Intelligence 37 (2009) 283–293

Contents lists available at ScienceDirect

Intelligence

A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence

Jill T. Shelton *,[1], Emily M. Elliott [2], B.D. Hill [3], Matthew R. Calamia [4], Wm. Drew Gouvier [2]

Louisiana State University, United States

ARTICLE INFO

Article history:
Received 31 March 2008
Received in revised form 5 November 2008
Accepted 7 November 2008
Available online 21 December 2008

Keywords:
Working memory
Memory assessment
Intelligence
Fluid abilities
Translational research

ABSTRACT

The working memory (WM) construct is conceptualized similarly across domains of psychology, yet the methods used to measure WM function vary widely. The present study examined the relationship between WM measures used in the laboratory and those used in applied settings. A large sample of undergraduates completed three laboratory-based WM measures (operation span, listening span, and n-back), as well as the WM subtests from the Wechsler Adult Intelligence Scale-III and the Wechsler Memory Scale-III. Performance on all of the WM subtests of the clinical batteries shared positive correlations with the lab measures; however, the Arithmetic and Spatial Span subtests shared lower correlations than the other WM tests. Factor analyses revealed that a factor comprising scores from the three lab WM measures and the clinical subtest, Letter-Number Sequencing (LNS), provided the best measurement of WM. Additionally, a latent variable approach was taken using fluid intelligence as a criterion construct to further discriminate between the WM tests. The results revealed that the lab measures, along with the LNS task, were the best predictors of fluid abilities.

Working memory (WM) was recently defined as "a temporary storage system under attentional control that underpins our capacity for complex thought" (Baddeley, 2007, p. 1). The strong relationship between WM and complex cognition underscores the key importance of this construct in many aspects of human behavior. WM has been heavily investigated by researchers, and has been shown to play a key role in complex behaviors such as reading comprehension (Daneman & Carpenter, 1980), the acquisition of language (Baddeley, Gathercole, & Papagno, 1998), and fluid abilities (Gray, Chabris, & Braver, 2003; Salthouse & Pink, 2008). Additionally, the importance of the construct has been noted across areas of psychology, as researchers in clinical psychology have evaluated the relationship of WM to deficits in schizophrenia (Barch, 2003) and depression (Harvey et al., 2004), social psychologists have assessed the role of WM in stereotype threat (Bonnot & Croizet, 2007), and neuropsychologists have assessed WM ability as a way to identify the early onset of Alzheimer's disease (Rosen, Bergeson, Putnam, Harwell, & Sunderland, 2002).

The prominence of the WM construct has resulted in the use of different measurement techniques that often vary substantially in their methodology. For example, cognitive and clinical psychologists typically define WM similarly, but use different methods to assess WM function. Cognitive psychologists use laboratory tasks that have been extensively analyzed for their reliability and construct validity (for review see Conway et al., 2005). Clinical psychologists often use psychometric indices, such as subscales of the Wechsler Adult Intelligence Scale (WAIS-III) and the Wechsler Memory Scale (WMS-III), to

* Corresponding author. Campus Box #1125, Department of Psychology, Washington University, St. Louis, MO 63130, United States. Tel.: +1 314 935 7610; fax: +1 314 935 7588.
E-mail address: jshelton@artsci.wustl.edu (J.T. Shelton).
[1] Now at Department of Psychology at Washington University in St. Louis, United States.
[2] Now at Department of Psychology, Louisiana State University, United States.
[3] Now at Department of Psychology at Brown University, United States.
[4] Now at Department of Psychology at the University of Iowa, United States.

0160-2896/$ – see front matter
doi:10.1016/j.intell.2008.11.005

measure WM function. The assumption is that the psychometric instruments used in the clinical setting accurately depict the WM construct discussed by cognitive psychologists. However, this assumption has not been fully tested. Ackerman, Beier, and Boyle (2005) commented on this issue in more general terms by saying that, “many intelligence measures have been developed with substantially greater criterion-related validity, as opposed to construct validity” (p. 31). The present study addressed these concerns by providing a systematic evaluation of the construct and criterion-related validity of various WM measures.

1. Laboratory assessment of WM

Laboratory studies often utilize tasks, such as the operation span task (Ospan; Turner & Engle, 1989), reading span task (Daneman & Carpenter, 1980), or the n-back task (Dobbs & Rule, 1989) to assess WM function. Complex span tasks, such as Ospan and reading span, require participants to retain a list of items (storage component), while simultaneously engaging in a secondary activity, such as solving math problems (processing component). It is assumed that the central executive distributes attentional resources to the memory system to enable an individual to meet the complex demands of the task. Storage and processing tasks were specifically designed to support the theoretical assumptions held about how the WM system operates. Furthermore, these tasks have been repeatedly shown to be reliable measures of WM that demonstrate excellent construct and criterion-related validity (for a list of the many higher order cognitive tasks that correlate with WM, see Conway et al., 2005, p.777).

The lag, or n-back, task is less commonly used to measure WM function in the laboratory setting; however, the demands inherent within this task make it a potentially good measurement tool. It was originally developed by Kirschner (1958) to examine general retrieval processes. More recently, researchers have used this task to examine WM function in neurological settings with brain injury patients (Cohen et al., 1994), in the aging population (Dobbs & Rule, 1989; Kwong See & Ryan, 1995), and to study focal attention (McElree, 2001). In the n-back task, participants are presented are asked to recall or recognize an item that fell in a particular serial position in a presented list. This task does not contain some of the features present in traditional WM measures (e.g., secondary processing task); however, several of its features do reflect important aspects of the WM construct. The items being presented must be actively maintained for later recall, while controlled attention is used to guide the retrieval process in meeting task demands (identifying the item located in a particular position in the list). The absence of a secondary processing component intermittent throughout the test trials is an important difference that likely leads to different strategies being used to perform these tasks. More specifically, the n-back task may not provide a clear indication of the capacity of WM, rather a person's ability to efficiently update the contents of WM to better maintain current task goals.

Recent studies have offered conflicting results on the utility of the n-back task as a valid measure of WM (Kane, Conway, Miura, & Colflesh, 2007; Shelton, Metzger, & Elliott, 2007). One key difference between these studies is the version of the task used; that is, participants were either asked to perform recall or recognition. For example, Shelton et al. (2007) used a recall version of the task, and observed strong relationships between n-back and Ospan performance. Kane et al. (2007), on the other hand, used a recognition version of the task and did not find strong relationships between n-back performance and performance on traditional storage and processing measures. It is clear that further evaluation of this task is needed to support its usefulness as a measure of WM. This is one sub-goal of the present research.

2. Clinical assessment of WM

WM is important clinically as it is impaired in a wide variety of neuropsychiatric conditions, including dementia (Collette, Van der Linden, & Salmon, 1999), attention-deficit hyperactivity disorder (ADHD; Pasini, Paloscia, Alessandrelli, Porfirio, & Curatolo, 2007), and schizophrenia (Fleming, Goldberg, Goldman, & Weinberger, 1995; Goldman-Rakic, 1994). Efficacy for therapeutic interventions in these groups is typically demonstrated by measured improvements in executive functions such as WM. Additionally, executive dysfunction has been shown to deleteriously impact a number of clinical factors such as functional outcome (Boyle et al., 2003), medication compliance (Hinkin et al., 2002), and capacity to give informed consent (Marson, Chatterjee, Ingram, & Harrell, 1996). Currently, it is unclear whether clinical tests of WM are assessing the same WM construct discussed in the experimental cognitive literature.

The WAIS-III (Wechsler, 1997a) is the most commonly used measure of intelligence in clinical settings (Camara, Nathan, & Puente, 2000; Heijden & Donders, 2003; Rabin, Barr, & Burton, 2005). This test generates four summary indices, one of which is the Working Memory Index. The WM Index is derived from the following WAIS-III subtests: Digit Span (Dspan), Arithmetic, and Letter/Number Sequencing (LNS). Performance on the Dspan subtest reflects a combined measure of accuracy in the forward and backward conditions. Although there have been some questions raised regarding this practice (see Reynolds, 1997), the combined score was used in the current research as this is common clinical methodology. It should also be pointed out that the Dspan task is considered a simple span test because it only contains a storage component, in contrast to complex span measures that also contain a secondary processing requirement.

The Arithmetic portion of the WM index consists of word problems read aloud to participants with increasing levels of difficulty with each new problem. Difficulty level is determined by the amount of information that has to be held in memory in order to successfully complete the problem. This task does place demands on the WM system; however, other factors likely contribute to accuracy on this task, such as mathematical efficiency (Stearns, Dunham, McIntosh, & Dean, 2004). Math anxiety has also been linked to impaired performance on WM tasks that contain a mathematical element, such as computation span (Ashcraft & Kirk, 2001).

In the LNS task, the experimenter reads mixed lists of digits and letters aloud to the participants and they are asked to recall this list in correct numeric and alphabetic order. This task involves additional processing requirements similar to that of traditional WM tasks. Haut, Kuwabara, Leach, and Arias (2000) examined LNS performance in a PET study and