1-1-2011

Comparison of Wechsler Memory Scale–Fourth Edition (WMS–IV) and Third Edition (WMS–III) dimensional structures: Improved ability to evaluate auditory and visual constructs

James B. Hoelzle
Marquette University, james.hoelzle@marquette.edu

Nathaniel W. Nelson
University of Minnesota

Clifford A. Smith

Accepted version. Journal *of Clinical and Experimental Neuropsychology,* Volume 33, Issue 3, pp 283-291 (March, 2011). DOI.

Comparison of Wechsler Memory Scale–Fourth Edition (WMS– IV) and Third Edition (WMS–III) Dimensional Structures: Improved Ability to Evaluate Auditory and Visual Constructs

James B. Hoelzle

Psychology Department, Marquette University
Milwaukee, WI

Nathaniel W. Nelson

Psychology Department, University of Minnesota
Minneapolis, MN

Clifford A. Smith

Independent Consultant, Iron Mountain
WI

Journal of Clinical and Experimental Neuropsychology, Vol. 33, No. 3 (2011): pg. 283-291. DOI.

Abstract

Dimensional structures underlying the Wechsler Memory Scale–Fourth Edition (WMS–IV) and Wechsler Memory Scale–Third Edition (WMS–III) were compared to determine whether the revised measure has a more coherent and clinically relevant factor structure. Principal component analyses were conducted in normative samples reported in the respective technical manuals. Empirically supported procedures guided retention of dimensions. An invariant two-dimensional WMS–IV structure reflecting constructs of auditory learning/memory and visual attention/memory (C1 = .97; C2 = .96) is more theoretically coherent than the replicable, heterogeneous WMS–III dimension (C1 = .97).This research suggests that the WMS–IV may have greater utility in identifying lateralized memory dysfunction.

The construct of memory is broad and diverse, and no single anatomical structure is comprehensively responsible for learning and storing all forms of sensory information (Lashley, 1950). For example, the striatum, cerebellum, and amygdale are believed to be integral for specific aspects of nondeclarative memory, whereas medial temporal structures and the diencephalon play significant roles in declarative memory (Bear, Connors, & Paradiso, 2001).The latter construct is most relevant to neuropsychological assessment and is often further differentiated by material-specific learning and recall. For example, researchers have suggested that auditory memory is differentially dependent on left temporal lobe structures, while visual/perceptual memory is differentially dependent on right temporal lobe structures (e.g., Gleiβner, Helmstaedter, & Elger, 1998; Milner, 1968). Psychometric memory tests demonstrate clinical utility by quantifying these distinct constructs, which informs differential diagnosis and treatment.

Factor analysis is one way to determine whether clinical instruments evaluate meaningful constructs such as auditory and visual/perceptual memory. A useful instrument should have an underlying structure that reflects diagnostically relevant constructs. However, in contrast to this position, based upon the results of numerous factor analytic studies that failed to differentiate between important immediate and delayed memory constructs, some researchers have suggested that factor analysis should not be implemented to evaluate memory instruments (e.g., see, Delis, Jacobson,

Bondi, Hamilton, & Salmon, 2003; Jacobson, Delis, Hamilton, Bondi, & Salmon, 2004; Millis, Malina, Bowers, & Ricker, 1999). The failure of data reduction methods to differentiate between these constructs is related to significant shared variance between immediate and delayed memory tasks (i.e., efficient delayed memory is to a degree dependent upon intact immediate memory).Given the shared variance between immediate and delayed memory tasks, it is inappropriate to expect, and highly unlikely, that corresponding factors would be observed. Incidentally, it also explains why well supported psychometric theories of cognitive ability, based largely upon the results of factor analytic studies (e.g., Carroll, 1993; McGrew 2009), do not include immediate and delayed constructs.[1]

While failure to reliably identify immediate and delayed memory constructs is an important methodological limitation to acknowledge when interpreting results or developing theory, it does not render the statistical data reduction approach useless. For example, consideration of discrepant Wechsler Memory Scale– Third Edition (WMS–III; Wechsler, 1997b) factor analytic studies illustrates how this methodological approach informs clinical practice and ultimately suggests that WMS–III index scores should be interpreted cautiously. The WMS–III technical manual initially reported that confirmatory factor analytic (CFA) results supported a five-factor model consisting of auditory immediate, auditory delayed, visual immediate, visual delayed, and working memory constructs. However, Millis et al. (1999) and Price, Tulsky, Millis, and Weiss (2002) could not replicate these analyses. Millis and colleagues attributed failure to replicate the previously described model to the very high correlations between immediate and delayed memory tasks. They also expressed concern that evaluation of visual memory might be "flawed" because of insufficient commonality between Faces and Family Pictures subtests. It is challenging to describe the WMS– III factor structure; the literature includes compelling factor analytic studies of the WMS–III that posit an underlying four-factor structure(Burton, Ryan, Axelrod, Schellenberger, & Richards, 2003; auditory, visual, working memory, and learning factors),three-factor structure (Millis et al., 1999, and Price et al., 2002;verbal,visual, and working memory factors),and two-factor structure (Wilde et al., 2003; general and working memory factors).

Consideration of WMS–III factor analytic literature is clinically relevant because it allows one to evaluate whether index scores are composed of relatively homogenous variance. This issue is especially relevant in clinical contexts that require documentation of lateralized memory functioning (e.g., presurgical evaluation for temporal lobotomy in the context of intractable epilepsy). For example, Wilde and colleagues' (2003) two factor solution does not reflect distinct constructs of auditory and verbal memory because there is insufficient commonality between visual subtests, Faces and Family Pictures, which is plausibly related to the Family Pictures subtests being verbally mediated. This finding is concerning and suggests that the interpretation of WMS–III visual memory indices may be confounded by construct irrelevant factors (e.g., verbal memory functioning). Heterogeneous variance within indices decreases sensitivity of the WMS–III and makes clear how being knowledgeable of the factor structure underlying any psychometric instrument is an important aspect of understanding diagnostic utility. Alternative WMS–III index scores have been developed because of this limitation, and interpretation of these indices may be clinically warranted (e.g., see Tulsky, Ivnik, Price, & Wilkins, 2003; Tulsky & Price, 2003).

The Wechsler Memory Scale–Fourth Edition (WMS– IV; Wechsler, 2009) was recently developed to improve upon several notable shortcomings of the WMS–III, including issues contributing to nonoptimal sensitivity to memory impairment (e.g., range restriction, problematic scoring floors, and verbally mediated visual memory tasks). The WMS–IV technical manual includes CFA results that support an a priori theoretical model of visual memory (Designs II and Visual Reproduction II subtests), visual working memory (Symbol Span and Spatial Addition subtests), and auditory memory (Logical Memory II and Verbal Paired Associates II subtests). A two-factor model consisting of visual (Designs II, Visual Reproduction II, Spatial Addition, and Symbol Span subtests) and auditory (Logical Memory II and Verbal Paired Associates II subtests) constructs was also supported. Fit indices were not statistically different between two-and three-factor models. The decision was made to include three WMS–IV index scores based on response processes evaluated, not necessarily the results of preliminary CFA. It is noteworthy that immediate and delayed memory subtests were not included in the initial analyses because correlations among immediate and delayed subtests were greater than the correlations among subtests within the same domain (e.g., WMS–IV Logical Memory and Verbal Paired Associates subtests).

Given the conflicting body of literature describing the WMS–III factor structure and the importance of psychometric properties on clinical decision making, we sought to compare underlying dimensional structures of the WMS– IV and WMS–III. Similar methodology was applied to normative data presented in respective technical manuals (Wechsler 1997b, 2009) and will permit direct and relevant comparison of factor structures. Findings will assist clinicians and researchers in determining whether the WMS–IV has a more coherent and clinically relevant factor structure than the WMS–III. Results will also be beneficial in further understanding psychometric properties of new and relatively unknown WMS–IV subtests: Designs, Symbol Span, and Spatial Addition.

Method

Participants

Data were obtained from the WMS–IV and WMS–III technical manuals (Wechsler, 1997b, 2009). The study made use of 18 normative samples that each included 100 individuals. WMS–IV data consisted of nine age-based correlation matrices that includedthefollowing10 subtest scores that contribute to primary index scores: Logical Memory I, Logical Memory II, Verbal Paired Associates I, Verbal Paired Associates II, Designs I, Designs II, Visual Reproduction I, Visual Reproduction II, Spatial Addition, and Symbol Span. WMS–III data consisted of nine age-based correlation matrices that included the following 11 subtest scores that contribute to primary index scores: Logical Memory I, Logical Memory II, Verbal Paired Associates I, Verbal Paired Associates II, Faces I, Faces II, Family Pictures I, Family Pictures II, Letter–Number Sequencing, Spatial Span, and Auditory Recognition Delayed.

Correlation matrices were composed of data collected from the follow age-based normative samples, 16–17-year-olds, 18–19-year-olds, 20–24-year-olds, 25–29-yearolds, 30–34-year-olds, 35–44-year-olds, 45–54-year-olds, 55–64-year-olds, and 65–69- year-olds, respectively.

While some researchers have suggested that a sample size of 100 is appropriate to conduct factor analyses (e.g., see, Gorsuch, 1983; Hatcher, 1994; Kline, 1979), others have recommended that larger samples are necessary (e.g., Cattell, 1978; Guilford, 1954). In reality, a well-selected set of test variables (i.e., those that are a good measure of a factor) can produce stable solutions across smaller samples (Velicer & Fava, 1998). Identification of a replicable factor solution across samples (methodology described below) dramatically decreased the likelihood that results were arbitrarily influenced by relatively modest sample sizes. As further protection against arbitrarily influenced results, previously described age-based normative samples were combined resulting in respective WMS–IV and WMS–III normative samples that each included 900 individuals. Supplemental analyses were conducted on the combined samples, and results were compared with those obtained from analysis of more narrow age-bands.

Immediate and delayed subtests were included in datasets to increase the number of marker variables analyzed. Differing from confirmatory factor analysis, exploratory approaches require a larger number of marker variables in datasets (Kim & Mueller, 1978).Typically, a minimum of three variables are needed to define a dimension (Goldberg & Velicer, 2006). If the methodological decision were made to analyze only immediate or delayed subtests, it would be highly unlikely that multifactor solutions would be identified due to a restricted number of auditory learning and memory variables. Given the restricted number of test variables available for analyses, we believe it was psychometrically desirable to analyze a combination of immediate and delayed subtests. Notably, we acknowledge a legitimate limitation of this decision is that correlations between immediate and delayed subtests are frequently higher than those within the same domain (e.g., see Millis et al., 1999).

Procedure

To supplement competing-model CFA conducted in the WMS–IV and WMS–III technical manuals (Wechsler 1997b, 2009), unrestrictive exploratory principal component analyses (PCAs) were conducted in each age-based sample. Oblique (oblimin) rotation was used because it is widely accepted that cognitive constructs are correlated with one another (Carroll, 1993; Deary, 2000).

Parallel analysis (PA) and Velicer's (1976) minimum average partial (MAP) procedure were used to determine the number of components underlying a set of variables. These methods improve upon several limitations of more traditional guidelines, such as Cattell's (1966) scree test or Kaiser's (1960) criterion (e.g., see Frazier & Youngstrom, 2007; Hoelzle & Meyer, 2009; Zwick & Velicer, 1982, 1986).

Briefly, PA determines the eigenvalues from random datasets containing the same number of "variables" and "cases" as the actual data. Components are retained if the actual eigenvalue is larger than the corresponding 95th percentile of eigenvalues generated across random datasets (Glorfeld, 1995; Longman, Cota, Holden, & Fekken, 1989). The MAP procedure is an iterative process focusing on the average squared partial correlation amongst test variables. The average squared partial correlation is computed prior to and after each subsequent extraction of a component. When a component is extracted that is composed of unique, variable specific variance the squared partial correlation increases, which suggests over extraction. Thus, the smallest average squared partial correlation value observed indicates the number of components to extract. The interested reader is referred to O'Connor (2000) for a more detailed description of PA and the MAP procedure.

Barrett's (2005) Orthosim software was used to determine the extent that extracted dimensions defined similar multidimensional space across age-based normative samples. Orthogonal vector matrix comparisons were conducted by maximally aligning two complete *m*-dimensional orthogonal solutions. Congruency coefficients range from −1.0 to 1.0 and represent the extent to which a fixed set of variables have similar component coefficients from one solution to the next.

Congruency coefficients >.90 typically indicate replicated factors, though both more restrictive and lenient benchmarks have also been proposed (e.g., see Barrett, 2005; Lorenzo-Seva & ten Berge, 2006; MacCallum, Widaman, Zhang, & Hong, 1999). Orthogonal rather than oblique solutions are matched to avoid artificial overfitting.[2] This procedure assisted in determining the most differentiated structure that was replicated across normative samples.

Results

WMS–IV

Eigenvalues corresponding with the first component in each age-based sample were much greater than the corresponding eigenvalue generated from 500 random datasets (mean difference was 2.91).Five of nine samples had second component eigenvalues greater than the corresponding PA eigenvalue (16–17-year-olds, 30–34-year-olds, 35–44-year-olds, 55–64-year-olds, 65–69-year-olds; mean difference was 0.26). PA did not support retention of three components in any age-based sample. In each instance, the third eigenvalue generated from random data was larger than the eigenvalue derived from normative data (mean difference was –0.27).Overall, PA supported retention of one or two components across age-based normative samples.

MAP procedure results were somewhat ambiguous and supported retention of one, two, or three WMS–IV components across age-based samples (see Figure1). Average squared partial correlations were somewhat invariant after extraction of the first three components (i.e., for many samples there was not a clear trajectory of decreasing or increasing average squared partial correlations). MAP values appear to increase after the third component is sequentially extracted, and more clearly after the fourth and fifth components, which reflects extraction of variable-specific variance (as opposed to common variance). MAP results suggest it would be inappropriate to retain four or more components.

Given support from MAP and PA for consideration of one-, two-, and three-dimensional solutions, congruency analyses were conducted to identify the most multidimensional solution that was consistent across age-based samples. Three-dimensional WMS–IV solutions were not consistent across samples and are challenging to concisely summarize. Congruency coefficients support initial observation that three-dimensional structures were inconsistent across samples. While the mean congruency coefficient for each component was >.90 (C1 = .97; C2 = .91; C3 = .91), nearly one third of the total congruency coefficients for Component 2 and Component 3 were <.90 (C2 = 25/72; C3 = 26/72), which indicates a meaningful degree of inconsistency across solutions. Given that three dimensional solutions were inconsistent across samples, mean pattern matrix loadings are not reported.[3]

The majority of samples produced three-dimensional structures that included rather specific dimensions of (a) Logical Memory subtests, (b) Verbal Paired Associates subtests, and (c) visual attention/memory. Designs sub-tests generally had the largest loadings on visual attention/memory dimensions. In four age-based samples Visual Reproduction subtests had similar, moderate loadings on two of the three dimensions (20–24-yearolds, 25–29-year-olds, 45–54-year-olds, 55–64-year-olds; pattern matrix loadings ranged from |.31| to |.65|). The oldest age-based sample (65–69-year-olds) produced a solution that reflected (a) Logical Memory, (b) Designs, and (c) Visual Reproduction subtests. Curiously, Verbal Paired Associates subtests had comparable, moderate loadings on dimensions that reflected Logical Memory and Visual Reproduction subtests in that solution (pattern matrix loadings varied from |.52| to |.56|).

A robust two-dimensional WMS–IV structure was consistent across age-based normative samples (C1 = .97; C2 = .96) and emphasized moderately correlated dimensions of (a) auditory learning/memory and (b) visual attention/memory. The auditory learning/memory dimension reflected Logical Memory and Verbal Paired Associates subtests, whereas the Designs subtests were primarily reflected on visual attention/memory dimensions. Visual Reproduction, Spatial Addition, and Symbol Span subtests also had large comparable loadings on dimensions that reflected visual attention/memory. Average pattern matrix loadings for each subtest were significant and distinct (i.e., all pattern matrix loadings >.60 on content-specific dimension and <.40 on non-content-specific dimension) and are presented in Table 1. Investigation of single dimension WMS–IV structures is precluded by the finding that two-dimensional solutions were invariant.

WMS–III

Across age-based samples, eigenvalues corresponding with first and second WMS–III components were greater than those generated from 500 random datasets (mean differences were 2.93 and 0.22, respectively). There was minimal support for the retention of three or four WMS–III components. Three age-based samples produced solutions with third component eigenvalues larger than the corresponding PA eigenvalue (16–17-year-olds, 18–19-year-olds, 25–29-year-olds; mean difference was 0.22); one age-based sample produced a solution with a fourth component eigenvalue that was comparable with the corresponding PA eigenvalue (16–17-year-olds; difference was 0.07). PA did not support retention of five components in any age-based sample. In each instance, the fifth eigenvalue generated from random data was larger than eigenvalues derived from normative data (mean difference was −0.35).Overall, WMS–III PA results largely supported retention of two components across samples, though in some samples there was support for retention of three components.

MAP procedure results were somewhat ambiguous (see Figure 2). Support is strongest for the retention of one component based upon the average squared partial correlations being lowest after extraction of one dimension. The MAP value appears to increase after extraction of second and third dimensions in all but three samples (18– 19-year-olds, 20–24-year-olds, and 55–64-year-olds). MAP procedure results do not support retention of our or more components as the average squared partial correlations trend up after the extraction of three components.

Next, PA-and MAP-supported models (single-, two-, and three-dimensional models) were reviewed to determine whether dimensional structures were consistent across age-based samples. Three-dimensional WMS–III structures were not replicable across samples

(C1 = .92; C2 = .85; C3 = .80).[4] Only one solution across the nine samples was consistent with what might have been intuitively expected—that is, reflected (a) verbal, (b) visual, and (c) working memory dimensions (20–24year-olds). Five of nine solutions consisted of (a) auditory learning and memory,(b) Family Pictures, and (c) Faces dimensions (18–19-year-olds, 25–29-year-olds, 30– 34-year-olds, 35–44- year-olds, 45–54-year-olds), whereas only one sample produced a solution that included markers of (a) visual attention and memory, (b) Logical Memory, and (c) Verbal Paired Associates subtests (16– 17-year-olds). The two oldest age-based samples (55– 64-year-olds, 65–69-year-olds) produced solutions that included dimensions reflecting combinations of auditory and visual subtests.

WMS–III two-dimensional solutions were also not replicable across samples (C1 = .94, C2 = .84). The most frequently observed two-factor solution included general and facial memory dimensions (25–29-year-olds, 30–34-year-olds, 35–44-year-olds, 45–54- year-olds, 65– 69-year-olds). The general memory dimension in these solutions largely reflected Logical Memory and Verbal Paired Associates subtests. Distinct from that commonly observed solution, the sample composed of 18–20-yearolds produced a solution with general and family picture memory dimensions, and the sample composed of 55–64year-oldsproducedasolutionreflecting general memory and Logical Memory subtests. The sample composed of 20–24-year-olds produced a solution that reflected constructs of general and working memory. Only one solution (16–17-year-olds) included coherent auditory and visual memory dimensions. Overall, it is difficult to summarize two-and three-dimensional WMS–III structures, and congruency coefficients reflect the notable inconsistency. Significant variability across solutions precludes presentation of average pattern matrix loadings.

Next, a single component was extracted from each WMS–III normative age-based sample. This dimension was found to be replicable (C1= .97), and average pattern matrix loadings are presented in Table 2. The dimension most significantly reflected Logical Memory, Verbal Paired Associates, and Family Pictures subtests. It is notable that visual memory subtests had strikingly different mean pattern matrix loadings on the single dimension. Faces were the only subtests with average pattern matrix loadings <.40.

Supplemental Analyses

To investigate whether previously conducted analyses were arbitrarily influenced by modest sample sizes, age-based samples were combined for supplemental analyses. Results of PA and the MAP procedure recommended retaining 1 or 2 components in combined WMS–IV and WMS–III normative samples. Average congruency coefficients between combined and age-based WMS–IV samples support an invariant two-factor solution consisting of auditory learning/memory and visual attention/memory (C1 =.98; C2 =.98). Average congruency coefficients between combined and age-based WMS–III samples are acceptable for a two-dimensional solution (C1 =.98; C2 =.91), though notably the three youngest samples exhibit unacceptable congruence with the total sample (C1 =.96; C2 =.84). WMS–III single-factor solutions were consistent between combined and age-based samples (C1 =.98). Respective WMS–IV and WMS–III pattern matrix loadings are nearly identical to those previously presented (see Tables 1 and 2; average difference between pattern matrix loadings =|.02|; maximum pattern matrix loading difference =|.05|). Thus, it does not appear that previously presented findings are attributable to use of moderately sized age-based samples.

Discussion

Clinical practice is in part guided by evaluation of whether revised psychometric measures represent an improvement over preexisting measures. In fact, the American Psychological Association (APA) Ethical Guidelines (American Psychological Association, 2002) requires providers to use current, updated versions of psychometric tests, as it is assumed that there is an incremental increase in validity and reliability with updated versions. The primary aim of this study was to compare the dimensional structures underlying the WMS–IV and WMS–III using identical methodology. A replicable and theoretically relevant multidimensional structure consisting of auditory learning/memory (Logical Memory and Verbal Paired Associates subtests) and visual attention/memory (Visual Reproduction, Designs, Spatial Addition, and Symbol Span subtests) was observed underlying the WMS–IV. This structure is preferable to the replicable, heterogeneous WMS–III dimension identified. Clinicians can have confidence that WMS–IV auditory and visual subtests "hang together" in a more coherent manner than auditory and visual subtests of the WMS–III.

A more coherent and multidimensional WMS–IV factor structure is likely related to the inclusion of new Designs, Spatial Addition, and Symbol Span subtests. Encouragingly, current results do not suggest that these subtests are verbally mediated. Including these new sub-tests as opposed to WMS–III Family Picture, Letter– Number Sequencing, and Digit Span subtests significantly increased the likelihood that a dimension reflecting visual attention/memory would be observed. There are clear clinical advantages for using these logical and efficient WMS–IV markers when evaluating memory along conceptually distinct dimensions of auditory and visual memory. These dimensions may be useful in clinical contexts to localize modality-specific memory functioning, though the relative value of these dimensions remains an important topic to explore further in diverse clinical samples.

An additional way to evaluate whether the WMS–IV factor structure represents an improvement from the WMS–III is considering how the results can be integrated with clinical theory. For example, one of the most comprehensive and complete theories of cognitive abilities, the Cattell–Horn–Carroll cognitive abilities framework (McGrew, 2009; McGrew & Flanagan, 1998), includes distinct constructs of auditory and visual memory. The WMS–IV factor structure can be easily integrated with this theoretical model, where as it would be significantly more challenging to do so with the WMS–III.

It is noteworthy that the replicable WMS–IV factor structure consisting of auditory learning/memory and visual attention/memory is inconsistent with WMS–IV indices (Auditory Memory, Visual Memory, and Visual Working Memory). Empirically supported factor retention procedures provided only weak support for retention of three WMS–IV factors across age-based normative samples. Though inconsistent across normative samples, the most common three-factor model included dimensions that reflected (a) visual attention/ memory, (b) Logical Memory subtests, and (c) Verbal Paired Associates subtests. It is somewhat unclear how WMS–IV Visual Memory and Visual Working

Memory indices are different from one another, though encouragingly, the respective indices did not contain construct irrelevant factors(in contrast, tasks included within the WMS–III visual memory indices are likely verbally mediated).

The failure to identify distinct WMS–IV visual memory and visual working memory dimensions could plausibly relate to the relatively small number of visual working memory marker variables included in normative datasets (Spatial Addition and Symbol Span subtests). This fact is important to recognize because at least three marker variables are typically needed to potentially identify a unique, corresponding factor (Velicer & Fava, 1998). Thus, failure to identify a "visual working memory" dimension might be the result of an inadequate number of visual working memory marker variables, rather than conceptual overlap with the construct of visual memory. Future research could explore this issue by conducting factor analysis on datasets that include WMS–IV subtests and additional measures of visual working memory. Regardless, determining whether Visual Working Memory and Visual Memory Indices have distinct clinical and physiological correlates will be especially useful in better understanding the diagnostic utility of new WMS–IV visually mediated subtests.

With respect to the WMS–III, multidimensional structures were inconsistent across age-based normative samples, and a replicable heterogeneous single factor solution has unclear clinical utility. Though inconsistent across samples, the most frequently observed two-dimensional WMS–III structure reflected general memory and facial memory. This finding is consistent with explicit concern expressed by some researchers that WMS–III visual memory tasks, Family Pictures and Faces, are different from one another (e.g., Millis et al., 1999; Wilde et al., 2003).The more verbally mediated visual memory task, Family Pictures, was reflected on a general memory dimension (that largely reflected auditory learning and memory), whereas immediate and delayed Faces subtests were uniquely reflected on a second dimension.

Researchers are discouraged from conducting CFA using current findings as an a priori specified model. A conceptually similar WMS–IV model consisting of visual and auditory memory constructs has previously been evaluated in the technical manual (Wechsler, 2009), and goodness-of-fit statistics were not meaningfully different from an a priori model consistent with the WMS– IV indices. Notably, those analyses included only delayed subtests. WMS–III CFA literature suggests that including all WMS–IV test variables (i.e., immediate and delayed subtests) will likely yield inadmissible parameter estimates because of high correlations between immediate and delayed measures.

WMS–III CFA studies have produced rich results and informed clinicians of the possibility that index scores may include construct irrelevant variance. Similar efforts to explore the psychometric properties of the WMS–IV are warranted and are likely to produce clinically relevant information. For instance, conducting CFA in clinical samples might inform whether the three-factor (Visual Memory, Visual Working Memory, Auditory Memory) or two-factor(Visual Memory, Auditory Memory) model described in the technical manual (Wechsler, 2009) is superior. Also, it is currently unknown whether alternative a priori models might more optimally describe the underlying structure of the WMS–IV, or how combined CFA analyses with the Wechsler Adult Intelligence Scale– Fourth Edition (WAIS–IV; Wechsler, 2008) might alter conceptualization of the revised measures.

Optimal replication of these findings would include applying identical methodology to clinical samples. There is conflicting evidence whether clinical and nonclinical samples should produce similar factor structures (e.g., see, Bowden, 2004; Delis et al., 2003; Wilde et al., 2003). Supporting the position that similar structures can be identified across diverse samples, Bowden and colleagues (2008) reported measurement equivalence of the Wechsler Adult Intelligence Scale– Third Edition (Wechsler, 1997a) and the WMS–III across normative and clinical samples (attention-deficit/ hyperactivity disorder; learning disorders).Encouragingly, there is a body of literature developing that highlights the importance of using empirically supported factor retention strategies such as PA and the MAP procedure. Similar dimensional structures have been found underlying psychological measures across normative and clinical samples when these guidelines are applied (Hoelzle & Meyer, 2009; O'Connor, 2002). It would be worthwhile to investigate whether these findings are relevant to neuropsychology. In other words, efforts to determine whether psychometric proprieties of neuropsychological measures are similar across diverse samples with localized or lateralized cerebral dysfunction would only improve clinical assessment.

In summary, in contrast to the replicable WMS–III single-factor solution, the underlying replicable WMS–IV factor structure is multidimensional and coherent and reflects important modality-specific constructs of auditory and visual memory. Findings support the WMS–IV as an improved, useful instrument to evaluate auditory and visual memory. Additional research is needed to evaluate the clinical utility of these dimensions and to identify how WMS–IV Visual Memory and Visual Working Memory indices are diagnostically relevant and unique from one another.

References

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, *57*, 1060–1073.

Barrett, P. T. (2005). Orthosim 2 (Version 2.0): Target-comparison matrix fitting [Computer software and manual]. Retrieved January 13, 2006, from
http://www.pbarrett.net/orthosim2.htm

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2001). Neuroscience: Exploring the brain (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Bowden, S. C. (2004). The role of factor analysis in construct validity: Is it a myth? Journal of the International Neuropsychological Society, 10, 1018–1019.

Bowden, S.C., Gregg, N., Bandalos, D.,Davis, M., Coleman,C., Holdnack, J. A., et al. (2008).Latent mean and covariance differences with measurement equivalence in college students with developmental difficulties versus the Wechsler Adult Intelligence Scale–III/Wechsler Memory Scale–III normative sample. Educational and Psychological Measurement, 68, 621–642.

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 standardized sample. Archives of Clinical Neuropsychology, 18, 629–641.

Carroll, J.B. (1993). Human cognitive abilities: A survey of factor analytic studies. New York: Cambridge.

Cattell,R.B. (1966).The scree test for the number of factors. Multivariate Behavioral Research, 1, 245–267.

Cattell, R. B. (1978). The scientific use of factor analysis. New York: Plenum.

Deary, I. J. (2000). Looking down on human intelligence: From psychometrics to the brain. Oxford, UK: Oxford University Press.

Delis, D.C., Jacobson, M.,Bondi, M.W., Hamilton, J. M.,& Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal and mixed clinical populations: Lessons from memory assessment. Journal of International Neuropsychological Society, 9, 936–946.

Frazier, T.W., & Youngstrom, E. A. (2007).Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? Intelligence, 35, 169–182.

Gleiβner, U., Helmstaedter, C., & Elger, C.E. (1998).Right hippocampal contribution to visual memory: A presurgical and postsurgical study in patients with temporal lobe epilepsy. Journal of Neurology, Neurosurgery, and Psychiatry, 65, 665– 669.

Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurement, 55, 377–393.
Goldberg, L. R., & Velicer, W. F. (2006). Principles of exploratory factor analysis. In S.Strack (Ed.), Differentiating normal and abnormal personality (2nd ed., pp. 209–237). New York: Springer.
Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hall.
Hatcher, L. (1994). A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. Cary, NC: SAS Institute.
Hoelzle, J.B., & Meyer, G.J. (2009).The invariant component structure of the Personality Assessment Inventory (PAI) Full Scales. Journal of Personality Assessment, 91, 175–186.
Jacobson, M.W., Delis, D.C., Hamilton, J. M.,Bondi, M.W., & Salmon, D.P. (2004).How do neuropsychologists define cognitive constructs? Further thoughts on limitations of factor analysis used with normal or mixed clinical populations. Journal of the International Neuropsychological Society, 10, 1020–1021.
Kaiser, H.F. (1960).The application of electric computers to factor analysis. Educational and Psychological Measurement, 20, 141–151.
Kim, J., & Mueller, C. W. (1978). Factor analysis: Statistical methods and practical issues. Newbury Park, CA: Sage Publications.
Kline, P. (1979). Psychometrics and psychology. London: Academic Press.
Lashley, K.S. (1950).In search of the engram. Symposia of the Society for Experimental Biology, 4, 454–482.
Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G.C. (1989). A regression equation for the parallel analysis criterion in principal components analysis: Mean and 95th percentile eigenvalues. Multivariate Behavioral Research, 24, 59–69.
Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. European Journal of Research Methods for the Behavioral and Social Sciences, 2, 57–64.
MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84–99.
McGrew, K.S.(1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 151–180). New York: Guilford.

McGrew, K.S.(2009).CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. Intelligence, 37, 1–10.

McGrew, K. S., & Flanagan, D. P. (1998). The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment. Boston: Allyn & Bacon.

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.

Milner, B. (1968). Visual recognition and recall after right temporal-lobe excision in man. Neuropsychologia, 6, 191–209.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavioral Research Methods, Instruments & Computers, 23, 396–402.

O'Connor, B.P. (2002).The search for dimensional structure differences between normality and abnormality: A statistical review of published data on personality and psychopathology. The Journal of Personality and Social Psychology, 83, 962–982.

Price, L.R., Tulsky, D.,Millis, S.,& Weiss, L. (2002).Redefining the factor structure of the Wechsler Memory Scale–III: Confirmatory factor analysis with cross-validation. Journal of Clinical and Experimental Neuropsychology, 24, 574–585.

Tulsky, D. S., Ivnik, R. J., Price, L. R., & Wilkins, C. (2003). Assessment of cognitive functioning with the WAIS–III and WMS–III: Development of a six-factor model. In D. S. Tulsky, D.H. Saklofske, G.J. Chelune, R. K. Heaton, R.J. Ivnik, R. Bornstein, A. Prifitera, & M.F. Ledbetter (Eds.), Clinical interpretation of the WAIS–III and WMS–III (pp. 147–179).Orlando, FL: Academic Press.

Tulsky, D. S., & Price, L. R. (2003). The joint WAIS–III and WMS–III factor structure: Development and cross-validation of a six-factor model of cognitive functioning. Psychological Assessment, 15, 149–162.

Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 31, 321–327.

Velicer, W.F., & Fava, J.L. (1998). Affects of variable and subject sampling on factor pattern recovery. Psychological Methods, 3, 231–251.

Wechsler, D. (1997a). Wechsler Adult Intelligence Scale—Third Edition (WAIS–III). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). Wechsler Memory Scale—Third Edition (WMS–III) administration and scoring manual. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV).San Antonio, TX: Pearson.

Wechsler, D. (2009). Wechsler Memory Scale—Fourth Edition (WMS–IV) technical and interpretive manual. San Antonio, TX: Pearson.
Wilde, N.J., Strauss, E., Chelune, G.J., Hermann, B.P., Hunter, M., Loring, D.W., et al. (2003).Confirmatory factor analysis of the WMS–III in patients with temporal lobe epilepsy. Psychological Assessment, 15, 56–63.
Zwick, W. R., & Velicer, W.F. (1982).Factors influencing for rules for determining the number of components to retain. Multivariate Behavioral Research, 17, 253–269.
Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99, 432–442.

Notes

1. The Cattell–Horn–Carroll cognitive abilities model does differentiate between broad constructs of "long-term storage and retrieval" and "short-term memory," though the latter construct is more consistent with the notion of working memory or attention in neuropsychology (McGrew, 1997).
2. WMS–IV and WMS–III oblique and orthogonal solutions were largely consistent. Orthogonal solutions can be obtained from J. Hoelzle upon request.
3. Inconsistent solutions may be obtained by contacting J. Hoelzle.
4. Congruency coefficients did not meaning fully improve when the Auditory Recognition Delayed score was excluded from analyses.

Acknowledgements

We are thankful to NCS Pearson, Inc. for making WMS–IV and WMS–III normative data available for these analyses. It was requested that any use of these normative data bear the following notices: Standardized data from the Wechsler Memory Scale, Third Edition (WMS–III). Copyright © 1997 NCS Person, Inc. Used with permission. All rights reserved. Standardized data from the Wechsler Memory Scale, Fourth Edition (WMS–IV). Copyright ©2009 NCS Person, Inc. Used with permission. All rights reserved. Address correspondence to James B. Hoelzle, Department of Psychology, Marquette University, PO Box 1881, Milwaukee, WI, 53201-1881, USA (Email: james.hoelzle@marquette.edu).

Appendix

Figure 1: Results for the minimum average partial procedure with Wechsler Memory Scale–Fourth Edition (WMS–IV) age-based normative data.

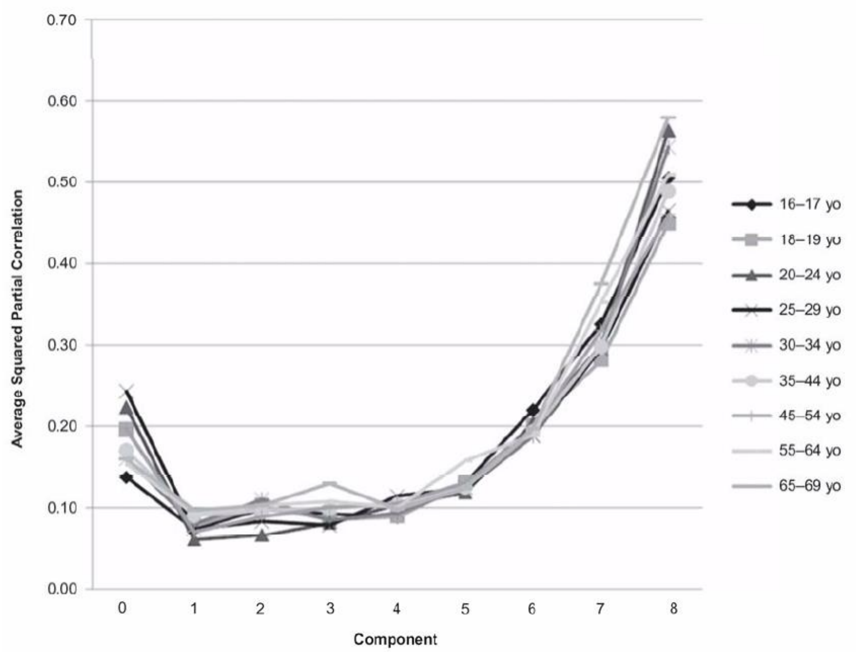

TABLE 1
Average WMS–IV pattern matrix loading across nine normative samples

	Component	
WMS–IV subtest	*1*	*2*
Logical Memory I	**.87**	−.01
Logical Memory II	**.89**	−.07
Verbal Paired Associates I	**.67**	.19
Verbal Paired Associates II	**.62**	.24
Designs I	−.07	**.86**
Designs II	−.08	**.80**
Visual Reproduction I	.11	**.69**
Visual Reproduction II	.13	**.61**
Spatial Addition	.06	**.63**
Symbol Span	.19	**.63**

Note. WMS–IV = Wechsler Memory Scale–Fourth Edition. Loadings ≥ .40 are in bold. Average correlation between components was .40.

Figure 2: Results for the minimum average partial procedure with Wechsler Memory Scale–Third Edition (WMS–III) age-based normative data.

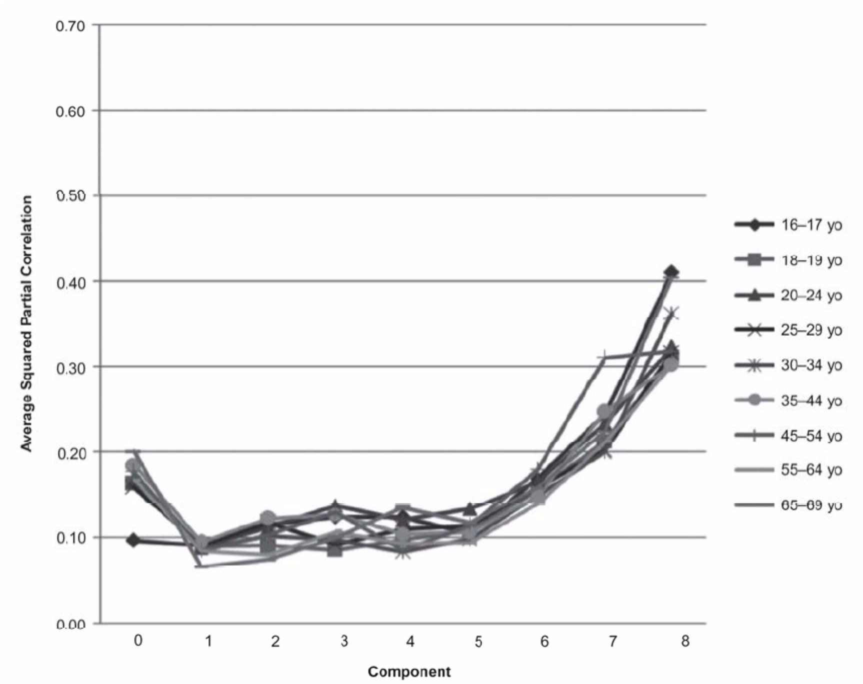

TABLE 2
Average WMS–III pattern matrix loading across nine normative samples

WMS–III subtest	*Component 1*
Logical Memory I	**.80**
Logical Memory II	**.82**
Auditory Recognition Delayed	**.73**
Verbal Paired Associates I	**.72**
Verbal Paired Associates II	**.68**
Family Pictures I	**.65**
Family Pictures II	**.68**
Letter–Number Sequencing	**.57**
Spatial Span	**.47**
Faces I	.36
Faces II	.39

Note. WMS–III = Wechsler Memory Scale–Third Edition. Loadings ≥ .40 are in bold.

