

## Chapter 13

# Neuropsychological Assessment Using Virtual Environments: Enhanced Assessment Technology for Improved Ecological Validity

Thomas D. Parsons

University of Southern California,  
Los Angeles, California  
tparsons@ict.usc.edu

**Abstract.** Although today's neuropsychological assessment procedures are widely used, neuropsychologists have been slow to embrace technological advancements. Two essential limitations have resulted from this refusal of technological adaptation: First, current neuropsychological assessment procedures represent a technology that has barely changed since the first scales were developed in the early 1900s. Second, while the historical purpose of clinical neuropsychology was differential diagnosis of brain pathology, technological advances in other clinical neurosciences have changed the neuropsychologist's role to that of making ecologically valid predictions about the impact of a given patient's neurocognitive abilities and disabilities on everyday functioning. After a brief discussion of current applications of computer-based neuropsychological assessment, there is a discussion of an increasingly important topic in recent decades—the design of ecologically valid neuropsychological instruments to address real world outcomes. Finally, there is an exploration of virtual reality environments for ecologically valid neuropsychological assessments that make use of current technological advances. It is concluded that a future possible virtual reality-based neuropsychological assessment battery will combine the control and rigor of technologically advanced computerized laboratory measures, the psychometric rigor (i.e., veridicality) of traditional paper-and-pencil assessments, and verisimilitude approximating real life situations.

## 1 Introduction

Clinical neuropsychology is one of the fastest growing specialty areas in psychology. Clinical neuropsychologists study brain-behavior relationships in both persons with normal cognitive functioning and persons with brain injury or disease. Over the course of the last several decades, clinical neuropsychology has gained increasing recognition as a discipline with relevance to a number of diverse practice areas (e.g., neurology, neurosurgery, psychiatry, and family medicine) as well as neuroscience specific research areas (e.g., behavior, learning, and individual differences). As a result, neuropsychologists must apply a working understanding of psychology, physiology, and

neurology to assess, diagnose, and treat patients with neurological, medical, neurodevelopmental, psychiatric, and cognitive disorders.

### **1.1 Neuropsychological Assessments**

Clinical neuropsychologists take wide-ranging measurements of multiple cognitive domains, including ability to attend and encode information, receptive and expressive language, problem-solving skills, reasoning and conceptualization abilities, learning, memory, perceptual-motor skills (e.g., visuospatial organization; visual-motor coordination), speed of processing, intelligence, academic/vocational skills, behavior, emotions, and personality. As can be seen neuropsychological assessments sample a range of functions. This is an important component in that most neuropsychological measures are not "pure", in that they do not assess one skill only. For example, on a timed task in which the person being assessed is asked to copy figures, a number of domains are being assessed (i.e., motor, visual perceptual, attentional and speed of processing). Take for example, results of a person that showed difficulty on a measure of math skills. These results may reflect limits in, for example, understanding of numerical concepts, remembering math facts, understanding the language of mathematics, remembering which operations to apply when, visualizing concepts, sequencing (e.g., performing the right steps in the right order), and/or attending to visual details (e.g., operational sign, place, columns of numbers). As a result, neuropsychological assessments are structured to investigate aspects or subcomponents of the person's performance that will clarify the nature of their individual strengths and weakness across multiple neuropsychological domains. In addition to using neuropsychological assessments (i.e., tests of neurocognitive, behavioral, and emotional functioning) to form hypotheses regarding a person's neurological functioning, neuropsychologists carefully consider how these factors interact with the individual's psychosocial environment. Data gleaned from these neuropsychological assessments yield a variety of inferences that can be correlated to the person's everyday functioning.

The neuropsychological examination has historically been characterized as both a refinement and an extension of the neurological examination [1]. Much of what is now considered part of neuropsychological assessment originated from attempts of late nineteenth and early twentieth century physicians to improve evaluation of the cognitive capacities of persons with brain disease (e.g., Broca [2] and Wernicke [3] aphasics). Part of this has to do with the fact that many widely used neuropsychological tests were constructed before the advent of neuroimaging and emergence of much of the currently available information relating altered behavior to brain dysfunction. A major problem for the neurological examination is the lack of any standardized method of giving or scoring the neurological exam's procedures. In many cases, the description of administrative procedures is vague. Moreover, neurologists many times change procedures for individual patients. Even when administrative procedures are clear, scoring procedures are not. Scoring is determined by the personal assessment of the neurologist based on experience and knowledge rather than on any normative data. Despite these major problems, the neurological examination clearly possesses a high degree of face validity as well as a strong theoretical foundation. Neuropsychologists have made great strides in proper standardization, scoring, and validation. As such, the revised neurological exam (i.e., neuropsychological assessment) has

become a major tool of both clinical and experimental neuropsychology. The work of A. R. Luria, the Russian neuropsychologist, is a good example of a neuropsychological evaluation derived from a neurological approach. According to Luria, performance on a neuropsychological assessment reveals patterns or “functional systems” of interacting neurological areas that are responsible for a given behavior. Each neurological area of the brain participates in numerous such functional systems. Luria aimed to show that the cognitive sequelae of a brain injury reflects an interruption to the execution of any functional system that includes the injured areas(s). The most significant flaw in the original Luria battery is a lack of standard administration and scoring that has precluded an assessment of its validity. However, there have been attempts to overcome these deficiencies by developing an objective form, combining Luria's procedures with the advantages of a standard test battery (i.e., Luria-Nebraska Neuropsychological Battery, also known as LNNB).

The Halstead-Reitan Battery was developed specifically to detect “organic” dysfunction and differentiate between patients with and without brain damage (e.g., to distinguish “organic” from “functional” disorders). Over the years, tests have been designed in concert with evolving information regarding the mediation of behavior by specific structures or circuits provide greater insight into the integrity or disintegration of neurologic function. Extensive experience with these instruments provides a basis for interpreting the tests in neurologic terms. Another widely used battery of tests includes the Wechsler Scales. While commonly used by neuropsychologists, it is important to note that the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Memory Scale (WMS) were developed without the specific intention of using them as instruments to assess brain function and detect brain disorders, but extensive experience with these instruments has provided a basis for interpreting the tests in neurologic terms. Most current neuropsychological assessment approaches use several of the traditional tests in combination with newer techniques developed specifically to evaluate neurocognitive activities and provide insight into brain function in different disease states. Hence, the range of domains assessed by clinical neuropsychologists has expanded tremendously in recent decades to include areas beyond behavioral neurology and the traditional differentiation between organic and functional conditions in psychiatry. Although there are aspects of neuropsychological assessment that are similar to the conventional evaluation of the behavioral neurologist, neuropsychological measures have the advantage of standardization and psychometric rigor.

## 1.2 Standardization of Neuropsychological Assessment Measures

An important factor in the development of clinical neuropsychology is the establishment of standardized assessment measures capable of identifying the neurocognitive effects of brain dysfunction. Standardized assessment in neuropsychology is largely due to its historic development from Alfred Binet's tests of intelligence [4, 5, 6, 7] and the United States's entry into the World War I in 1917 [8]. During this time Robert Yerkes, Arthur Otis, and the American Psychological Association developed a group administered version of the Stanford-Binet (i.e., Army Alpha), and a novel group administered assessment composed of nonverbal tasks (i.e., Army Beta). Yerkes [9] preferred a point-scale methodology (i.e., tests selected based upon specified functions) over Binet's age-scale approach (i.e., tasks fluctuate with age and

developmental level). Ultimately, the Army group administered measures reflecting an amalgamation of Yerkes's point-scale approach and Binet's task-specific approach to measuring cognitive performance. Further, a performance scale developed by David Wechsler was included in an Army battery, [10] that was made up of subtests developed primarily by Binet and World War I psychologists.

A major shift in testing occurred when Wechsler applied testing procedures (i.e., group and individual) developed for normal functioning persons to the construction of a clinical test battery. Following World War I, Wechsler assembled the Wechsler-Bellevue battery, which included both Verbal and Performance Scales. By the 1940s a number of specialized neurocognitive tests were available to clinicians for assessing the mental capacities of persons with brain disease [11, 12, 13, 14, 15, 16]. The additive effects of these tests provided the foundation for today's neuropsychological assessment procedures [17, 18, 19, 20, 21, 22].

### 1.3 Psychometric Rigor of Neuropsychological Assessment Measures

Neuropsychologists tend to emphasize the psychometric rigor of neuropsychological assessments though their emphasis upon reliability and validity. By reliability, neuropsychologists are referring to the consistency with which the same information is obtained by a given neuropsychological test or set of neuropsychological tests. For scores to be reliable, they should (in the absence of intervening variables like illness, injury, new learning) remain stable. Some examples of reliability include: 1) Interrater reliability: the reliability of test scores when administered by different examiners; 2) Intra-rater reliability: the reliability of test scores when the test is given by the same examiner on more than one occasion; and 3) Test-retest reliability: the reliability of test scores when given to the same patient on different occasions. Hence, reliability of tests refers to the consistency with which the same information is obtained if the test is given by different examiners (interrater reliability), by the same examiner on more than one occasion (intrarater reliability), or to the same patient on different days (test-retest reliability). It is important to note that the reliability for neuropsychological tests of memory is consistently lower than for other types of neuropsychological tests. In such cases, the clinician must bear in mind that conclusions drawn from the tests may be more variable than is desirable.

When neuropsychologists refer to validity, they are referring to how well the test measures what it purports to measure. Specific types of validity that may be questioned include the following: 1) Construct validity: validity of the test to measure what it is supposed to measure; 2) Concurrent validity: strength of the correlation of a new test with existing tests or independent measures of the construct in question; 3) Face validity: appearance ("looks like") of a test to measure what it is supposed to measure; 4) Localization validity: validity of a test to localize focal lesions accurately; and 5) Ecologic validity – validity of the test to predict real life ability and performance on activities of daily living. Hence, validity data can be used by the neuropsychologist to answer questions of psychometric rigor: construct validity (do memory tests assess memory?), concurrent validity (do new tests come to the same conclusions as established tests?), localization validity (do test results localize focal lesions?), diagnostic validity (do tests accurately diagnose disease?), and ecologic validity (do test results predict real life performance?). It is important to note that the

neuropsychologist is careful in the use of the terms reliability and validity. When a neuropsychologist refers to reliability as a concept, it is best to understand this as a reference to scores not to tests and when the neuropsychologist refers to validity, it is to the interpretation of performance on tests and not to tests themselves. For example, the question, "Is a given neuropsychological test a valid test?" is not something that can be readily answered. Instead, the neuropsychologist asks about interpretation of performance. As such, much of what is discussed within validity is in fact a reference to ecological validity (more on this below).

As mentioned above, much of what is now considered part of neuropsychological assessment originated from attempts of late nineteenth and early twentieth century physicians to improve evaluation of the cognitive capacities of persons with brain disease. As such, during a period focusing on localization, neuropsychologists received referrals from neurosurgeons to psychometrically localize brain damage. As a result, developed measures were based upon a localization paradigm that focused upon double dissociation—two neocortical areas are functionally dissociated by two behavioral measures, each measure is affected by a lesion in one neocortical area and not the other [23, 24]. Given the importance of neuropsychological assessment for lesion localization it became increasingly important that neuropsychological assessment have enhanced psychometric rigor. In addition to the reliability and validity issues mentioned above, this also includes issues of sensitivity and specificity. By sensitivity, neuropsychologists are referring to a test's ability to detect even the slightest expression of abnormalities in neurological (primarily central nervous system) function. Sensitivity is understood as a reflection of the neuropsychological test's ability to identify persons with a disorder. This is often referred to as true positive rate. By specificity, neuropsychologists are referring to the ability of a neuropsychological test to differentiate patients with a certain abnormality from those with other abnormalities or with no abnormality. This is often referred to as true negative rate. A score on any test can be a true positive, false positive, true negative, or false negative. For a score to be true positive, it must have high sensitivity to dysfunction, allowing dysfunctions to be detected. If a score on any test is false positive, it indicates sensitivity to dysfunction, but lacks specificity to a particular dysfunction. A score on any test can be a true negative if it has high specificity, allowing negative to be distinguished from others. If a score on any test is false negative, this indicates a lack of sensitivity, without regard to specificity of the test. For any evaluation, it is important to understand the rates of each of the four categories of results. The ability to identify brain dysfunction varies greatly among neuropsychological tests and is determined by the fidelity with which the neuropsychological test distinguishes normal from abnormal function and by the specific type of deficit that the patient exhibits. The WAIS, for example, has no memory subtests and is necessarily insensitive to memory-related deficits, whereas it has demonstrated sensitivity to disorders affecting visuospatial, calculation, and attentional abilities. In general, tests that are timed, requiring the patient to complete the test in a specified period, have greater sensitivity to diffuse or multifocal cerebral changes than untimed tests.

In summary, in a number of ways, clinical neuropsychology can be viewed as representing a synthesis of the best features of neurological, psychiatric, and psychological examination procedures, whereby the systematic neurological assessment of functional cortical and subcortical systems is combined with the precise scaling of

psychometric measurement. Neuropsychological assessment allows the examiner to reduce the subjectivity in traditional neurological examinations by conducting assessments that lead to quantifiable standardized scores, thereby increasing the reliability of the assessment as well as allowing for a more sensitive baseline for comparisons across time. Further, availability of normative data and use of standardized administration procedures allow neuropsychological evaluation to be more sensitive than unstructured mental status testing in the detection of mild cognitive disturbances.

#### **1.4 Paradigm Shift in Neuropsychological Assessment**

It is important to note, however, that with the advent of neuroimaging, the need for neuropsychologists to localize brain damage has been greatly reduced. Unfortunately, many neuropsychologists continue to rely on “localization” as the chief basis for validating neuropsychological tests. As Ronald Ruff has contended, although neuroimaging caused the role of neuropsychology to shift from localization to documentation of neuropsychological deficits for prediction of real world functioning, clinical neuropsychologists many times fail to develop ecologically oriented assessments and continue to use localizationist-developed test batteries [25].

Although today’s neuropsychological assessment procedures are widely used, neuropsychologists have been slow to adjust to the impact of technology on their profession. Two essential limitations have resulted from this refusal of technological adaptation: First, current neuropsychological assessment procedures represent a technology that has barely changed since the first scales were developed in the early 1900s (i.e., Binet and Simon’s first scale in 1905 and Wechsler’s first test in 1939). In order for neuropsychologists to fully embrace the development of new batteries that take real world functioning (i.e., ecological validity) seriously, there is a need for them to move beyond cosmetic changes to standardized tests to computerized measures. However, neuropsychologists have historically resisted embracing technological advances in computation. While neuropsychology emphasizes its role as a science, its technology is not progressing in pace with other science-based technologies. Second, while the historical purpose of clinical neuropsychology was differential diagnosis of brain pathology, technological advances in other clinical neurosciences (e.g., the development of neuroimaging) have changed the neuropsychologist’s role to that of making ecologically valid predictions about the impact of a given patient’s neurocognitive abilities and disabilities on everyday functioning.

The organization of this chapter is as follows. In section two, a brief discussion of current applications of computer-based neuropsychological assessment are described. In section three, there will be a discussion of an increasingly important topic in recent decades—the design of ecologically valid neuropsychological instruments to address real world outcomes. The utility of a virtual environment (VE) for ecologically valid neuropsychological assessments that make use of current technological advances is discussed in section four. A summary of concluding remarks is given in section five.

## **2 Computer-Based Neuropsychological Assessment**

In this section, a brief discussion of current applications of computer-based neuropsychological assessment is described, as follows.

## 2.1 Automated Neuropsychological Assessments

Computer-based neuropsychological assessments offer a number of advantages over traditional paper-and-pencil testing: increased standardization of administration; increased accuracy of timing presentation and response latencies; ease of administration and data collection; and reliable and randomized presentation of stimuli for repeat administrations [26]. It is important to note that neuropsychology as a field has not embraced anarchoprimitivism. In fact, in the 1980s there was some initial interest in computerization of various assessment measures and neuropsychologists transferred a number of paper-and-pencil measures to the personal computer platform. Initial attempts at assessing the equivalence of these measures to traditional tests were made [27]. A few examples of computerized versions of traditional paper-and-pencil neuropsychological tests include: the Raven's Colored Progressive Matrices [28]; the Peabody Picture Vocabulary Test [29]; Category Test subtest of the Halstead Reitan Battery [30]; and the Wisconsin Card Sorting Test [31]. In the past decade, a number of computerized tests on neurocognitive function have been developed: CogSport [32], ImPACT [33], ANAM [34], and HeadMinder [35].

Perhaps the most widely used computerized neuropsychological assessment battery is the Automated Neuropsychological Assessment Metrics (ANAM) battery, which has been given widely to civilian and military populations. In fact, the ANAM battery has been given to over 400,000 predeployed soldiers. The ANAM battery is the result of 30 years of computerized psychological test development sponsored primarily by the U.S. Military—decendent of the joint services Unified Tri-Service Cognitive Performance Assessment Battery (UTCPAB). Currently, ANAM is a major focus on neuropsychological assessment for Military Service Members. Much of this emphasis comes from injuries resulting from conflicts in which Service Members may be injured by explosions, resulting in concussions (mTBI). The ANAM is a validated computer-based tool designed to detect speed and accuracy of attention, memory, and thinking ability. It records a Service Member's performance through responses provided on a computer. It is being conducted prior to deployment and can be used to identify and monitor changes in function. A specialized neuropsychological battery for use in blast injury cases has been developed. The ANAM Traumatic Brain Injury (TBI) Battery is a selection of tests from the ANAM library designed to aid in the assessment of general cognitive function following a head injury. Some of the most useful research on ANAM and TBI has evolved from an ongoing project conducted by the Defense and Veterans Brain Injury Center (DVBIC). DVBIC has extensive databases on selected ANAM tests and has been using the precursor to the ANAM TBI Battery for a number of years. One normative study of ANAM TBI Battery test modules with over 2,000 paratrooper recruits and another study based on over 5,000 recruits revealed consistent results. These studies have provided some of the largest and finest neuropsychological assessment databases for military personnel available and attest to the cost-effective leveraged value of DoD-sponsored ANAM test development and application.

Recently, customized modifications to the ANAM TBI Test Battery made by the University of Oklahoma's Center for the Study of Human Operator Performance (C-SHOP) for the U.S. Army have resulted in the ANAM4 TBI MIL Battery. The ANAM4 TBI and the ANAM4 TBI MIL Batteries do not differ with regard to the

actual ANAM4 test modules presented or the order of test module presentation. The differences between these test modules reside in customized demographic features and characteristics of the ANAM Performance Report, which provide relevance and ease of integration with unique DoD medical records systems and clinical applications. The ANAM4 TBI MIL Battery provides precise, objective, automated measures of fundamental neurocognitive functions including response speed, attention/concentration, immediate and delayed memory, spatial processing, and decision processing speed and efficiency. Importantly, these qualities of the ANAM4 TBI MIL Battery are consistent with past applications of computer-based testing and TBI, with normative work conducted by DVBIC, and with the Clinical Practice Guidelines and Recommendations published by the Defense and Veterans Brain Injury Center Working Group on the Acute Management of Mild Traumatic Brain Injury in Military Operational Settings.

## 2.2 Outdated Technology

Despite these computerized versions of traditional paper-and-pencil neuropsychological tests, the vast majority of current neuropsychological assessment procedures represent a technology that has not changed since the first scales developed in the early 1900s (e.g., Binet and Simon's first scale in 1905 and Wechsler's in 1939). For the past few decades, the Wechsler scales (in various manifestations; e.g., WAIS-R, WAIS III) have been the most widely used neuropsychological tests [36, 37, 38]. While automated versions were developed of the original WAIS [39] in 1969 and again in 1980 [40], these automations provided only rudimentary stimulus presentation and limited data recording. Since the 1980s, the automated versions are all but abandoned and now the focus is upon slight revisions of the paper-and-pencil versions with computerized scoring. In fact, the latest revisions of the Wechsler scales (e.g., Wechsler Adult Intelligence Scale—Third Edition [41]; Wechsler Intelligence Scale for Children—Fourth Edition [42]) offer little more than cosmetic change and improved standardization. This lack of technological advancement of the Wechsler scales is important because according to a 2005 study surveying assessment practices and test usage patterns among 747 North American, doctorate-level clinical neuropsychologists, the Wechsler Scales were the most frequently used tests in their neuropsychological assessments [43].

Robert Sternberg pointed out over a decade ago the discrepancy between progress in cognitive assessment measures like the Wechsler scales and progress in other areas of technology [44]. Sternberg used the example of the now obsolete black and white televisions, vinyl records, rotary-dial telephones, and the first commercial computer made in the United States, UNIVAC I to illustrate the lack of technological progress in the standardized testing industry. According to Sternberg, currently used standardized tests differ little from tests that have been used throughout this century. For example, while the first edition of the Wechsler Adult Intelligence Scale appeared some years before UNIVAC, the Wechsler scales (and similar tests) have hardly changed at all (aside from primarily cosmetic changes) compared to computers. Although one may argue that innovation in the computer industry is different from innovation in the standardized testing industry, there are still appropriate comparisons. For example, whereas millions of dollars spent on technology in the computer industry typically

reflects increased processing speed and power; millions of dollars spent on innovation in the testing industry tends to reflect the move from multiple choice items to fill-in-the-blank items. Sternberg's statements are as true now as they were over a decade ago. While neuropsychology emphasizes its role as a science, its technology is not progressing in pace with other clinical neurosciences. Sternberg also points out neurocognitive testing needs progress in ideas, not just new measures, for delivering old technologies.

### **3 Ecologically Valid Neuropsychological Instruments**

The historical development of clinical neuropsychology reveals that while the initial purpose of neuropsychological assessment was diagnosing persons with brain injury or disease, and then describing brain-behavior relationships, today clinical neuropsychologists are increasingly being asked to make prescriptive statements about everyday functioning [45]. This new role for neuropsychologists has resulted in increased emphasis upon the ecological validity of neuropsychological instruments. To establish ecological validity of neuropsychological measures, neuropsychologists focus on demonstrations of either (or both) verisimilitude and veridicality [46]. By verisimilitude, ecological validity researchers are emphasizing the need for the data collection method to be similar to real life tasks in an open environment. For the neuropsychological measure to demonstrate veridicality, the test results should reflect and predict real world phenomena [47, 48, 49].

#### **3.1 Ecological Validity: Need to Incorporate Advanced Technology**

In addition to the controversy related to whether or not current indices found on commonly used paper-and-pencil neuropsychological tests give us sufficient detail for prediction of the potential everyday difficulties likely to be faced by patients [50], a dearth of research has addressed the degree to which neuropsychological testing is ecologically valid [51]. There is also the issue of how should neuropsychologists go about improving the ecological validity of neuropsychological tests. Some neuropsychologists prefer a veridicality approach, in which results gleaned from neuropsychological measures are combined with behavioral observations, rating scales, and self-report measures (e.g., Everyday Memory Questionnaire [52]). A problem for this approach, however, is that while rating scales tend to have satisfactory reliability, they have weak correlations with performance measures (i.e., relatively low validity) [53]. Other neuropsychologists take the verisimilitude approach, in which completely new measures are developed that more closely approximate everyday activities and behaviors (e.g., Rivermead Behavioral Memory Test [54]; Behavioral Assessment of the Dysexecutive Syndrome [55]; and the Test of Everyday Attention [56]). Review of the ecological validity of neuropsychological tests has provided support for the superiority of verisimilitude tests as the results from these measures tended to be more consistently related to the outcome measures than the traditional paper-and-pencil tests. However, a problem for the verisimilitude approach is that these instruments do not appear to be migrating from research laboratories into the applied settings of clinical neuropsychologists [57]. An additional problem for this approach is that

although these neuropsychologists have developed instruments that more closely approximate skills required for everyday functioning, have not made use of advances in computer technology. As a result, they are in danger of continuing the negative trend that deemphasizes psychology's role as a science. As Sternberg has contended, neurocognitive testing needs progress in ideas, not just new measures, for delivering old technologies.

#### **4 Virtual Environments Offer Advanced Ecological Validity**

While standard neuropsychological measures have been found to have adequate predictive value, their ecological validity may diminish predictions about real world functioning. Traditional neurocognitive measures may not replicate the diverse environment in which persons live. Additionally, standard neurocognitive batteries tend to examine isolated components of neuropsychological ability, which may not accurately reflect distinct cognitive domains [58, 59] Although today's neuropsychological assessment procedures are widely used, neuropsychologists have been slow to adjust to the impact of technology on their profession. While there are some computer-based neuropsychological assessments that offer a number of advantages over traditional paper-and-pencil testing (e.g., increased standardization of administration; increased accuracy of timing presentation and response latencies; ease of administration and data collection; and reliable and randomized presentation of stimuli for repeat administrations), the ecological validity of these computer-based neuropsychological measures is less emphasized. As mentioned above, only a handful of neuropsychological measures have been developed with the specific intention of tapping into everyday behaviors like navigating one's community, grocery shopping, and other activities of daily living. Of those that have been developed, even fewer make use of advances in computer technology.

Virtual reality (VR) is as an advanced computer interface that allows humans to become immersed within a computer-generated simulation. Potential VR use in assessment and rehabilitation of human cognitive processes is becoming recognized as technology advances. Since VEs allow for precise presentation and control of dynamic perceptual stimuli (visual, auditory, olfactory, gustatory, ambulatory, and haptic conditions), they can provide ecologically valid assessments that combine the veridical control and rigor of laboratory measures with a verisimilitude that reflects real life situations. Additionally, the enhanced computation power allows for a range of the accurate recording of neurobehavioral responses in a perceptual environmental that systematically presents complex stimuli. Such simulation technology appears to be distinctively suited for the development of ecologically valid environments, in which three-dimensional objects are presented in a consistent and precise manner. As a result, subjects are able to manipulate three dimensional objects in a virtual world that proffers a range of potential task demands.

Virtual reality applications that focus on component cognitive processes, including attention processes [60, 61, 62], spatial abilities [63, 64, 65] memory [66], and executive functions [67], are now being developed and tested. The increased ecological validity of neurocognitive batteries that include assessment using virtual scenarios may aid differential diagnosis and treatment planning. Within a VE, it is possible to

systematically present cognitive tasks targeting neuropsychological performance beyond what are currently available using traditional methods [68]. Reliability of neuropsychological assessment can be enhanced in VEs by better control of the perceptual environment, more consistent stimulus presentation, and more precise and accurate scoring [69]. Virtual Environments may also improve the validity of neurocognitive measurements via the increased quantification of discrete behavioral responses, allowing for the identification of more specific cognitive domains [70]. Virtual environments could allow for neurocognition to be tested in situations that are more ecologically valid. Participants can be evaluated in an environment that simulates the real world, not a contrived testing environment [71].

According to Rizzo et al [72], the application of VR to neuropsychological assessment is distinctively important because it represents the potential for more than a simple linear extension of existing computer technology for human use. For Rizzo, it is important that VR does more than simply automate the paradigms of the past. Instead, VEs provide a paradigm shift for the future. Here, Rizzo is reflecting Neisser's [73] contention that the findings from many traditional cognitive assessments have not been demonstrated to generalize beyond the narrow laboratory context. However, there is an essential tension between persons striving for ecological validity and persons interested in maintaining experimental control. For example, Banaji and Crowder [74] have contended that the ecological approach to neurocognitive research is inconsequential and that scientific progress necessitates greater emphasis on experimental control. This seems to hold especially true for much of the work that has been done in virtual and augmented reality because the focus of ecological validity tends to be upon verisimilitude and not veridicality. As Banaji and Crowder have challenged, if neurocognitive measures fail to establish internal validity, then one can conclude nothing from study findings. Likewise, if VR-based neuropsychological assessments do not take seriously the importance of veridicality, we have attractive simulations (i.e., verisimilitude), but do not have an ability to reliably and validly predict a person's performance on real world activities (i.e., veridicality).

There are a number of neuropsychology researchers would agree with Neisser that there are legitimate concerns about the verisimilitude (or ecological validity) of neuropsychological assessments. However, while the issue of ecological validity has been discussed in the literature, little has been done to remedy this situation. Instead, there are attempts to simply enhance the external validity of neuropsychological assessments. The concepts of external and ecological validity are related but not interchangeable. External validity involves the extent to which findings from research studies can be generalized across a variety of persons, times, and settings as well as to generalizations to specific persons, times, and settings. Given that traditional paper-and-pencil neuropsychological measures were developed for localization and the focus was upon double dissociation, enhancements tend to reflect endeavours to increase external validity. Hence, they do not typically require experimental conditions to mirror real life conditions. Neuropsychological measures are quite basic in their presentation and do not appear concerned with the level of verisimilitude found in VEs. Instead, they strive to be externally valid—to be consistently predictive of behavior exhibited in the real world.

As mentioned above, though, Banaji and Crowder have contended that the ecological approach to neurocognitive research is insignificant and that scientific progress

necessitates greater emphasis on experimental control. Unfortunately, much VR research supports this dichotomy. While verisimilitude is a major emphasis in reported studies using VR for psychology and neuropsychology, much less emphasis is placed upon veridicality—reliability, validity, and psychometric properties. In a recent meta-analysis of VR studies, Parsons and Rizzo [75] sought to examine the magnitude of changes in affective functioning that occurred following virtual reality exposure therapy (VRET). Although the results of the meta-analysis revealed that VRET had statistically large effects across affective domains, findings must be interpreted with caution given the inconsistencies in the research designs across studies. Many of the VRET studies did not include control groups, and many were not randomized clinical trials. As a result, the authors had diminished confidence that affective enhancements were directly related to or caused by VRET. Additionally, even though Parsons and Rizzo attempted to identify possible moderators of affective improvements, this was not possible because necessary information was either not reported or on occasions where it was reported it was done so in insufficient detail. This lack of information related to affective improvements and presence, immersion, anxiety and/or phobia duration, demographics (e.g. age, gender, and ethnicity) may reflect a limited range of values given the selection criteria employed by most studies. Thus, the findings of this meta-analysis may not generalize to patients with anxiety disorders in general. Similarly, a host of other factors that could not be directly analyzed might moderate affective regulation, including differences among treatment centers in terms of beliefs about best practices concerning VRET, timing of sessions, and concurrent psychopharmacological treatment.

It could be argued that the challenge for neuropsychologists using VEs is to develop techniques that simultaneously satisfy the demands of internal validity, external validity, and ecological validity. Hence, the development of an ecologically valid VE should include psychometric rigor (i.e. internal validity, external validity) as well as verisimilitude and veridicality (i.e. ecological validity). Achieving such standards requires consideration of a number of issues: 1) Correspondence: the tasks performed within VEs should correspond to the pertinent aspects of real world activities and environments; 2) Representativeness: the tasks developed should be representative of persons who are performing the tasks; 3) Expedience: research problems should have practical consequences on real world functioning if they are to be components of verisimilitude and veridicality; and 4) Relevance: outcome measures need to have relevance to the practical problem being investigated.

#### **4.1 Correspondence**

The tasks performed within VEs should correspond to the pertinent aspects of real-world activities and environments. This is reflected in attempts by neuropsychologists to establish the ecological validity of their measures through demonstrations of verisimilitude. Within a VE, it is possible to systematically present cognitive tasks targeting different domains of performance beyond what are currently available using traditional methods. Verisimilitude of cognitive assessment can be enhanced in VR by better control of the perceptual environment, more consistent stimulus presentation, and more precise and accurate scoring. The validity of VR-based neuropsychological measurements may be improved via the quantification of more discrete behavioral responses,

allowing for the identification of more specific cognitive domains. It may also be possible that VR-based neuropsychological assessments will allow for neurocognition to be tested in situations that are more ecologically valid. Participants can be evaluated in an environment that simulates the real world, not a contrived testing environment. Neuropsychological evaluations that include VR technology will bring verisimilitude—the quality or state of appearing to be true—to synthetic experiences. Verisimilitude emphasizes the need for the data collection method to be similar to real life tasks in an open environment. Neuropsychologists interested in such VR-based measures will design immersive experiences that proffer compelling interactive narratives that simulate the real world.

#### **4.2 Representativeness**

The tasks developed should be representative of persons who are performing the tasks. In addition to consideration of tasks and outcome measures, it is also important to consider subject populations. Specifically, it is important to ensure that these populations are representative of persons performing the tasks. A crucial aspect of quality VR-based neuropsychological studies is that authors adequately report the sampling strategy used so that readers may assess whether the sample reported upon is representative of the target population. This aspect of study design is critical because poor sampling will undermine the generalizability of the study and/or reduce the validity in situations where sampling bias is introduced. For study findings to be generalizable to the population as a whole, the sample must be representative of the population from which it is drawn. As such, the ecologically valid assessment should be culturally sensitive and have a design that reflects appropriate language—language-appropriate interventions that emphasize greater familiarity with cultural knowledge. Also, the visually mediated environments found within virtual reality emphasize the need for VR researchers to develop environments that make use of symbols and concepts shared by the population in question. This is an important issue for VR researchers in a multicultural society. Therefore it is important that neuropsychological assessments adequately handle cultural information (e.g. values, customs, and traditions). In addition to the development of future virtual simulations, there may also be need that existing treatment manuals be adapted to incorporate cultural values and validate the uniqueness of the particular ethnic group [76] While typical neuropsychological tests make efforts at such generalizability, an ecologically valid assessment will align the neurocognitive assessment with the target population's or the assessed person's socioculture milieu.

#### **4.3 Expedience**

The research problems should have practical consequences on real world functioning if they are to be components of veridicality. For the neuropsychological measure to demonstrate veridicality, the test results should reflect and predict real world phenomena. Research problems should be expedient instruments whose worth is measured by something more than the mere verisimilitude standard, in which the concepts and theories somehow mirror reality. For veridicality, the standard of measurement is

more an issue of pragmatic instrumentalism, in which significance is measured by how effective VR-based neuropsychological assessment results are at explaining and predicting future neurocognitive performance and activities of daily living. A common methodological procedure for establishing the veridicality might be to ensure the construct validity of virtual reality-based neuropsychological measures using the multitrait-multimethod matrix, in which a given construct is measured by multiple methods to develop a method-by-measure matrix [77]. The multitrait-multimethod matrix offers a method for assessment of the convergent and divergent validity of a measure by examining its pattern of correlations with other measures. In a VR-based neuropsychological study, for example, all the included scales might measure the construct of memory, but vary either on the level of memory impairment or the aspect of memory being assessed. Convergent validity coefficients (assessing memory domain) derived from the VR-based memory score and the traditional neuropsychological measures of memory could be assessed to see if they are significantly (statistical significance) larger than correlations of different measures assessing domains other than memory within the same array of measures. Evidence for discriminant validity would be indicated when correlations of different scales assessed using different measures were lower than the convergent validity coefficients [78].

#### **4.4 Relevance**

A condition essential for obtaining ecologically valid answers to neurocognitive assessment questions regarding real world functioning, is the inclusion of criterion tasks that include significant aspects of real world tasks and environments. While computerized assessments are helpful in examining performance on laboratory tasks, these measures are insufficient when attempting to obtain resolutions to practical problems. It is important to note that current neuropsychological tests (paper-and-pencil as well as computerized tests) do help provide insight into the source or nature of individual differences in performance. However, exclusive reliance on laboratory measures is rarely sufficient to gain insight into the types of problems neuropsychologists encounter when receiving a referral for a neuropsychological assessment. Hence, for VR-based neuropsychological measures to meet the ecological issue of relevance there is need for neuropsychologists to understand that experience includes both particulars and relations between those particulars. Therefore both the particulars of the neurocognitive domains and the relations between these domains that result from interaction within the real world deserve a place in our explanations. Any VR-based neuropsychological measure is flawed if it stops at assessing a neurocognitive domain and fails to explain how meaning, values, intentionality, and activities of daily living can arise from that.

In general, ecologically valid VR research should focus on real world practical problems such as issues related to return to work, driving competence, self-care, education and training, interface design, and treatment. A valid research question might look at the ways in which individual differences (within a given clinical population) impact VR-based neuropsychological assessments for determining the patient's ability to perform activities of daily living.

## 5 Conclusions

Much of this manuscript has dealt with two essential limitations that have resulted from neuropsychology's refusal of technological adaptation: First, current neuropsychological assessment procedures represent a technology that has barely changed since the first scales were developed in the early 1900s. Second, while the historical purpose of clinical neuropsychology was differential diagnosis of brain pathology, technological advances in other clinical neurosciences have changed the neuropsychologist's role to that of making ecologically valid predictions about the impact of a given patient's neurocognitive abilities and disabilities on everyday functioning.

Within this manuscript, virtual reality-based neuropsychological assessments have been presented as technologically advanced and ecologically valid neuropsychological instruments that may be used to address real world outcomes. A primary focus of this manuscript has been the explication of the necessary characteristics for ecologically valid VR assessment of neuropsychological functioning. To design such ecologically valid VR tools one must identify representative real world tasks. Following the identification of representative tasks, the characteristics of these tasks must be defined and a set of goal-directed activities established. Next, a virtual scenario that represents the task may be developed so that neuropsychological performance may be investigated within ecologically valid research environments. Neuropsychologists will need to subject these VEs to techniques that simultaneously satisfy the demands of internal validity, external validity, and ecological validity. Again the achievement of such standards requires insurance that the tasks performed within VEs correspond to the pertinent aspects of real world activities and environments. The tasks developed should be representative of persons who are performing the tasks. The research problems should have practical consequences on real world functioning if they are to be components of verisimilitude and veridicality. Finally, the outcome measures need to have relevance to the practical problem being investigated.

In conclusion, the advanced computer interfaces found in virtual reality allow humans to become immersed within a computer-generated environment. Enhanced accuracy in recording, coding, and storing of a range of neurobehavioral responses elicited from complex stimuli will allow for improved data analytics and predictive validity. It is argued that virtual environments may be uniquely suited for assessment of daily activities, allowing for presentation of three-dimensional objects in a consistent and precise manner, which subjects can then manipulate depending on a range of task demands. The precise presentation and control of dynamic perceptual stimuli (visual, auditory, olfactory, gustatory, ambulatory, and haptic conditions) in the virtual environment allows neuropsychologists the opportunity to develop statistically and clinically significant tasks within a virtual world. It is believed that a future possible neuropsychological assessment battery will make use of virtual reality to combine the control and rigor of technologically advanced computerized laboratory measures, the psychometric rigor (veridicality) of traditional paper-and-pencil assessments, and approximate the verisimilitude of real life situations.

## References

1. Benton, A.L.: Some problems associated with neuropsychological assessment. *Bulletin of Clinical Neuroscience* 50, 11–15 (1985)
2. Broca, P.: Sur le siège de la faculté du langage articulé. *Bulletine de Société Anthropologique* 6, 337–339 (1865)
3. Wernicke, C.: *Der Aphasische Symptomkomplex*. Cohn & Weigart, Breslau (1874)
4. Binet, A., Henri, V.: La psychologie individuelle [Individual psychology]. *L'Annee Psychologique* 2, 411–465 (1896)
5. Binet, A., Henri, V.: La fatigue intellectuelle [Intellectual tiredness]. Schleicher, Paris (1898)
6. Binet, A., Simon, T.: Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [New methods to diagnose the intellectual level of abnormal]. *L'Annee Psychologique* 11, 245–366 (1905)
7. Binet, A., Simon, T.: Le développement de l'intelligence chez les enfants [The development of intelligence in children]. *L'Annee Psychologique* 14, 1–94 (1908)
8. Anastasi, A., Urbina, S.: *Psychological testing*, 7th edn. McMillan, New York (1997)
9. Yerkes, R.M.: Behaviorism and genetic psychology. *Journal of Philosophy, Psychology, and Scientific Methods* 14, 154–160 (1917)
10. Yoakum, C.S., Yerkes, R.M.: *Army mental tests*. Holt, New York (1920)
11. Conkey, R.: Psychological changes associated with head injuries. *Archives of Psychology* 33, 5–22 (1938)
12. Shipley, W.C.: A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology* 9, 371–377 (1940)
13. Benton, A.L., Howell, I.L.: The use of psychological tests in the evaluation of intellectual function following head injury. *Psychosomatic Medicine* 3, 138–151 (1941)
14. Goldstein, K., Scheerer, M.: Abstract and concrete behavior: an experimental study with special tests. *Psychological Monograph* 43, 1–151 (1941)
15. Wells, F.L., Ruesch, J.: *Mental Examiner's Handbook*. Psychol. Corp., New York (1942)
16. Hunt, H.F.: A practical clinical test for organic brain damage. *Journal of Applied Psychology* 27, 375–386 (1943)
17. Halstead, W.C.: *Brain and Intelligence*. Univ. Chicago Press, Chicago (1947)
18. Reitan, R.M.: Investigation of the validity of Halstead's measure of biological intelligence. *Archives of Neurology* 73, 28–35 (1955)
19. Lezak, M.D.: *Neuropsychological Assessment*, 3rd edn. Oxford University Press, New York (1995)
20. Spreen, O., Strauss, E.: *A compendium of neuropsychological tests: Administration, norms and commentary*, 2nd edn. Oxford University Press, New York (1998)
21. Mitrushina, M.N., Boone, K.L., D'Elia, L.: *Handbook of normative data for neuropsychological assessment*. Oxford University Press, New York (1999)
22. Heaton, R.K., Grant, I., Matthews, C.G.: *Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Psychological Assessment Resources, Odessa (1991)
23. Teuber, H.-L.: Physiological psychology. *Annual Review of Psychology* 6, 267–296 (1955)
24. Pribram, K.H.: *Languages of the brain*. Prentice-Hall, Englewood Cliffs (1971)
25. Ruff, R.M.: A friendly critique of neuropsychology: Facing the challenges of our future. *Archives of Clinical Neuropsychology* 18, 847–864 (2003)

26. Schatz, P., Browndyke, J.: Applications of computer-based neuropsychological assessment. *Journal of Head Trauma Rehabilitation* 17, 395–410 (2002)
27. Eckerman, D.A., Carroll, J.B., Forre, C.D., Guillen, M., Lansman, M.: An approach to brief field testing for neurotoxicity. *Neurobehavioral Toxicology & Teratology* 7, 387–393 (1985)
28. Waterfall, R.C.: Automating standard intelligence tests. *J. Autovisual. Media Med.* 2, 21–24 (1979)
29. Space, L.G.: A console for the interactive on-line administration of psychological tests. *Behav. Res. Meth. Instr.* 7, 191–193 (1975)
30. Beaumont, J.G.: The validity of the Category Test administered by on-line computer. *Journal of Clinical Psychology* 31, 458–462 (1975)
31. Heaton, R.: Wisconsin Card Sorting Test: Computer Version 3 for Windows. Psychological Assessment Resources, Lutz (1999)
32. CogSport. Parkville, Victoria, Australia: CogState, Ltd (1999)
33. Lovell, M.R., Collins, M.W., Podell, K., Powell, J., Maroon, J.: ImPACT: Immediate Post-Concussion Assessment and Cognitive Testing. NeuroHealth Systems, LLC, Pittsburgh (2000)
34. Johnson, D.R., Vincent, A.S., Johnson, A.E., Gilliland, K., Schlegel, R.E.: Reliability and construct validity of the Automated Neuropsychological Assessment Metrics (ANAM) mood scale. *Archives of Clinical Neuropsychology* 1, 73–85 (2008)
35. Erlanger, D.M., Feldman, D.J., Kutner, K.: Concussion Resolution Index. HeadMinder, Inc., New York (1999)
36. Hartlage, L.C., Telzrow, C.F.: The practice of clinical neuropsychology in the US. *Archives of Clinical Neuropsychology* 2, 200–202 (1980)
37. Guilmette, T.J., Faust, D., Hart, K., Arkes, H.R.: A national survey of psychologists who offer neuropsychological services. *Archives of Clinical Neuropsychology* 5, 373–392 (1990)
38. Lees-Haley, P.R., Smith, H.H., Williams, C.W., Dunn, J.T.: Forensic neuropsychological test usage: an empirical survey. *Archives of Clinical Neuropsychology* 11, 45–51 (1996)
39. Elwood, D.L., Griffin, R.: Individual intelligence testing without the examiner. *Journal of Consulting and Clinical Psychology* 38, 9–14 (1972)
40. Vincent, K.R.: Semi-automated full battery. *Journal of Clinical Psychology* 36, 437–446 (1980)
41. Wechsler, D.: Wechsler Adult Intelligence Scale, 3rd edn. WAIS-3. Harcourt Assessment, San Antonio (1997)
42. Wechsler, D.: Wechsler Intelligence Scale for Children—, 4th edn. WISC-IV. Harcourt Assessment, San Antonio (2003)
43. Rabin, L., Barr, W., Burton, L.: Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology* 20, 33–65 (2005)
44. Sternberg, R.J.: Intelligence and lifelong learning: What's new and how can we use it? *American Psychologist* 52, 1134–1139 (1997)
45. Long, C.J.: Neuropsychological tests: A look at our past and the impact that ecological issues may have on our future. In: Sbordone, R.J., Long, C.J. (eds.) *Ecological Validity of Neuropsychological Testing*, pp. 1–14. GR Press/St. Lucie Press, Delray Beach, FL (1996)
46. Franzen, M.D., Wilhelm, K.L.: Conceptual foundations of ecological validity in neuropsychological assessment. In: Sbordone, R.J., Long, C.J. (eds.) *Ecological validity of neuropsychological testing*, pp. 91–112. St Lucie Press, Boca Raton (1996)

47. Chaytor, N., Schmitter-Edgecombe, M.: The Ecological Validity of Neuropsychological Tests: A Review of the Literature on Everyday Cognitive Skills. *Neuropsychology Review* 13, 181–197 (2003)
48. Silver, C.H.: Ecological validity of neuropsychological assessment in childhood traumatic brain injury. *Journal of Head Trauma Rehabilitation* 15, 973–988 (2000)
49. Ready, R.E., Stierman, L., Paulsen, J.S.: Ecological Validity of Neuropsychological and Personality Measures of Executive Functions. *Clinical Neuropsychologist* 15, 314–323 (2001)
50. Wilson, B.A.: Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied and Preventive Psychology* 2, 209–215 (1993)
51. Nussbaum, P.D., Goreczny, A., Haddad, L.: Cognitive correlates of functional capacity in elderly depressed versus patients with probable Alzheimer's disease. *Neuropsychological Rehabilitation* 5, 333–340 (1995)
52. Sunderland, A., Harris, J.E., Baddeley, A.D.: Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior* 22, 341–357 (1983)
53. Larrabee, G.J., Crook, T.H.: The ecological validity of memory testing procedures: Developments in the assessment of everyday memory. In: Sbordone, R.J., Long, C.J. (eds.) *Ecological validity of neuropsychological testing*, pp. 225–242. GR Press/St. Lucie Press, Delray Beach, FL (1996)
54. Wilson, B.A., Cockburn, J., Baddeley, A.: *The Rivermead Behavioral Memory Test*. Thames Valley Test Co./Gaylord/National Rehabilitation Services, Reading/England MI (1985)
55. Wilson, B.A., Alderman, N., Burgess, P., Emslie, H., Evans, J.J.: *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*, Thames Valley Test Company. Bury St. Edmunds, England (1996)
56. Robertson, I.H., Ward, T., Ridgeway, V., Nimmo-Smith, I.: *The Test of Everyday Attention*. Thames Valley Test Co./Gaylord/National Rehabilitation Services, Suffolk/UK/MI (1994)
57. Rabin, L.A., Burton, L.A., Barr, W.B.: Utilization rates of ecologically oriented instruments among clinical neuropsychologists. *Clinical Neuropsychologist* 5, 727–743 (2007)
58. Parsons, T.D., Rizzo, A.A., van der Zaag, C., McGee, J.S., Buckwalter, J.G.: Gender and cognitive performance: a test of the common cause hypothesis. *Aging, Neuropsychology, & Cognition* 12, 78–88 (2005)
59. Parsons, T.D., Rizzo, A.A., Buckwalter, J.G.: Backpropagation and regression: comparative utility for neuropsychologists. *Journal of Clinical & Experimental Neuropsychology* 26, 95–104 (2004)
60. Parsons, T.D., Bowerly, T., Buckwalter, J.G., Rizzo, A.A.: A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology* 13, 363–381 (2007)
61. Parsons, T.D., Cosand, L., Courtney, C., Iyer, A., Rizzo, A.A.: *Neurocognitive Workload Assessment using the Virtual Reality Cognitive Performance Assessment Test*. LNCS (LNAI), vol. 5639, pp. 243–252 (2009)
62. Parsons, T.D., Rizzo, A.A.: *Neuropsychological Assessment of Attentional Processing using Virtual Reality*. *Annual Review of CyberTherapy and Telemedicine* 6, 23–28 (2008)

63. Parsons, T.D., Larson, P., Kratz, K., Thiebaut, M., Bluestein, B., Buckwalter, J.G.: Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia* 42, 555–562 (2004)
64. Wolbers, T., Weiller, C., Buchel, C.: Neural foundations of emerging route knowledge in complex spatial environments. *Brain Research Cognitive Brain Research* 21, 401–411 (2004)
65. Morganti, F., Gaggioli, A.: A virtual reality extended neuropsychological assessment for topographical disorientation: a feasibility study. *Journal of NeuroEngineering and Rehabilitation* 4(26) (2007)
66. Parsons, T.D., Rizzo, A.A.: Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test. *Cyberpsychology and Behavior* 11, 17–25 (2008)
67. Elkind, J.S., Rubin, E., Rosenthal, S., Skoff, B., Prather, P.: A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: an analysis of preliminary results. *CyberPsychology & Behavior* 4, 489–496 (2001)
68. Rizzo, A.A., Schultheis, M.T., Kerns, K., Mateer, C.: Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation* 14, 207–239 (2004)
69. Rizzo, A.A., Buckwalter, J.G.: Virtual reality and cognitive assessment and rehabilitation: the state of the art. In: Riva, G. (ed.) *Virtual reality in neuro-psycho-physiology*, pp. 123–146. IOS Press, Amsterdam (1997)
70. Rizzo, A.A., Kim, G.: A SWOT analysis of the field of Virtual Rehabilitation and Therapy. *Presence: Teleoperators and Virtual Environments* 14, 1–28 (2005)
71. Giuseppe, R., Fabrizia, M., Andrea, G.: Presence and rehabilitation: toward second-generation virtual reality applications in neuropsychology. *Journal of NeuroEngineering and Rehabilitation* 1(9) (2004)
72. Neisser, U.: Memory: What are the important questions? In: Gruneberg, M.M., Morris, P.E., Sykes, R.N. (eds.) *Practical Aspects of Memory*, pp. 3–24. Academic Press, San Diego (1978)
73. Banaji, M., Crowder, R.G.: The bankruptcy of everyday memory. *American Psychologist* 44, 1185–1193 (1989)
74. Cook, T.D., Campbell, D.T.: *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin, Boston (1977)
75. Parsons, T.D., Rizzo, A.A.: Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. *Journal of Behavior Therapy and Experimental Psychiatry* 39, 250–261 (2008)
76. Bernal, G., Bonilla, J., Bellido, C.: Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with Hispanics. *Journal of Abnormal Child Psychology* 23, 67–82 (1995)
77. Campbell, D.T., Fiske, D.W.: Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin* 56, 81–105 (1959)
78. Parsons, T.D., Rizzo, A.A.: Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test. *Cyberpsychology and Behavior* 11, 17–25 (2008)