acronym "CNAD" be spelled out as "computerized neuropsychological assessment devices" for clarity in the secondary source.

device websites will discover broad variation in the extent to which such information is provided by developers of commercially available CNADs.

Regarding safety and efficacy, developers of CNADs should be explicit about how the test should be used in deriving diagnostic or prognostic statements. Some CNADs provide simple, easy-to-use reports that represent performance levels in color-coded fashion, with "red" indicating a cause for concern, and "green" suggesting normal results, similar in principle to the results of a metabolic panel. Though this simplifies the use of the device, it can obscure the need to consider other data (e.g., additional laboratory, neuroimaging, or clinical data) needed to establish a complex diagnosis and it contains little guidance regarding differential diagnostic considerations in the individual case. Test developers should provide sufficient data to allow users to determine if the CNAD has been previously applied to the problem/condition considered by the user, so that users can determine the suitability of the CNAD for their settings and needs. At the same time, users are expected to implement CNADs in a responsible and appropriate manner. For example, it would not be appropriate to implement a CNAD that had been developed for concussion assessment and management in a clinical practice for use as a dementia screening device in the absence of empirical data supporting its efficacy for that use.

(2) End-User Issues

Position Statement—Developers of CNADs are expected to provide a clear definition of the intended end-user population, including a description of the competencies and skills necessary for effective and accurate use of the device and the data it provides.

Discussion—Some devices are specifically intended as self-test instruments (with the examinee as the end-user), whereas others require test user qualifications similar to those imposed on examiner-administered neuropsychological tests. Still other CNADs are intended for use by health care providers who possess varying knowledge of psychometric principles and/or neuropsychological expertise. Although test administration is likely to be less affected by this lack of knowledge if appropriate orientation to the use of and training on the specific test is undertaken, interpretation of the data generated by the measure may be more substantially affected. Dependent on the intended use or application of the test, a lack of knowledge regarding psychometric properties of the measure, test behavior, associated medical or behavioral data to support interpretation, and neuropsychological expertise, may present a specific challenge to the general health care provider and create a risk to the patient with whom the test is used.

CNADs can be appropriately *administered* by paraprofessionals or technically trained staff who may lack the education, training, or experience necessary to integrate or interpret test results. However, unlike examiner-administered testing, many CNADs are intended to support clinical interpretations rendered by practitioners who have little or no expertise, training, or experience in psychometrics or clinical neuropsychology. The safe and competent use of CNADs requires a link to professionals trained in the use of psychometric techniques in the differential diagnostic setting. The appropriate process of *test interpretation* involves an integration of quantitative test findings with information from medical records, including disease course, functional impairment, comorbid illnesses, history, and other relevant factors. Also, an understanding that multiple factors separate from central nervous system disease or injury (e.g., premorbid abilities, general health, neuropsychiatric and emotional status, medications, fatigue, and effort) can affect performance on cognitive tests is critical to accurate interpretation of test results. Bypassing careful clinical interpretation may lead to potential misuse of the data or failure to consider potential clinical or methodological issues that could influence the results (American

Psychological Association, 1986). This issue is also relevant to the application of CNADs in settings in which a professional is not available to make behavioral observations of the examinee during the testing session. Indeed, some CNADs are designed to be administered to a patient or client with minimal or no direct observation by a trained examiner, and some CNADs that do involve observation by the examiner are intended for use by professionals or paraprofessionals with limited education, training, or experience in neuropsychological assessment. In these instances, behavioral indicators of emotional, motivational, or mental status issues that might complicate test interpretation may be inadvertently missed.

Test developers have taken one of two broad approaches to the user qualifications issue. Some CNADs require appropriate licensure or certification in relevant healthcare fields (e.g., psychology, medicine), thus making the device available to an "expert user". Another approach has been to develop an interpretive algorithm within the device software that essentially creates an "expert system". In this approach, the program itself contains clinical actuarial routines that generate clinical findings and recommendations. However, this poses two challenges for end users: (a) they might lack the knowledge and training to independently evaluate the accuracy of the output and/or the claims the developer makes regarding the test results; and (b) the proprietary interpretive routines might be opaque or ill-defined methodologically, obscuring critical evaluation by the end user or scientific community. Test developers should make sufficient information available to the end user (without compromising proprietary information or trade secrets) so that an independent evaluation of the validity of the interpretive report can be made, and so that the utility of the test in actual clinical practice can be independently evaluated.

In considering the broader context in which CNADs are applied today, it is important to distinguish between neuropsychological ***testing*** (utilizing cognitive tests to obtain behavioral samples of abilities in memory, concentration, or executive function) and neuropsychological ***assessment*** (providing a comprehensive evaluation of an individual that integrates test results with history, symptoms, behavioral observations, physical findings, and other aspects of the examinee's situation to yield interpretive statements about the underlying causes of the patient's performance pattern; Matarazzo, 1990). The interpretation of CNAD results requires similar specialized training and expertise in clinical neuropsychology as the interpretation of examiner-administered neuropsychological tests. Moreover, the interpretation of computerized test results, like their examiner-administered counterparts, occurs in the context of knowledge of relevant information from the social, medical, mental health, educational, and occupational history of the examinee (Matarazzo, 1990). Because of this, the specific interpretive statements generated by CNAD software may not apply to the individual examinee. If applied in practice, these automated reports should be carefully reviewed by someone with expertise in neuropsychological test interpretation for accuracy and relevance related to each individual examinee in each specific case. Consistent with professional competence, clinicians "do not promote the use of psychological assessment techniques by unqualified persons, except when such use is conducted for training purposes with appropriate supervision" (APA, 2010, Ethical Standard 9.07, Assessment by Unqualified Persons).

(3) Technical (Hardware/Software/Firmware) Issues

Position Statement—Test developers should provide users with sufficient technical information to insure that the local installation of a CNAD will produce data that can be accurately compared to that which exists in the test's normative database.

Discussion—As is true of examiner-administered assessment instruments, CNADs are developed within a specific environment that helps define the domains to which the test and

its results can be generalized. Technical aspects of the computing environment in which the test was developed may critically affect how the test performs when applied in clinical settings (Cernich, Brennana, Barker, & Bleiberg, 2007). Such aspects include the computer or tablet's operating system, the speed of the central processing unit (CPU), the amount of available memory, how the program clock interacts with the system clock (McInnes & Taylor, 2001), resolution and refresh rate of the display (Gofen & Mackeben, 1997), characteristics of the user interface (Gaver, 1991), and other aspects of the operating system environment (Forster & Forster, 2003; Myors, 1999; Plant & Turner, 2009; Plant, Hammond, & Turner, 2004). Even subtle differences between the performance of the test in standardization and its performance when locally installed on the user's computer, tablet, or handheld device can influence whether the test performs as advertised. For example, performance indices that rely on millisecond distinctions between groups become less discriminative if the operating environment is clouded by operating system interference, security verification, and/or commonly scheduled program updates that interfere with timing resolution (Creeger, Miller, & Paredes, 1990).

CNADs installed on a user's local machine should duplicate with sufficient accuracy the computing environment in which the normative performance data for the test were established. If this fidelity cannot be demonstrated or confirmed, users have reason to doubt the results of the CNAD. Test developers are expected to provide specific guidance to users that will enable them to determine whether their local installation meets certain technical criteria, including a clear description of the necessary hardware and software configuration, and a developer-provided diagnostic that allows the user to determine that the test has been properly installed and will operate with fidelity to the normative installation.

(4) Privacy, Data Security, Identity Verification, and Testing Environment

Position Statement—Ultimately, maintaining patient privacy and security is the responsibility of the healthcare professional who collects, stores, and/or transmits personal health information, and users of CNADs should have appropriate knowledge about information technology that assures that patient rights are protected. Maintaining this responsibility requires the provision of detailed information about how the CNAD collects and stores patient data. If data is to be transmitted to remote servers or databases for normative referencing or automated report generation, users need to understand how that data is protected from security intrusion, corruption, or other threats to data integrity and privacy. Test developers provide a procedure to verify the identity of examinees who complete a CNAD remotely.

Discussion—Some CNADs store patient data files on a local hard disk, whereas others utilize a "store and forward" web interface in which the patient's data are collected locally and then uploaded via a web connection after testing is completed (Cernich et al., 2007). Users need detailed information about security precautions in place when such data are transmitted, stored, and accessed, and what procedures are in place in the event of inadvertent data loss. Because users have legal obligations to examinees imposed by the HIPAA Security Rule, civil rights legislation, and ethical guidelines, they also need assurances about the security and privacy of data that are transmitted over the web to remote databases (American Psychological Association, 2007). Prevailing law and best practice requires the use of encryption technologies that offer a measure of protection from unauthorized intrusion. Users need to be informed about, and aware of, the unique characteristics of electronic data, how to protect privacy when transmitting data to remote sites, and challenges that exist in disposing of electronic data.

By design, CNADs can be administered remotely, and identity verification can pose particular challenges for implementation of computerized measures when used in this way. Though this does not pose particular problems for in-person administration of a CNAD, remote use, especially in situations where there is no proctor to verify identity, presents logistical and ethical issues (Naglieri, Drasgow, Schmit, et al., 2004). In internet based applications, individuals could be represented by accomplices or assisted by individuals to either feign or enhance their test performance. Even with security protocols that include provision of personal information to verify identity, an informed accomplice could assist. In certain settings (e.g., Department of Defense), systems are being developed wherein the person being assessed is identified by their personal identification verification card or common access card (CAC). This is a federal access card that includes password protection and personal information about the individual, including a biometric identifier (fingerprint; Department of Defense, 2008). Though these systems are restricted in nature, they may provide a template by which individuals could be identified for remote testing in a secure and authentic manner. Although it is not suggested that remote CNAD's must contain such sophisticated biometric identification routines, it is important that developers address identity verification concerns in a thoughtful way so that users can be reasonably assured that the remotely-assessed examinee is who s/he purports to be.

Remote testing, of course, creates special challenges relating to the reliability and validity of the results—given that it can be difficult to control the environment in which the administration occurs. For example, tasks that are dependent on precise presentation of stimuli or that require motor responses may be performed differently if the examinee is lounging on a couch using a laptop than if the examinee is seated in an office environment at a well-lit desk. Test developers should provide general guidance regarding characteristics of the testing environment that are reasonably likely to affect test performance so that users can advise examinees about such environmental considerations.

(5) Psychometric Development Issues

Position Statement—CNADs are subject to the same standards and conventions of psychometric test development, including descriptions of reliability, validity, and clinical utility (accuracy and diagnostic validity), as are examiner-based measures. Psychometric information should be provided to potential users of the CNAD in a manner that enables the user to ascertain the populations and assessment questions to which the test can be appropriately applied. Test developers should provide psychometric data relevant to the claimed purpose or application of the test. The actual data provided may vary depending on whether the test's claimed purpose is to provide a description of cognitive functions or domains versus assisting with the identification of the cognitive sequelae of specific diseases, injuries, or conditions. When established examiner-administered tests are offered in a computerized version, new psychometric data that describe the CNAD version are required. Information about how the data is scored, transformed, and analyzed to produce the CNADs output statistics should be provided with sufficient clarity so that users understand the meaning of the results they produce.

Discussion—Prevailing ethical standards (APA, 2010) state that, "Psychologists who develop tests and other assessment techniques use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias, and recommendations for use" (Standard 9.05). Although these standards are not binding on non-psychologist developers, the fact remains that, in order to be useful and meaningful in practice, all cognitive tests must meet minimum psychometric standards for reliability and validity. Reliability refers to the consistency of a test's output and pertains to both *test scores* and the *clinical inferences* derived from test

scores (c.f., Franzen, 1989, 2000). Reliability can be evaluated through several kinds of evidence, including (a) consistency across test items (internal consistency), (b) consistency over time (test retest reliability or test stability), (c) consistency across alternate forms (alternate form reliability), and (d) consistency across raters (inter-rater reliability).

Validity refers to the degree to which a test measures the construct(s) it purports to measure. According to classical test theory (c.f., Downing & Haladyna, 2006), types of validity include (a) content validity, (b) criterion-related validity (e.g., concurrent and predictive validity), and (c) construct validity (e.g., convergent and discriminant validity). Validity may be defined as the extent to which theory and empirical evidence support the *interpretation(s)* of test scores when they are *used as intended* by the test developer or test publisher (American Psychological Association, 1999; Messick, 1989; Pedhazer & Pedhazer Schmelkin, 1991). In other words, validity is a property of the *interpretation or meaning* attached to a test score within a specific context of test usage, not a property of a given test (Cronbach, 1971, p. 447; c.f., Franzen, 1989; Franzen, 2000; Urbina, 2004).

Reliability and validity are not unitary psychometric constructs. Instead, they are measured in studies in different clinical contexts with diverse populations. Moreover, reliability and validity should be viewed as a matter of degree rather than in absolute terms, and tests must be re-evaluated as populations and testing contexts change over time (Nunnally & Bernstein, 1994). Developers of CNADs are encouraged to update their psychometric studies and their normative databases over time. Working knowledge of reliability and validity, and the factors that impact those psychometric constructs, is a central requirement for responsible and competent test use, whether the measure is used for diagnostic or research purposes.

Neuropsychological tests yield scores that are derived from a comparison of a person's performance to the performance of a healthy normative sample, clinical samples, one's own expected level of performance or, in the case of symptom validity tests, research participants who had been given specific instructions to perform in a certain manner. The quality and representativeness of normative data can have a major effect on the clinical interpretation of test scores (c.f., Mitrushina, Boone, Razani, & D'Elia, 2005). APA (2010) Ethical Standard 9.02 (Use of Assessments), section (b) states, "Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation."

It cannot be assumed that the normative data obtained for an examiner-administered test apply equally well to a computerized version of the same test, due to changes in the method used to conduct the administration and variations in computer familiarity according to patient demographics. Studies of the comparability of computerized measures that are adaptations of examiner administered tests indicate that there are substantive differences in some samples (Berger, Chibnall, & Gfeller, 1997; Campbell et al., 1999; Choca & Morris, 1992; Ozonoff, 1995; Tien et al., 1996), further demonstrating the need for new normative data obtained with the computerized test. As we have indicated above, a computerized test adapted from an examiner administered test is a ***new test***. As a result, it is essential that ***new*** normative data with adjustments for the pertinent demographic variables be established for computerized tests. The relevant standard from the APA 2010 Ethics Code, Standard 9.02 (Use of Assessments) section (a) states that, "Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques."

Prior to using tests diagnostically, it is important to have information relating to their accuracy for that purpose (Newman & Kohn, 2009). Operating characteristics (sensitivity, specificity, positive predictive power, and negative predictive power) are important and should be considered when using a test in a specific clinical setting. *Sensitivity* is the accuracy of a test for identifying a condition of interest (i.e., the proportion with the condition that is correctly identified); such cases are considered true positive results. *Specificity* is the proportion of people who do not have the condition who are correctly identified; such cases are considered true negative results. *Positive predictive value* (PPV) is the probability that a person has a disease or condition given a positive test result; that is, the proportion of individuals with positive test results who are correctly identified by the test. *Negative predictive value* (NPV) is the probability that a person does not have a disease or condition given a negative test result; that is, the proportion of individuals with negative test results who are correctly identified as not having the condition. PPV and NPV are related to the base rate, or prevalence in a given population, of the condition/disease that one is trying to identify. For example, a sensitive test may result in many false positive results (low PPV) if the prevalence of the condition of interest is low. Research relating to sensitivity, specificity, and diagnostic classification accuracy (Retzlaff & Gibertini, 2000) is an important foundation for the proper use of CNADs just as it is for examiner-administered neuropsychological tests. Research relating to diagnostic validity must be evaluated with a critical eye toward the actual diagnostic question to which the CNAD will be applied. For example, it is important to know whether a measure is useful in differentiating patients with dementia from neurologically intact individuals, or whether it can also be useful in making the more difficult distinction between those with dementia and those with mild cognitive impairment.

Data included in the technical manual should be appropriate to the use of the test intended by the developer. Dependent on the claim made, information related to reliability, validity, and diagnostic classification should be included to allow the user to evaluate the utility of that specific CNAD in fulfilling its claimed purpose. Also, test developers are encouraged to make clear to the user what other information is required in the applied context in order to use the test most appropriately. This can be done, for example, by relating performance on the test to prevailing diagnostic algorithms with proven validity in the clinical literature.

(6) Examinee Issues: Cultural, Experiential, and Disability Factors

Position Statement—Test developers should provide appropriate normative information that allows the user to determine whether the CNAD can be given to patients from different racial, ethnic, age, and educational backgrounds. Some patients with cognitive, motor, or sensory disabilities might have difficulty completing a computerized test in the manner intended by the developer. In addition, individual differences in computer use and familiarity may affect how examinees interact with devices, utilize response modalities, or respond to stimuli. Test developers are encouraged to provide documentation that such factors have been accounted for during test standardization and validation, and should provide guidance to users with regard to how motor, sensory, or cognitive impairment in targeted patient populations may affect their performance on the test. It is particularly important to specify conditions under which the test should not be used in patients with motor, sensory, or cognitive impairment.

Discussion—As with examiner-administered neuropsychological assessment, computerized testing has limitations regarding the scope of information that can be obtained and the validity of data that are collected. One key aspect of this issue is how the physical, psychiatric, or neurologic condition of the patient affects his or her ability to interact with the computer interface. For example, computerized assessment places demands on the

examinee's academic skills (e.g., reading), cognitive function (e.g., attention), and sensory and motor functioning (e.g., rapid reaction time) that may influence the results if the examinee has disabilities or limitations in one or more of these areas. If the examinee does not comprehend task instructions, the results of the test will not be valid, and if the program requires speeded motor responses as a proxy for cognitive processing speed, patients with bradykinesia, tremors, or hemiparesis may be significantly compromised for reasons apart from impairments in the targeted construct. With the hemiparetic patient, validity or reliability of the measure might be diminished if they use their non-dominant hand to manipulate the mouse or provide a motor response. As with examiner-administered tests, numerous similar examples can be envisioned that complicate the quality of the data that emerge from CNADs (Hitchcock, 2006).

A key issue with many CNADs is that they may not include plans for consistent observation of the examinee by a trained examiner. Therefore, clinically useful information may be missed relating to task engagement, display of emotion, frustration, or tendency to give up easily when confronted with more challenging test items. Significant individual differences exist in computer use and familiarity (Iverson, Brooks, Ashton, Johnson, & Gualtieri, 2009), and results from computerized versus examiner-administered testing may be different in computer-competent versus computer-naïve populations (Feldstein et al., 1999).

Computerized assessment is constrained by the current hardware and software limitations of the field. Consequently, assessment of some important and sensitive aspects of cognitive functioning, such as free recall (vs. recognition) memory, expressive language, visual-constructional skills, and executive functioning may be difficult to incorporate into a CNAD. Clinicians utilizing CNADs in practice are responsible for recognizing the limitations of this testing approach and for appropriately documenting the impact such factors may have upon their findings. In situations involving examinees who require special testing accommodations as a result of sensorimotor limitations, aphasia, dyslexia, confusion, or variable cooperation, or in those instances that require the assessment of individuals who are less facile or comfortable with computers and tablets, examiner-administered testing may be advantageous or preferred.

(7) Use of Computerized Testing and Reporting Services

Position Statement—Professionals "select scoring and interpretation services (including automated services) on the basis of evidence of the validity of the program and procedures as well as on other appropriate considerations" (APA, 2010, Ethical Standard 9.09, Test Scoring and Interpretation Services, section b). Those "who offer assessment or scoring services to other professionals accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use" (APA, 2010, Ethical Standard 9.09, Test Scoring and Interpretation Services, section a). Professionals "retain responsibility for the appropriate application, interpretation, and use of assessment instruments, whether they score and interpret such tests themselves or use automated or other services" (APA, 2010, Ethical Standard 9.09, Test Scoring and Interpretation Services, section c).

Discussion—Professionals who lack training and expertise in clinical assessment might be tempted to simply accept the content of automated reports that provide descriptive or interpretive summaries of test results, and to incorporate textual output from CNADs into their standard clinical reports. This might occur because clinicians assume, uncritically or without sufficient evidence, that such summaries accurately reflect an individual patient's status and that the scientific bases of such interpretations have been established in the clinical setting. Practitioners are encouraged to evaluate the accuracy and utility of

automated clinical reports in light of the total corpus of information available on the patient, including symptom reports, functional abilities, personal and family history, and other relevant factors. Automated reports are best viewed as an important *resource* for knowledgeable professionals, rather than as a *substitute* for necessary and sufficient expertise.

(8) Checks on Validity of Responses and Results

Position Statement—Examinee compliance, cooperation, and sufficient motivation are essential to the process of obtaining valid neuropsychological test data (American Academy of Clinical Neuropsychology, 2007; Bush et al., 2005; Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). Developers of CNADs are encouraged to address these issues during test development and standardization. It is important for test developers to consider carefully the role of motivation and effort when conducting computerized testing. This is particularly true for CNADs intended for use by professionals unfamiliar with the signs and consequences of reduced effort on cognitive test performance. Test developers are encouraged to (a) provide information on how poor effort can be identified by patterns of performance on the CNAD, or (b) make specific recommendations about additional tests or procedures that can be concurrently conducted to evaluate examinee effort.

Discussion—Over the past few decades, research on effort and its effects on the validity of neuropsychological test results has dominated forensic neuropsychology, and significant interest has been devoted to understanding the role of effort and motivation in producing impairments on a wide variety of neuropsychological tests (Boone, 2007; Larrabee, 2007; Sweet, King, Malina, Bergman, & Simmons, 2002). Effort has been shown to substantially influence neurocognitive test scores, and in some studies, the variance attributable to effort is greater than that attributable to injury severity or other variables more directly related to underlying pathophysiology (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Green, Rohling, Lees-Haley, & Allen, 2001; Stevens, Friedel, Mehen, & Merten, 2008; West, Curtis, Greve, & Bianchini, 2011). These findings lead to the inescapable conclusion that carefully considering patient motivation and effort is a mainstream part of clinical practice (Boone, 2009). Without some form of assurance that the examinee has put forth adequate effort in completing neuropsychological tests, the clinician cannot interpret low test scores as evidence of impairment in a particular neurocognitive ability.

The assessment of effort requires the use of empirically derived indicators. Behavioral observations made during testing by a trained examiner are also useful but may suffice only when the lack of cooperation and effort affects overt behavior. Test developers are encouraged to provide users with procedural guidance about how to identify poor effort on the CNAD. This can be done by documenting a builtin measure of effort that has been appropriately validated within the CNAD, or by providing specific recommendations regarding other validated tests of effort that should be administered along with the CNAD.

CONCLUSIONS

This position paper is intended to provide guidance for test developers and users of CNADs that will promote accurate and appropriate use of computerized tests in a way that maximizes clinical utility and minimizes risks of misuse. We fully recognize the tension that exists between industry and professional users in bringing CNADs to market. On the one hand, there is substantial need to improve access to neurocognitive testing for underserved patients who, by virtue of economic, socioeconomic, geographical, logistical, or cultural reasons are not referred for, or cannot access, needed services. On the other hand, the development of CNADs is a complex enterprise; the tests themselves measure complex

constructs, the technical and information technology issues that ensure appropriate installation of the test in the local environment are nuanced, and the manner in which different patient groups interact with the assessment device introduces important sources of variance that can affect the interpretation of the test results.

Although there are clear differences between stand-alone computerized platforms of common examiner-administered tests (e.g., Wisconsin Card Sorting Test) and full-fledged computerized testing systems, all developers of CNADs are encouraged to provide users with core information regarding (a) test reliability, validity, accuracy, and utility; (b) technical specifications, including how to insure that the local installation faithfully duplicates the environment in which normative data was collected; (c) methods to protect privacy and data integrity; (d) the minimal qualifications of those who can install, administer, or interpret the test; (e) further requirements regarding utilization of computerized or actuarial reporting services; (f) information on who can and cannot benefit from undergoing assessment; (g) what the test claims to be able to do for the patient and/or professional user and (h) guidance with regard to how submaximal effort affects test results and how to interpret results when the examinee intentionally or unintentionally underperforms due to reasons other than neurocognitive compromise.

Computerized neuropsychological assessment devices (both individual tests and test batteries) are expected to meet the same psychometric standards of adequate reliability and validity for the intended clinical populations as examiner-administered neuropsychological tests. Adaptation of an examiner-administered test for computers or tablets should be accompanied by the development of normative or equivalency data for the computerized version; a computerized version is a new test and not merely a slightly different format for an existing test. Expertise in the interpretation of computerized tests requires advanced knowledge of testing theory and the complex interaction of multiple factors that can affect performance on cognitive tests, aside from putative or clearly established injury to the brain. Such expertise is typically obtained from specialized education, training, and experience in clinical neuropsychology.

Qualified test users understand that CNAD results must be interpreted in the context of relevant history, other test findings, and data available from other disciplines. For test results to be considered valid, all neuropsychological testing, including computerized testing, requires adequate motivation and cooperation from examinees.

It is clear that the competent use of appropriately developed computerized neuropsychological measures will serve an increasingly important role in the evaluation of a variety of patient populations. The use of CNADs clearly has a role in bringing valid and effective neuropsychological evaluation techniques to underserved populations. However, such application should proceed with an understanding that effective use of such techniques is not "plug and play", but in fact requires attention to a broad range of factors that determine whether the test will be useful, accurate, and appropriate in the intended setting. Users and consumers of CNADs must be mindful that ethical and clinically useful practice requires that such tests meet appropriate quality and efficacy criteria, and that those employing CNADs have the education, training, and experience necessary to interpret their results in a manner that will best meet the needs of the patients they serve.

Acknowledgments

REFERENCES

American Academy of Clinical Neuropsychology. American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. The Clinical Neuropsychologist. 2007; 21(2):209–231. [PubMed: 17455014]

American Psychological Association. Record keeping guidelines. American Psychologist. 2007; 62:993–1004. [PubMed: 18085845]

American Psychological Association. Standards for educational and psychological testing. American Psychological Association; Washington, DC: 1999.

American Psychological Association. Guidelines for computer-based tests and interpretations. American Psychological Association; Washington, DC: 1986.

Anger WK, Storzbach D, Binder LM, Campbell KA, Rohlman DS, McCauley L, et al. Neurobehavioral deficits in Persian Gulf veterans: evidence from a population-based study. Portland Environmental Hazards Research Center. Journal of the International Neuropsychological Society. 1999; 5(3):203–212. [PubMed: 10217920]

APA. Ethical Principles of Psychologists and Code of Conduct (2010 Amendments). 2010. from http://www.apa.org/ethics/code/index.aspx

Berger SG, Chibnall JT, Gfeller JD. Construct validity of the computerized version of the Category Test. Journal of Clinical Psychology. 1997; 53(7):723–726. [PubMed: 9356901]

Bleiberg J, Cernich AN, Cameron K, Sun W, Peck K, Ecklund PJ, et al. Duration of cognitive impairment after sports concussion. Neurosurgery. 2004; 54(5):1073–1078. discussion 1078–1080. [PubMed: 15113460]

Bleiberg J, Garmoe WS, Halpern EL, Reeves DL, Nadler JD. Consistency of within-day and across-day performance after mild brain injury. Neuropsychiatry. Neuropsychology, and Behavioral Neurology. 1997; 10(4):247–253.

Bolfer C, Casella EB, Baldo MV, Mota AM, Tsunemi MH, Pacheco SP, et al. Reaction time assessment in children with ADHD. Arquivos de Neuro-Psiquiatria. 2010; 68(2):282–286. [PubMed: 20464301]

Boone KB. The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. The Clinical Neuropsychologist. 2009; 23:729–741. [PubMed: 18949583]

Boone, KB., editor. Assessment of feigned cognitive impairment: A neuropsychological perspective. Guilford Press; New York: 2007.

Broglio SP, Ferrara MS, Macciocchi SN, Baumbartner TA, Elliott R. Test-retest reliability of computerized concussion assessment programs. Journal of Athletic Training. 2007; 42:509–514. [PubMed: 18174939]

Brooks BL, Iverson GL, Sherman EM, Roberge MC. Identifying cognitive problems in children and adolescents with depression using computerized neuropsychological testing. Applied Neuropsychology. 2010; 17(1):37–43. [PubMed: 20146120]

Bush SS, Ruff RM, Troster AI, Barth JT, Koffler SP, Pliskin NH, et al. Symptom validity assessment: practice issues and medical necessity NAN policy & planning committee. Archives of Clinical Neuropsychology. 2005; 20(4):419–426. [PubMed: 15896556]

Campbell KA, Rohlman DS, Storzbach D, Binder LM, Anger WK, Kovera CA, et al. Test-retest reliability of psychological and neurobehavioral tests self-administered by computer. Assessment. 1999; 6(1):21–32. [PubMed: 9971880]

Cernich A, Reeves D, Sun W, Bleiberg J. Automated Neuropsychological Assessment Metrics sports medicine battery. Archives of Clinical Neuropsychology. 2007; 22(Suppl 1):S101–114. [PubMed: 17118625]

Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. Archives of Clinical Neuropsychology. 2007; 22S:S39–S48. [PubMed: 17097851]

Chamberlain SR, Robbins TW, Winder-Rhodes S, Muller U, Sahakian BJ, Blackwell AD, et al. Translational approaches to frontostriatal dysfunction in attention-deficit/hyperactivity disorder

using a computerized neuropsychological battery. Biological Psychiatry. 2011; 69(12):1192–1203. [PubMed: 21047621]

Choca J, Morris J. Administering the Category Test by computer: Equivalence of results. The Clinical Neuropsychologist. 1992; 6(1):9–15.

Collie A, Makdissi M, Maruff P, Bennell K, McCrory P. Cognition in the days following concussion: comparison of symptomatic versus asymptomatic athletes. Journal of Neurology, Neurosurgery, and Psychiatry. 2006; 77(2):241–245.

Collins MW, Lovell MR, Iverson GL, Ide T, Maroon J. Examining concussion rates and return to play in high school football players wearing newer helmet technology: a three year prospective cohort study. Neurosurgery. 2006; 58(2):275–286. [PubMed: 16462481]

Constantinou M, Bauer L, Ashendorf L, Fisher JM, McCaffrey RJ. Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tasks? Archives of Clinical Neuropsychology. 2005; 20:191–198. [PubMed: 15708729]

Creeger CP, Miller KF, Paredes DR. Micromanaging time: Measuring and controlling timing errors in computer-controlled experiments. Behavior Research Methods, Instruments, and Computers. 1990; 22:34–79.

Cronbach, LJ. Test validation. Educational measurement. 2nd ed.. Thorndike, RL., editor. American Council on Education; Washington, DC: 1971. p. 443-507.

Crook, TH.; Kay, GG.; Larrabee, GJ. Computer-based cognitive testing.. In: Grant, I.; Adams, KM., editors. Neuropsychological Asessment of Neuropsychiatric and Neuromedical Disorders. 3rd ed.. Oxford University Press; New York: 2009. p. 84-100.

Doniger GM, Dwolatzky T, Zucker DM, Chertkow H, Crystal H, Schweiger A, et al. Computerized cognitive testing battery identifies mild cognitive impairment and mild dementia even in the presence of depressive symptoms. American Journal of Alzheimer's Disease and Other Dementias. 2006; 21(1):28–36.

Doniger GM, Zucker DM, Schweiger A, Dwolatzky T, Chertkow H, Crystal H, et al. Towards practical cognitive assessment for detection of early dementia: a 30-minute computerized battery discriminates as well as longer testing. Current Alzheimer Research. 2005; 2(2):117–124. [PubMed: 15974907]

Dorion AA, Sarazin M, Hasboun D, Hahn-Barma V, Dubois B, Zouaoui A, et al. Relationship between attentional performance and corpus callosum morphometry in patients with Alzheimer's disease. Neuropsychologia. 2002; 40(7):946–956. [PubMed: 11900746]

Downing, SM.; Haladyna, TM., editors. Handbook of test development. Lawrence Erlbaum Associates Publishers; Mahwah, NJ: 2006.

Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, et al. Validity of the Mindstreams computerized cognitive battery for mild cognitive impairment. Journal of Molecular Neuroscience. 2004; 24(1):33–44. [PubMed: 15314247]

Feldstein SN, Keller FR, Protman RE, Durham RL, Klebe KJ, Davis HP. A comparison of computerized and standard version of the Wisconsin Card Sorting Test. The Clinical Neuropsychologist. 1999; 13:303–313. [PubMed: 10726602]

Forster KI, Forster JC. DMDX: A Windows display proram with millisecond accuracy. Behavior Research Methods, Instruments, and Computers. 2003; 35:116–124.

Franzen, MD. Reliability and validity in neuropsychological assessment. Plenum Press; New York, NY: 1989.

Franzen, MD. Reliability and validity in neurological assessment. 2nd ed.. Kluwer Academic/Plenum Press; New York, NY: 2000.

Gaver, WW. Technology affordances. In CHI 1991 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching through Technology. Association for Computing Machinery; New York: 1991.

Gofen A, Mackeben M. An introduction to accurate display timing for PC's under "Windows". Spatial Vision. 1997; 10(4):361–368. [PubMed: 9176945]

Green P, Rohling ML, Lees-Haley PR, Allen LM. Effort has a greater effect on test scores than severe brain injury in compensation claimants. Brain Injury. 2001; 15:1045–1060. [PubMed: 11712951]

Green P, Rohling ML, Lees-Haley PR, Allen LM 3rd. Effort has a greater effect on test scores than severe brain injury in compensation claimants. Brain Injury. 2001; 15(12):1045–1060. [PubMed: 11712951]

Gualtieri CT, Johnson LG. Neurocognitive testing supports a broader concept of mild cognitive impairment. American Journal of Alzheimer's Disease and Other Dementias. 2005; 20(6):359–366.

Gualtieri CT, Johnson LG. Efficient allocation of attentional resources in patients with ADHD: maturational changes from age 10 to 29. Journal of Attentional Disorders. 2006; 9(3):534–542.

Gualtieri CT, Johnson LG. A computerized test battery sensitive to mild and severe brain injury. Medscape Journal of Medicine. 2008; 10(4):90. [PubMed: 18504479]

Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. The Clinical Neuropsychologist. 2009; 23(7):1093–1129. [PubMed: 19735055]

Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, Participants C. American Academy of Clinical Neuropsychology Consensus Conference statement on the neuropsychological assessment of effort, response bias, and malingering. The Clinical Neuropsychologist. 2009; 23(7):1093–1129. [PubMed: 19735055]

Hitchcock E. Computer access for people after stroke. Topics in Stroke Rehabilitation. 2006; 13:22–30. [PubMed: 16987789]

Iverson GL, Brooks BL, Ashton VL, Johnson LG, Gualtieri CT. Does familiarity with computers affect computerized neuropsychological test performance? Journal of Clinical and Experimental Neuropsychology. 2009; 31:594–604. [PubMed: 18972312]

Iverson GL, Brooks BL, Collins MW, Lovell MR. Tracking neuropsychological recovery following concussion in sport. Brain Injury. 2006; 20(3):245–252. [PubMed: 16537266]

Iverson GL, Brooks BL, Langenecker SA, Young AH. Identifying a cognitive impairment subgroup in adults with mood disorders. Journal of Affective Disorders. 2011; 132:360–367. [PubMed: 21439647]

Iverson GL, Brooks BL, Lovell MR, Collins MW. No cumulative effects for one or two previous concussions. British Journal of Sports Medicine. 2006; 40(1):72–75. [PubMed: 16371496]

Kramer JJ. On the question of professional standards for computer-based test interpretation. American Psychologist. 1987; 42:889–890.

Larrabee, GJ. Assessment of malingered neuropsychological deficits. Oxford University Press; New York: 2007.

Marx BP, Brailey K, Proctor SP, MacDonald HZ, Graefe AC, Amoroso P, et al. Association of Time Since Deployment, Combat Intensity, and Posttraumatic Stress Symptoms With Neuropsychological Outcomes Following Iraq War Deployment. Archives of General Psychiatry. 2009; 66(9):996–1004. [PubMed: 19736356]

Matarazzo JD. Clinical psychological tst interpretations by computer: Hardware outpaces software. Computers in Human Behavior. 1985; 1:235–253.

Matarazzo JD. Computerized clinical psychological test interpretations: Unvalidated plue all mean and no sigma. American Psychologist. 1986; 41:96. [PubMed: 3954244]

Matarazzo JD. Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. American Psychologist. 1990; 45:999–1017. [PubMed: 2221576]

McInnes WJ, Taylor TL. Millisecond timing on PC's and Macs. Behavior Research Methods Instruments and Computers. 2001; 31(1):129–136.

McLay R, Spira J, Reeves D. Use of computerized neuropsychological testing to help determine fitness to return to combat operations when taking medication that can influence cognitive function. Military Medicine. 2010; 175(12):945–946. [PubMed: 21265298]

Messick, S. Validity.. In: Linn, RL., editor. Educational measurement. 3rd ed.. American Council on Education and Macmillan; New York: 1989. p. 13-103.

Mitrushina, M.; Boone, KB.; Razani, J.; D'Elia, LF. Handbook of normative data for neuropsychological assessment. 2nd ed.. Oxford University Press; New York: 2005.

Moore JL, McAuley JW, Long L, Bornstein R. An Evaluation of the Effects of Methylphenidate on Outcomes in Adult Epilepsy Patients. Epilepsy & Behavior. 2002; 3(1):92–95. [PubMed: 12609358]

Myors B. Timing accuracy of PC programs running under DOS and Windows. Behavior Research Methods, Instruments, & Computers. 1999; 31:322–328.

Naglieri JA, Drasgow F, Schmit M, Handler L, Prifitera A, Margolis A, Velasquez R. Psychological testing on the internet: New problems, old issues. American Psychologist. 59(3):150–162. (date). [PubMed: 15222858]

Newman, TB.; Kohn, MR. Evidence-based diagnosis. Cambridge University Press; New York: 2009.

Nunnally, JC.; Bernstein, IH. Psychometric theory. 3rd ed.. McGraw-Hill, Inc.; New York: 1994.

Ozonoff D. Environmental medicine for all: getting there form here. Lancet. 1995; 346(8979):860. [PubMed: 7564668]

Pedhazer, EJ.; Pedhazer Schmelkin, L. Measurement, design, and analysis: An integrated approach. Lawrence Erlbaum Associates Publishers; Hillsdale, NJ: 1991.

Peterson SE, Stull MJ, Collins MW, Wang HE. Neurocognitive function of emergency department patients with mild traumatic brain injury. Annals of Emergency Medicine. 2009; 53(6):796–803. e791. [PubMed: 19167782]

Plant RR, Turner G. Millisecond precision psychological research in a world of commodity computers: new hardware, new problems? Behavior Research Methods. 2009; 41(3):598–614. [PubMed: 19587169]

Plant RR, Hammond N, Turner G. Self-validating presentation and response timing in cognitive paradigms: how and why? Behavioral Research Methods, Instrumentation, and Computers. 2004; 36(2):291–303.

Polderman TJ, van Dongen J, Boomsma DI. The relation between ADHD symptoms and fine motor control: a genetic study. Child Neuropsychology. 2011; 17(2):138–150. [PubMed: 21113825]

Raymond PD, Hinton-Bayre AD, Radel M, Ray MJ, Marsh NA. Assessment of statistical change criteria used to define significant change in neuropsychological test performance following cardiac surgery. European Journal of Cardiothoracic Surgery. 2006; 29(1):82–88. [PubMed: 16337395]

Reise SP, Waller NG. Item response theory and clinical measurement. Annual Review of Clinical Psychology. 2009; 5:27–48.

Retzlaff PD, Callister JD, King RE. Clinical procedures for the neuropsychological evaluation of U.S. Air Force pilots. Military Medicine. 1999; 164(7):514–519. [PubMed: 10414068]

Retzlaff, PD.; Gibertini, M. Neuropsychometric issues and problems.. In: Vanderploeg, RD., editor. Clinician's guide to neuropsychological assessment. Lawrence Erlbaum Associates, Publishers; Mahwah, NJ: 2000. p. 277-299.

Slick DJ, Tan JE, Strauss E, Mateer CA, Harnadek M, Sherman EM. Victoria Symptom Validity Test scores of patients with profound memory impairment: nonlitigants case studies. The Clinical Neuropsychologist. 2003; 17(3):390–394. [PubMed: 14704889]

Stevens A, Friedel E, Mehen G, Merten T. Malingering and uncooperativeness in psychiatric and psychological assessment: prevalance and effects in a German sample of claimants. Psychiatric Research. 2008; 157:191–200.

Sweeney JA, Kmiec JA, Kupfer DJ. Neuropsychologic impairments in bipolar and unipolar mood disorders on the CANTAB neurocognitive battery. Biological Psychiatry. 2000; 48(7):674–684. [PubMed: 11032979]

Sweet JJ, King JH, Malina AC, Bergman MA, Simmons A. Documenting the presence of forensic neuropsychology at national meetings and in relevant professional journals from 1990 to 2000. The Clinical Neuropsychologist. 2002; 16:481–494. [PubMed: 12822057]

Thomas ML. The value of Item Response Theory in clinical assessment: A review. Assessment. 2010; 18(3):291–307. [PubMed: 20644081]

Tien AY, Spevack TV, Jones DW, Pearlson GD, Schlaepfer TE, Strauss ME. Computerized Wisconsin Card Sorting Test: comparison with manual administration. The Kaohsiung Journal of Medical Sciences. 1996; 12(8):479–485. [PubMed: 8774117]

Tornatore JB, Hill E, Laboff JA, McGann ME. Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. Journal of Neuropsychiatry and Clinical Neurosciences. 2005; 17(1):98–105. [PubMed: 15746489]

Urbina, S. Essentials of psychological testing. John Wiley & Sons; Hoboken, NJ: 2004.

Van Kampen DA, Lovell MR, Pardini JE, Collins MW, Fu FH. The "value added" of neurocognitive testing after sports-related concussion. American Journal of Sports Medicine. 2006; 34(10):1630–1635. [PubMed: 16816151]

Vasterling JJ, Proctor SP, Amoroso P, Kane R, Heeren T, White RF. Neuropsychological outcomes of army personnel following deployment to the Iraq war. Journal of the American Medical Association. 2006; 296(5):519–529. [PubMed: 16882958]

West LK, Curtis KL, Greve KW, Bianchini KJ. Memory in traumatic brain injury: the effects of injury severity and effort on the Wechsler Memory Scale-III. Journal of Neuropsychology. 2011; 5:114–125. [PubMed: 21366889]

Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimer's & Dementia. 2008; 4(6):428–437.

Wouters H, de Koning I, Zwinderman AH, van Gool WA, Schmand B, Buiter M, et al. Adaptive cognitive testing in cerebrovascular disease and vascular dementia. Dementia and Geriatric Cognitive Disorders. 2009; 28(5):486–492. [PubMed: 19940481]

Wouters H, Zwinderman AH, van Gool WA, Schmand B, Lindeboom R. Adaptive cognitive testing in dementia. International Journal of Methods in Psychiatric Research. 2009; 18(2):118–127. [PubMed: 19507163]