ELSEVIER

Archives of Clinical Neuropsychology 21 (2006) 327–337

Archives of CLINICAL NEUROPSYCHOLOGY

Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations

Donna M. Spooner*, Nancy A. Pachana

School of Psychology, The University of Queensland, Brisbane, QLD 4072, Australia

Accepted 28 April 2006

Abstract

The focus of the discipline of neuropsychology is shifting towards a greater emphasis on understanding the relationship between assessment results and performance of everyday tasks (ecological validity). To date, the literature has highlighted the importance of this concept in the assessment of patients with brain injury or disease (e.g. in rehabilitation and forensic settings). This paper presents the argument that there is another important area in which the ecological validity of neuropsychological assessments should be considered: in clinical outcomes studies using neurologically intact participants. For example, determining the extent to which a medical procedure or intervention affects performance of everyday cognitive tasks can provide useful information that can potentially guide decision-making regarding treatment options. It is argued that tests designed with ecological validity in mind (the verisimilitude approach), as opposed to traditional tests, may be most effective at predicting everyday functioning. Explanations are proposed as to why researchers may be reluctant to use tests with verisimilitude in favor of more traditional measures.
© 2006 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Ecological validity; Neuropsychological assessment; Neuropsychological research

Ecological validity, the "functional and predictive relationship between the patient's performance on a set of neuropsychological tests and the patient's behavior in a variety of real world settings" (Sbordone, 1996, p. 16), received little attention in the neuropsychological literature until the late 1980s. It has since become recognized as a vital concept for our profession, clarifying our understanding and interpretation of test findings (Chaytor & Schmitter-Edgecombe, 2003; Long & Kibby, 1995; Troster, 2000), offering suggestions for improved training of neuropsychologists focused on functional applications of assessment findings (Johnstone & Farmer, 1997; Johnstone & Frank, 1995), and guiding the evolution of our profession.

1. The changing role of neuropsychology

Historically, neuropsychological assessments were regarded as an essential means of detecting and localizing neuropathology. Specifically, the neuropsychologist would draw educated inferences about the likely etiology of the underlying brain dysfunction based on the patient's cognitive strengths and weaknesses. However, the role of neu-

* Corresponding author. Tel.: +61 7 3636 1668; fax: +61 7 3636 2632.
E-mail address: spooner@psy.uq.edu.au (D.M. Spooner).

0887-6177/$ – see front matter © 2006 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.acn.2006.04.004

ropsychology in this regard is diminishing as sophisticated neuroimaging techniques become more widely available, enabling medical specialists to gather more definitive information about the location and extent of brain injury for diagnostic purposes (Johnstone & Frank, 1995). While neuropsychological tests continue to contribute to diagnosis in some conditions, such as detecting the early stages of Alzheimer's disease (e.g. Derrer et al., 2002), they often reflect the activity of multiple brain regions or 'neural systems', which may limit their precision at localization. The nature of referral questions has also evolved in recent years as neuropsychological services are accessed by sources other than medical specialists, including lawyers, employers, insurance companies, and schools (Chaytor & Schmitter-Edgecombe, 2003; Long & Kibby, 1995).

Following from these developments, a new emphasis on greater consideration of the functional implications of neuropsychological test results has emerged. Specifically, neuropsychologists are now being asked to draw conclusions from assessments regarding patients' abilities to competently perform such tasks as living independently or returning to a previous occupation. It has become increasingly recognized that neuropsychologists must embrace this new role to ensure that we continue to offer highly valuable contributions in issues related to patient care and decision-making (Heinrichs, 1990; Johnstone & Farmer, 1997; Troster, 2000).

2. Approaches to ecological validity

Unfortunately, the conceptualization behind test design and, in many instances, the reasoning behind the neuropsychologist's choice of tests administered, has not reflected the shifting focus of neuropsychological assessment. Many older tests face limitations in utility or efficacy; often the same tests that were used to aid neuropathological diagnosis are now used to make predictions about the real-world functioning of individuals. Typically, the predictions made are based on the face validity of the tests and the past experience of the neuropsychologist in drawing inferences between the results and the patient's everyday life. This is of concern, as only limited research exists on the ecological validity of neuropsychological tests.

Two approaches appear to be most prevalent in establishing ecological validity: verisimilitude and veridicality (Franzen & Wilhelm, 1996). Verisimilitude refers to the similarity between the task demands of the test and the demands imposed in the everyday environment. Establishing verisimilitude requires the development of tests comprised of everyday cognitive tasks, such that inferences can be easily drawn from test results and the individual's likely ability to perform those tasks in daily life. Some progress has been made towards developing standardized instruments with adequate verisimilitude. The Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994), the Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996), the Rivermead Behavioral Memory Test (RBMT; Wilson, Cockburn, & Baddeley, 1985), and more recently, the Cambridge Test of Prospective Memory (CAMPROMPT; Wilson et al., 2004), were developed with this theoretical concept in mind to assess everyday skills in attention, executive functioning, memory, and prospective memory, respectively. These tests differ from traditional assessment instruments in their focus on identifying individuals with limited functional abilities rather than discriminating brain-injured from healthy people or determining the etiology of the brain dysfunction (Chaytor & Schmitter-Edgecombe, 2003).

In contrast, veridicality refers to the extent to which results on an assessment instrument are related to scores on other measures that predict the performance of real-world tasks (Franzen & Wilhelm, 1996). To determine veridicality, the relationship between performance on traditional assessment instruments and measures of everyday functioning (e.g. employment status, clinician's ratings, behavioral observations) is assessed using statistical methods. An example of this approach is to examine the relationship between results on the Wechsler Memory Scales—Third Edition (WMS-III; The Psychological Corporation, 1997), a traditional memory test, and scores on a specific functional outcome measure, such as clinician's ratings of memory failures. This approach has been subject of some criticism as direct parallels between the demands of traditional tests and functional performance are often not always evident (e.g. Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999; Sunderland, Harris, & Baddeley, 1983; Wilson, 1993; Wilson, Cockburn, Baddeley, & Hiorns, 1989).

Recently, Chaytor and Schmitter-Edgecombe (2003) investigated the effectiveness of these two approaches in predicting everyday cognitive abilities. They reviewed empirical studies utilizing tests with verisimilitude, veridicality, or a combination of both approaches, to investigate everyday cognitive skills across the domains of general neuropsychological functioning (encapsulating a wide range of neuropsychological tests), memory, and executive functioning. One study also assessed the utility of the two approaches in predicting everyday skills in the domain of attention

(Higginson, Arnett, & Voss, 2000). The studies reviewed included a variety of clinical and neurological samples, including schizophrenia (e.g. Evans, Chua, McKenna, & Wilson, 1997), closed head injury (e.g. Kaitaro, Koskinen, & Kaipio, 1995), Alzheimer's disease (Johnson, 1994), multiple sclerosis (Higginson et al., 2000), frontal lobe damage (Dimitrov, Grafman, & Hollnagel, 1996), and stroke (Nadolne & Stringer, 2001). Some of the studies also included control groups comprised of normal, healthy participants (e.g. Dimitrov et al., 1996; Evans et al., 1997).

For each study reviewed, Chaytor and Schmitter-Edgecombe (2003) determined how effective the chosen approach was by examining how precisely the test findings related to the functional outcome measure(s) used in the study. For the domain of general neuropsychological functioning, they were unable to draw firm conclusions regarding the differential effectiveness of the two approaches due to poor specificity of the hypotheses in each of the studies reviewed. Comparison of the two approaches across the studies of memory and executive functioning was slightly obscured by the use of different outcome measures (e.g. clinicians ratings, self and informant questionnaires) as well as specific population effects (e.g. the finding of significant results in particular patient groups but not in others). However, for these two cognitive domains, Chaytor and Schmitter-Edgecombe (2003) concluded that there was some support for the superiority of the verisimilitude approach, as there was a tendency for the results from these tests to be more highly and consistently related to everyday cognitive performance than their traditional counterparts. In the single study examining attention, a test with verisimilitude was found to be a better predictor of general functional outcome than traditional tests of attention in patients with multiple sclerosis (Higginson et al., 2000). These findings suggest that the continued development of tests designed with ecological validity in mind is a worthwhile pursuit, and that use of such tests is important when an understanding of the individual's everyday cognitive functioning is required.

3. Consideration of ecological validity in the assessment of patients with brain impairment

Commentaries on ecological validity to date have primarily emphasized the increased consideration of this concept in assessments of neurologically impaired individuals, particularly in rehabilitative and forensic contexts. In rehabilitation settings, the primary goal is treatment planning rather than determining the type and location of cerebral abnormalities. To ensure appropriate treatment interventions are designed, the primary role of the neuropsychologist is to discuss the likely implications of the assessment findings on the patient's ability to carry out daily activities (Bennett, 2001; Wilson, 1993). This information is then used to guide the neuropsychologist and other health professionals (e.g. occupational therapists) on the type of rehabilitation required and the degree of recovery that can reasonably be expected.

In the forensic domain, also, the neuropsychologist must describe the functional implications of patients injuries when providing courtroom testimony. In this environment, it is clearly insufficient to merely discuss the patient's cognitive strengths and weaknesses. Rather, in order to make decisions regarding compensation, the court requires provision of information about how changes in the patient's cognitive, emotional, vocational, and social abilities affect his or her capacity to cope with the demands of everyday life (Long & Collins, 1997; Long & Kibby, 1995).

4. Ecological validity and research with neurologically intact samples

Given the changing focus of neuropsychology, the greater emphasis on the importance of ecologically based assessments in rehabilitation and forensic settings is promising. As this awareness continues to grow, it is hoped that neuropsychologists will continue to offer increased consideration to this concept in these areas of practice. However, there is a further area in which ecological validity has not been carefully considered: in clinical outcomes research studies using neurologically intact participants. In the body of literature promoting increased awareness of ecological validity, the relevance of the concept to populations other than those with some form of neuropathology has received little attention. It is our view, therefore, that neuropsychologists working in research are neglecting a major area in which consideration of this concept would also have important implications. Neuropsychologists are not engaged merely in clinical practice focused on assessment and rehabilitation of neurologically impaired patients. Rather, neuropsychologists spend a considerable proportion of their time engaged in research work (Sweet, Moberg, & Suchy, 2000; Sweet, Peck, Abramowitz, & Etzweiler, 2002), including examination of cognitive functioning in neurologically intact samples. Examples of such studies include those investigating the cognitive effects of sleep deprivation (e.g. Taylor & McFatter, 2003), neuropsychological performance following acute exposure to solvent-based paints (e.g. Morrow, Steinhauer, Condray, & Hodgson, 1997) and the effects of changes in blood glucose levels on cognitive performance (e.g. Donohoe & Benton, 1999). However, the most common type of studies utilizing neurologically intact samples

are outcome studies that seek to elucidate cognitive functioning following a clinical intervention, such as a medical procedure or administration of a particular pharmaceutical agent. Such studies include those that have investigated the effects of estrogen replacement therapy on memory performance in postmenopausal women (e.g. Polo-Kantola et al., 1998), cognitive performance following dialysis treatment in patients with renal disease (e.g. Griva et al., 2003), the effects of adjuvant chemotherapy on cognition in women with breast cancer (e.g. Brezden, Phillips, Abdolell, Bunston, & Tannock, 2000), and cognitive outcomes following administration of general anaesthetic (e.g. Johnson et al., 2002). These studies consistently measure cognitive functioning via administration of traditional neuropsychological tests or, more infrequently, experimental measures of cognitive functioning. These tests may be selected for their purported potential to highlight specific cognitive domains that may be affected by the particular intervention, and the availability, in many cases, of good normative data. However, often they are chosen simply because they were used in the past, with little consideration of the fact that they were developed primarily to fulfill a role in an era that emphasized diagnosis. This is a concern in clinical outcomes research, as these tests do not provide information about the degree of functional difficulty that may be experienced, thereby limiting effective evaluation of the effects of specific interventions on performance of everyday tasks.

Some researchers have expressed concern about the poor functional relevance of the findings of clinical outcome studies. In a comprehensive review of the amnesic effects of benzodiazepines, Barbee (1993) questioned the value of laboratory-based memory tasks when attempting to understand the impact of memory deficits in daily life, warning that such tasks have low relevance to everyday functioning. Also, in a recent paper suggesting improvements to the design of studies examining the cognitive effects of chemotherapy in patients with breast cancer, Tannock, Ahles, Ganz, and van Dam (2004) commented that many neuropsychological tests do not have ecological validity and recommended use of measures that "give insight into the problems that cognitive impairment causes in daily life" (p. 2235). Furthermore, it has been stated that tests demonstrating a clear relationship to meaningful change in everyday performance will receive greater acceptance as a means of documenting the efficacy of pharmaceutical interventions (Larabee & Crook 1996; Leber, 1986). Despite such comments, little has been done to improve our understanding of the effects of specific clinical interventions on everyday cognitive functioning of neurologically intact individuals. This is of concern given that an emphasis on evidence-based medicine has recently emerged and neuropsychological outcome studies are being increasingly used as the basis on which to evaluate the efficacy of medical interventions (Chelune, 2002). Furthermore, patients are demanding more comprehensive information about the possible side effects of pharmaceutical agents and other medical management procedures when making treatment decisions (Ford, Schofield, & Hope, 2003). Moreover, it is disappointing, given our increased awareness of ecological validity in other areas of neuropsychological practice.

5. The verisimilitude approach in clinical intervention studies using neurologically intact samples: a focus on memory assessment

The finding that tests with verisimilitude may be better predictors of everyday cognitive skills than traditional measures (Chaytor & Schmitter-Edgecombe, 2003) has important implications for the process of neuropsychological test selection. Although the researchers reached this conclusion somewhat tentatively, it provides some guidance to neuropsychologists in the absence of further empirical evidence. It implies that a different approach to the selection of tests may be required, not only in clinical settings, but also in research contexts, including studies using neurologically intact samples. In conducting clinical practice with individual neurologically impaired patients, ecological validity can be enhanced by consideration of behavioral observations and clinician ratings, in addition to test results. However, when conducting research with neurologically intact participants, it is not always possible or practical to obtain information on participants' everyday functioning aside from that gained via formal cognitive assessment. For this reason, selecting an assessment tool with demonstrated ecological validity is of particular importance. To highlight the use of the verisimilitude approach in neuropsychological research, we use the area of memory assessment as an example, since careful assessment of this domain is integral to understanding the overall functioning of an individual, and memory difficulties can affect a broad range of other cognitive abilities, including learning, reasoning, planning, and attention (Makatura et al., 1999).

The assessment of functional memory performance or "everyday memory" provoked much controversy in the past, particularly with regards to whether memory should be measured through standardized, laboratory-based methods or via more observational and naturalistic assessment practices (e.g. Banaji & Crowder, 1989; Conway, 1991; Neisser, 1978). However, the debate has since subsided (deWall, Wilson, & Baddeley, 1994) and tests with verisimilitude

highlight the successful merging of these approaches by being both standardized and ecologically relevant. Only a few memory tests have been developed with the concept of verisimilitude in mind. These tests focus on the practical aspects of memory by reflecting the types of situations that individuals commonly encounter on an everyday basis. They differ from traditional tests such as the WMS-III or memory component of the Luria–Nebraska Neuropsychological Battery in that they do not require the individual to remember clinical or experimental material, such as verbal paired associates or geometric figures.

The most widely used memory test developed with verisimilitude, and arguably the test with the greatest amount of empirical support, is the Rivermead Behavioral Memory Test (Wilson et al., 1985). This standardized test was developed for use with patients with brain-injury to assist decision-making regarding the individual's functional abilities, such as his or her capacity to leave hospital or return to a previous occupation. It consists of 12 subtests, each of which are analogues of everyday memory tasks, such as remembering names associated with faces, recalling a short story, remembering a route traced around the room, remembering to ask for a hidden belonging and recalling its location, remembering to deliver a message, and recognizing a series of faces. Studies employing the RBMT have revealed that it is a very effective tool for predicting everyday memory impairments in patients with head-injury (Makatura et al., 1999; Schwatrz & McMillan, 1989; Wilson et al., 1989), and patients who have suffered coma or post-traumatic amnesia (Geffen, Encel, & Forrestor, 1991). Furthermore, it has been found to be superior to conventional memory tests (the WMS and memory tests of the Luria–Nebraska Neuropsychological Battery) in predicting functional memory abilities in patients with Alzheimer-type dementia (Goldstein, McCue, Rogers, & Nussbaum, 1992).

However, the RBMT is not sufficiently sensitive to identify memory deficits in normal, healthy individuals, making it unsuitable for use in clinical outcomes studies with neurologically intact participants. Recognizing this limitation, deWall et al. (1994) developed a more demanding version by combining two parallel forms of the RBMT, later named the Rivermead Behavioral Memory Test—Extended Version (RBMT-E; Wilson et al., 1999). The purpose of doubling the amount of material to remember was to eliminate ceiling effects in the general population to enable increased sensitivity to mild memory problems. The RBMT-E was piloted with healthy middle-aged and elderly people, and was found to have sufficient sensitivity to detect the small differences in memory performance between these two groups that result from the effects of increasing age on memory (deWall et al., 1994). Two parallel versions of the test were later developed and normative data were constructed across three IQ and three age bands (Wilson et al., 1999).

The test developers recognized the potential value of the RBMT-E in research using neurologically intact participants, particularly to assess mild and/or transient everyday memory deficits following administration of a particular medication or imposition of a specific stressor. While they acknowledged that traditional, laboratory-based memory tests may be sensitive in determining whether a new medication has any effect on memory, they argued that "such standard measures are much less helpful in helping to decide whether any deficits observed are sufficiently marked as to outweigh any benefits obtained" (deWall et al., 1994, p. 151). The RBMT-E has also been found to differentiate between patients with brain-injury who achieve scores within the normal range on the RBMT (Wills, Clare, Shiel, & Wilson, 2000), lending further support to the test as a useful tool for detecting subtle everyday memory impairment.

Despite the RBMT-E being developed primarily to measure the everyday memory performance of neurologically intact individuals for research purposes, and its strong potential in fulfilling this role, it has been underutilized in the literature. To our knowledge, only one study to date has used the test for this purpose. Specifically, Stephens, Hamilton, and Pachana (2003) used the RBMT-E to assess everyday memory functioning in a New Zealand sample of 104 mid-life women and determine there were differences in performance between users of hormone therapy (HT) and non-users. The selection of the RMBT-E was particularly appropriate given that practical memory difficulties represent a very typical complaint of menopausal women (Warga, 1999). The results showed that HT-users performed significantly better on three of the RBMT-E subtests (immediate and delayed story recall, and a delayed recall task requiring the ability to remember to deliver a message) and the RBMT-E overall measure of everyday memory functioning (Stephens et al., 2003). In addition to providing a useful insight into the possible effects of HT on memory, the findings have important implications for mid-life women making decisions regarding management of menopausal symptoms, allowing consideration of the potential implications of HT use on everyday memory functioning in conjunction with other known risks and benefits of HT. Inclusion of the RBMT-E in research protocols following administration of other treatments or interventions would provide similarly ecologically relevant information. The data collected could then be used to guide decision-making regarding therapeutic options.

While memory has been selected as the focus here, consideration of the ecologically validity of instruments used in research studies also applies to the assessment of other cognitive domains. Two established tests developed with

verisimilitude to assess attention and executive functioning (the TEA and the BADS, respectively) have been well validated in groups of brain-impaired patients (Robertson et al., 1994; Wilson et al., 1996). The TEA consists of a series of ecologically relevant tasks such as searching maps for specific information, looking through telephone directories, and listening to lottery number broadcasts (Robertson et al., 1994). It has a high level of sensitivity, being able to discriminate between patients with minimal and mild Alzheimer's disease (Green, Hodges, & Baddeley, 1995). Furthermore, specific subtests are able to detect the presence of normal age effects in attentional skills in the normal population (Robertson et al., 1994) and it is purported to have a wide application of use, ranging from patients with early Alzheimer's disease to young, healthy individuals (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996). Thus, the TEA appears to have strong utility to detect subtle differences in the attentional abilities of neurologically intact individuals participating in clinical outcomes studies.

Similar to the TEA, the BADS also consists of tasks that are similar to everyday activities, such as developing a strategy to search for lost keys, organizing the completion of multiple tasks while adhering to specific rules, and developing a plan for visiting a series of locations at a zoo while again following stated rules (Wilson et al., 1996). Wilson, Evans, Emslie, Alderman, and Burgess (1998) describe the test as having potential utility in detecting subtle difficulties in planning and organization, particularly in individuals who appear to be cognitively well-preserved. However, it was primarily designed to assess executive disturbances in patients with brain-impairment. Thus, further investigation into the sensitivity of the test with neurologically intact individuals is required before it can be considered for use in research with such samples.

The development of tests with verisimilitude to assess other cognitive domains that are currently assessed using traditional tests, such as visuospatial processing and speed of information processing, may also prove to be highly useful. Deficits in these domains can have profound effects on the ability to competently perform everyday behaviors. At this stage, it is not known how effectively traditional tests of these domains predict the performance of everyday tasks requiring these skills, that is, whether the veridicality approach is adequate. If the construction of tests with verisimilitude is deemed necessary, the test developers should examine whether the instrument is sufficient to provide a broad range of scores reflective of the performance of normal, healthy participants. More difficult and sensitive versions of the test should be developed, where relevant, thereby making it suitable for use in research with neurologically intact samples.

6. A reluctance to move beyond traditional tests

Although we believe that the need to move beyond the sole use of traditional instruments in clinical outcomes studies with neurologically intact samples is clear, the challenge resides in convincing researchers to become more open to the increased use and development of tests which are demonstrated to be most ecologically valid. Like developments in other health sciences, changes in neuropsychological theories and views on assessment-based issues can take some time to gain acceptance and continue forth to influence research and practice. Generally, neuropsychologists have historically been cautious about adopting new methods and shifting perspective on the way in which they conduct aspects of their profession (Williams, 1988). Although neuropsychologists involved in work with brain-impaired patients are developing an increased appreciation of the value of ecologically valid assessments, it is our view that the field of neuropsychology as a whole cannot move forward until this message penetrates outcomes research with neurologically intact participants. We propose five explanations as to what may be contributing to and perpetuating neuropsychologists' resistance to incorporating measures with seemingly high ecological validity (i.e. tests with verisimilitude as opposed to traditional tests) into such research protocols. Again, the domain of memory functioning will be used to help demonstrate points, where relevant.

6.1. Neuropsychologists assume that traditional tests are ecologically valid

Part of the reluctance to use tests with verisimilitude stems from a widely held belief that traditional tests are ecologically valid. Generally, neuropsychologists assume that the veridicality approach is adequate and that they can simply draw inferences between performance on traditional tests and everyday functioning (Sbordone, 1996). This belief has perpetuated despite: (a) limited investigation into the ecological validity of traditional tests and; (b) the findings from the studies that *have* been conducted generally indicating that the relationship between traditional measures and everyday functioning is poor (e.g. Goldstein & McCue, 1995; Higginson et al., 2000; Nadolne & Stringer, 2001).

However, with the recent publication of Chaytor's and Schmitter-Edgecombe's (2003) review of ecological validity approaches, and their conclusion that there is general support for the use of tests with verisimilitude, this belief may change. The developing literature on ecologically valid approaches may encourage the development and increased use of such tests which have sufficient sensitivity to detect subtle impairments in the cognitive performance of neurologically intact individuals.

6.1.1. Test inertia

Williams (1988) coined the term "test inertia", another factor that may impede the willingness for neuropsychologists to be more open to advancements in the development of assessment tools. This term refers to the resistance of neuropsychologists to accept new test developments and their over-reliance on traditional assessment instruments, typically those acquired during graduate training. According to Williams (1988), psychologists hold a relatively conservative view of new assessment methods and may develop a "romantic attachment" (p. 125) to particular tests or batteries. This may be due, at least in part, to the high degree of familiarity with the administration, scoring, and interpretation that comes from using the same tests over an extended period of time. It may also stem from a tendency for some neuropsychologists to strongly subscribe to particular theories of assessment or schools of thought, such as the Halsteid–Reitan or Luria–Nebraska approach, and therefore not actively seek to be informed about new test developments, particularly those that challenge long-held beliefs about assessment practices. Unfortunately, this may slow progression towards the development and use of more up-to-date and effective instruments, thus impeding the growth of the field of neuropsychology as a whole. To help prevent potential stagnation, neuropsychologists need to ensure they remain informed about new assessment developments and remain open to learning of their purpose and to their use.

6.2. Perceptions of face validity

The degree of verisimilitude of a test may be seen to overlap with the concept of face validity (Franzen & Wilhelm, 1996). This form of validity that is not based on technical aspects of the test, but rather on whether it appears to measure what it purports to measure (Cohen & Swerdlik, 2005). The RBMT and RBMT-E, for example, have high levels of face validity in that they clearly assess the concept of memory as individuals understand it in their everyday life. While sufficient face validity is important for purposes of social acceptability, it is not a concept based on strong psychometric principles and is of secondary importance to more technical forms of validity. Thus, neuropsychologists may perceive that tests with verisimilitude are 'un-scientific' or have not been subjected to more stringent validity assessments. This belief may stem from previous debates on methods of everyday memory assessment in which approaches with high ecological validity (e.g. direct observational methods) were criticized for being low in external validity (e.g. Banaji & Crowder, 1989). Such beliefs may be contributing to the resistance of neuropsychology researchers to use such tests, but are based on perceptions that are typically incorrect. The validity of the RBMT, for example, has been rigorously demonstrated via a series of studies examining its correlation with other measures and ratings of memory functioning (Wilson et al., 1989). Neuropsychologists need to be aware that well-established tests with verisimilitude have sound psychometric properties and go beyond satisfying the more basic concept of face validity. This information needs to be considered when selecting tests for use in clinical outcomes studies.

6.3. Views on the overlap with occupational therapy

The profession of occupational therapy is concerned with the assessment and rehabilitation of functional skills, including understanding the impact of cognitive difficulties on activities of daily living (Bennett, 2001; Hagedorn, 2001). The development and use of tests with verisimilitude, with their emphasis on everyday function, can be seen to 'overlap' with the principles of occupational therapy. Thus, neuropsychologists may regard the use of these tests as crossing into the realms of another discipline, thereby de-emphasizing the unique contributions we have been trained to provide in the practice of psychometric assessment. Despite this overlap, neuropsychologists continue to make unique contributions through comprehensive understanding of brain–behavior relationships and critical evaluation of cognitive assessment and rehabilitation practices. Use of such tests extends the boundaries of practice to the ultimate benefit of the individuals or samples assessed. Bennett (2001) emphasized that the ecological validity of neuropsychological assessment can be enhanced in rehabilitation settings by welcoming input from other allied health professions (e.g.

occupational therapy, speech pathology). This multidisciplinary approach will likely infiltrate the research arena over the coming years, including clinical outcomes research with neurologically intact participants. Acceptance of the intersecting contributions of allied health disciplines will help pave the way for this to occur, but also encourage us to recognize and promote our specialist contributions.

6.4. An over-reliance on the belief that traditional tests measure specific constructs

In research contexts, as in clinical practice, it is simple and convenient to report findings in terms of performance on specific and discrete domains of functioning rather than to make broader statements about capacity to function in daily life. The WMS-III, for example, is purported to measure specific domains of memory functioning as represented by its separate indices. In contrast, the RBMT-E provides one overall measure of everyday memory performance by classifying individuals into one of five groups (exceptionally good, good, average, poor, or impaired). On initial consideration, reporting scores on the different indices of the WMS-III before and after a clinical intervention would appear to provide more definitive and comprehensive information to researchers about the effects of the intervention on memory functioning, than would the RBMT-E. For this reason, the use of a traditional test such as the WMS-III may be more appealing. However, reporting results in terms of scores on specific domains can be problematic, as the purported domains may not have been established through appropriate statistical methods. More specifically, in some cases, the labeling of cognitive domains is not based on demonstrated construct validity (Johnstone & Frank, 1995). The Wechsler Memory Scale—Revised (WMS-R; Wechsler, 1987), for example, provides separate indices for Visual Memory, Verbal Memory, General Memory, Attention/Concentration, and Delayed Recall. An exploratory factor analysis of the scale indicated that two factors (General Memory and Attention/Concentration) provided a best fit of the data in the standardization sample and a mixed clinical sample (Wechsler, 1987). However, the other three indices were retained based on their face validity and because of their potential clinical utility, even though they were not found to represent distinct constructs.

A similar problem can be seen with the Wide Range Assessment of Memory and Learning (WRAML; Adams & Sheslow, 1990), a test of memory for children. The authors conducted a principal components analysis and reported three factors representing the domains of Verbal Memory, Visual Memory, and Learning. However, a visual learning test was found to load more strongly on the Visual Memory factor than the Learning Index, and a verbal memory subtest loaded more on the Learning Index than Verbal Memory (Johnstone & Frank, 1995; Spreen & Strauss, 1998). Thus, the widely held assumptions behind what these tests measure, and the labels applied to domains of functioning, may be dubious at best.

Similar problems in using labels to denote specific ‘constructs’ are evident in the confirmatory factor analytic studies of the WMS-III. Specifically, different methodological approaches used in these studies has led to published findings of a three-factor (Millis, Malina, Bowers, & Ricker, 1999), four-factor (The Psychological Corporation, 1997), and five-factor model (Burton, Ryan, Axelrod, Schellenberger, & Richards, 2003), each claiming to best fit the data. Each of the factors was purported to reflect a particular construct and labeled as such. These conflicting findings, resulting from methodological differences, clearly indicate that the application of labels to cognitive domains is not necessarily reflective of unambiguous empirical findings. Researchers, therefore, should be wary about using such labels when reporting study findings. Moreover, if they opt to use a traditional neuropsychological instrument (e.g. the WMS-III) instead of a more recently devised test with verisimilitude, they should be aware that this decision may be based on a false assumption that results from the use of traditional tests can be reported in terms of definitive cognitive constructs.

7. Concluding comments

The arguments presented in this paper are not intended to imply that traditional neuropsychological tests have no purpose in clinical outcomes studies with neurologically intact samples. We recognize that traditional tests play an important role in measuring performance across a broad range of cognitive functions, often with the benefit of good reliability, validity, and comprehensive normative data. However, given that the focus of neuropsychological assessment is shifting towards developing a greater understanding of performance of everyday behaviors, and in light of the findings of Chaytor and Schmitter-Edgecombe’s (2003) review, tests of verisimilitude may be the most effective to use to fulfill the new role that is emerging.

There appears to be no reason why tests of verisimilitude and traditional tests cannot amicably co-exist in research protocols. One option would be to develop hypotheses about what areas of cognition are likely to be affected by a specific pharmaceutical agent or medical intervention, based on previous studies using traditional tests. Following this, selected tests with verisimilitude can be incorporated into the research protocol to determine the degree to which deficits in those cognitive domains are likely to impact functioning in everyday life. At this stage, verisimilitude approaches to the assessment of memory (the RBMT-E) and attention (the TEA) appear to be well-validated for use with neurologically intact samples. However, further research is required to establish appropriate methods for the ecological assessment of executive functioning and other cognitive domains of importance. Specifically, future studies should investigate whether the BADS is sensitive to subtle differences in the performance of neurologically intact samples before it is deemed appropriate for the assessment of executive functioning in such groups. Furthermore, research investigating the ecological validity of existing, traditional tests of other domains (e.g. visuospatial processing, speed of information processing) will help determine whether new measures, developed with verisimilitude in mind, are required. There are some early indications that virtual reality tasks (which incorporate the principles of verisimilitude) are an effective and ecologically valid means of assessing specific cognitive functions in patients with brain impairment (e.g. Titov & Knight, 2005). With further refinement and research, this method of assessment may also prove to be useful in clinical outcomes studies with neurologically intact individuals. Such studies are becoming increasingly relevant in current health care systems, and it is hoped that the arguments presented in this paper will provide the impetus for further research and encourage researchers to consider in greater depth the ecological relevance of their findings. Neuropsychologists have the potential to contribute greatly to improving the understanding of the functional implications of cognitive difficulties, and the pursuit of these research goals can only serve to enhance the field as a whole.

References

Adams, W., & Sheslow, D. (1990). *WRAML manual*. Wilmington, DE: Jastak Associates.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44*, 1185–1193.

Barbee, J. G. (1993). Memory, benzodiazepines, and anxiety: Integration of theoretical and clinical perspectives. *Journal of Clinical Psychiatry*, *54*(Suppl.), 86–97.

Bennett, T. L. (2001). Neuropsychological evaluation in rehabilitation planning and evaluation of functional skills. *Archives of Clinical Neuropsychology*, *16*, 237–253.

Brezden, C. B., Phillips, K.-A., Abdolell, M., Bunston, T., & Tannock, I. F. (2000). Cognitive function in breast cancer patients receiving adjuvant chemotherapy. *Journal of Clinical Oncology*, *18*, 2695–2701.

Burton, D. B., Ryan, J. J., Axelrod, B. N., Schellenberger, T., & Richards, H. M. (2003). A confirmatory factor analysis of the WMS-III in a clinical sample with cross validation in the standardization sample. *Archives of Clinical Neuropsychology*, *18*, 629–641.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*, 181–197.

Chelune, G. J. (2002). Making neuropsychological outcomes research consumer friendly: A commentary on Keith et al. (2002). *Neuropsychology*, *16*, 422–425.

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement* (6th ed.). New York: McGraw-Hill.

Conway, M. A. (1991). In defense of everyday memory. *American Psychologist*, *46*, 19–26.

Derrer, D. S., Howieson, D. B., Mueller, E. A., Camicioli, R. M., Sexton, G., & Kaye, J. A. (2002). Memory testing in dementia: How much is enough? *Journal of Geriatric Psychiatry and Neurology*, *13*, 1–6.

deWall, C., Wilson, B. A., & Baddeley, A. D. (1994). The Extended Rivermead Behavioral Memory Test: A measure of everyday memory performance in normal adults. *Memory*, *2*, 149–166.

Dimitrov, M., Grafman, J., & Hollnagel, C. (1996). The effects of frontal lobe damage on everyday problem solving. *Cortex*, *32*, 357–366.

Donohoe, R. T., & Benton, D. (1999). Cognitive functioning is susceptible to the level of blood glucose. *Psychopharmacology*, *145*, 378–385.

Evans, J. J., Chua, S. E., McKenna, P. J., & Wilson, B. A. (1997). Assessment of the Dysexecutive Syndrome in schizophrenia. *Psychological Medicine*, *27*, 1–12.

Ford, S., Schofield, T., & Hope, T. (2003). What are the ingredients for a successful evidence-based patient choice consultation? A qualitative study. *Social Science and Medicine*, *56*, 589–602.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 91–112). Boca Raton, FL: St Lucie Press.

Geffen, G. M., Encel, J. S., & Forrester, G. M. (1991). Stages of recovery during post-traumatic amnesia and subsequent everyday memory deficits. *Cognitive Neuroscience and Neuropsychology*, *2*, 105–108.

Goldstein, G., & McCue, M. (1995). Differences between patient and informant functional outcome ratings in head-injured individuals. *International Journal of Rehabilitation and Health*, *1*, 25–35.

Goldstein, G., McCue, M., Rogers, J., & Nussbaum, P. D. (1992). Diagnostic differences in memory test based predictions of functional capacity in the elderly. *Neuropsychological Rehabilitation*, *2*, 307–317.

Green, J. D. W., Hodges, J. R., & Baddeley, A. D. (1995). Autobiographical memory and executive functioning in early dementia of the Alzheimer's type. *Neuropsychologia*, *33*, 1647–1670.

Griva, K., Newman, S. P., Hankins, M., Davenport, A., Hansraj, S., & Thompson, D. (2003). Acute neuropsychological changes in hemodialysis and peritoneal dialysis patients. *Health Psychology*, *22*, 570–578.

Hagedorn, R. (2001). *Foundations for practice in occupational therapy* (3rd ed.). London, England: Churchill Livingstone.

Heinrichs, R. W. (1990). Current and emergent applications of neuropsychological assessment: Problems of validity and utility. *Professional Psychology: Research and Practice*, *21*, 171–176.

Higginson, C. I., Arnett, P. A., & Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology*, *15*, 185–204.

Johnson, J. L. (1994). Episodic memory deficits in Alzheimer's disease: A behaviorally anchored scale. *Archives of Clinical Neuropsychology*, *9*, 337–346.

Johnson, T., Monk, T., Rasmussen, L. S., Abildstrom, H., Houx, P., Korttila, K., et al. (2002). Postoperative cognitive dysfunction in middle-aged patients. *Anesthesiology*, *96*, 1351–1357.

Johnstone, B., & Farmer, J. E. (1997). Preparing neuropsychologists for the future: The need for additional training guidelines. *Archives of Clinical Neuropsychology*, *12*, 523–530.

Johnstone, B., & Frank, R. G. (1995). Neuropsychological assessment in rehabilitation: Current limitations and applications. *NeuroRehabilitation*, *5*, 75–86.

Kaitaro, T., Koskinen, S., & Kaipio, M. L. (1995). Neuropsychological problems in everyday life: A 5-year follow-up study of young severely closed-head-injured patients. *Brain Injury*, *9*, 713–727.

Larabee, G. J., & Crook, T. H. (1996). The ecological validity of memory testing procedures: Developments in the assessment of everyday memory. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 225–242). Boca Raton, FL: St Lucie Press.

Leber, P. (1986). Establishing the efficacy of drugs with psychogeriatric indications. In T. Crook, R. T. Bartus, S. Ferris, & S. Gershon (Eds.), *Treatment development strategies for Alzheimer's disease* (pp. 1–14). Madison, CT: Mark Powley Associates.

Long, C. J., & Collins, L. F. (1997). Ecological validity and forensic neuropsychological assessment. In R. J. McCaffrey, A. D. Williams, J. M. Fisher, & L. C. Laing (Eds.), *The practice of forensic neuropsychology: Meeting challenges in the courtroom* (pp. 153–164). New York: Plenum.

Long, C. J., & Kibby, M. Y. (1995). Ecological validity of neuropsychological tests: A look at neuropsychology's past and the impact that ecological issues may have on its future. *Advances in Medical Psychotherapy*, *8*, 59–78.

Makatura, T. J., Lam, C. S., Leahy, B. J., Castillo, M. T., & Kalpakjian, C. Z. (1999). Standardized memory tests and the appraisal of everyday memory. *Brain Injury*, *13*, 355–367.

Millis, S. R., Malina, A. C., Bowers, D. A., & Ricker, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale—III. *Journal of Clinical and Experimental Neuropsychology*, *21*, 87–93.

Morrow, L. A., Steinhauer, S. S., Condray, R., & Hodgson, M. (1997). Neuropsychological performance of journeymen painters under acute solvent exposure and exposure-free conditions. *Journal of the International Neuropsychological Society*, *3*, 269–275.

Nadolne, M. J., & Stringer, A. Y. (2001). Ecological validity in neuropsychological assessment: Prediction of wayfinding. *Journal of the International Neuropsychological Society*, *7*, 675–682.

Neisser, U. (1978). Memory: What are the important questions? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 3–24). London: Academic Press.

Polo-Kantola, P., Portin, R., Polo, O., Helenius, H., Irjala, K., & Erkkola, R. (1998). The effect of short-term estrogen replacement therapy on cognition: A randomized, double-blind, cross-over trial in postmenopausal women. *Obstetrics and Gynecology*, *91*, 459–466.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention*. Bury St. Edmunds, England: Thames Valley Test Company.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). Structure of normal human attention: The Test of Everyday Attention. *Journal of the International Neuropsychological Society*, *2*, 525–534.

Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 15–41). Boca Raton, FL: St Lucie Press.

Schwatrz, A. F., & McMillan, T. M. (1989). Assessment of everyday memory after severe head injury. *Cortex*, *25*, 665–671.

Spreen, O., & Strauss, I. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Harcourt Brace & Company.

Stephens, C., Hamilton, Y. M., & Pachana, N. A. (2003). Hormone replacement therapy and everyday memory in mid-aged New Zealand women. *New Zealand Journal of Psychology*, *32*, 13–21.

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behaviour*, *22*, 341–357.

Sweet, J. J., Moberg, P. J., & Suchy, Y. (2000). Ten-year follow-up survey of clinical neuropsychologists: Part I. Practices and beliefs. *The Clinical Neuropsychologist*, *14*, 18–37.

Sweet, J. J., Peck, E. A., Abramowitz, C., & Etzweiler, S. (2002). National Academy of Neuropsychology/Division 40 of the American Psychological Association Practice survey of clinical neuropsychology in the United States, Part I: Practitioner and practice characteristics, professional activities, and time requirements. *The Clinical Neuropsychologist*, *16*, 109–127.

Tannock, I. F., Ahles, T. A., Ganz, P. A., & van Dam, F. S. (2004). Cognitive impairment associated with chemotherapy for cancer: Report of a workshop. *Journal of Clinical Oncology*, *22*, 2233–2239.

Taylor, D. J., & McFatter, R. M. (2003). Cognitive performance after sleep deprivation: Does personality make a difference. *Personality and Individual Differences*, *33*, 1179–1193.

The Psychological Corporation. (1997). *WAIS-III–WMS-III technical manual*. San Antonio, TX: Harcourt Brace.

Titov, N., & Knight, R. G. (2005). A computer-based procedure for assessing functional cognitive skills in patients with neurological injuries. *Brain Injury*, *19*, 315–322.

Troster, A. I. (2000). Clinical neuropsychology, functional neurosurgery, and restorative neurology in the next millennium: Beyond secondary outcome measures. *Brain and Cognition*, *42*, 117–119.

Warga, C. (1999). *Menopause and the mind*. New York: The Free Press.

Wechsler, D. (1987). *Wechsler Memory Scale—Revised Manual*. San Antonio, TX: The Psychological Corporation.

Williams, J. M. (1988). Everyday cognition and the ecological validity of intellectual and neuropsychological tests. In J. M. Williams & C. J. Long (Eds.), *Cognitive approaches to neuropsychology* (pp. 123–141). New York: Plenum Press.

Wills, P., Clare, L., Shiel, A., & Wilson, B. A. (2000). Assessing subtle memory impairments in the everyday memory performance of brain injured people: Exploring the potential of the Extended Rivermead Behavioral Memory Test. *Brain Injury*, *14*, 693–704.

Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied and Preventative Psychology*, *2*, 209–215.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioral Assessment of the Dysexecutive Syndrome*. Bury St. Edmunds: Thames Valley Test Company.

Wilson, B. A., Clare, L., Cockburn, J. M., Baddeley, A. D., Tate, R., & Watson, P. (1999). *The Rivermead Behavioral Memory Test—Extended Version*. Bury St. Edmunds: Thames Valley Test Company.

Wilson, B. A., Cockburn, J. M., & Baddeley, A. D. (1985). *The Rivermead Behavioral Memory Test*. Reading, England/Gaylord, MI: Thames Valley Test Company/National Rehabilitation Services.

Wilson, B. A., Cockburn, J., Baddeley, A., & Hiorns, R. (1989). The development and validation of a test battery for detecting and monitoring everyday memory problems. *Journal of Clinical and Experimental Neuropsychology*, *11*, 855–870.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. W. (1998). The development of an ecologically valid test for assessing patients with a Dysexecutive Syndrome. *Neuropsychological Rehabilitation*, *8*, 213–228.

Wilson, B. A., Shiel, A., Foley, J., Emslie, H., Groot, Y., Hawkins, K., et al. (2004). *Cambridge Test of Prospective Memory*. Bury St. Edmunds, England: Thames Valley Test Company.