Perceptual and Motor Skills, 2003, 97, 1209-1218. © Perceptual and Motor Skills 2003

DIGIT SPAN AS A MEASURE OF EVERYDAY ATTENTION: A STUDY OF ECOLOGICAL VALIDITY[1]

GARY GROTH-MARNAT AND SONYA BAKER

Curtin University

Summary.—This study investigated the effectiveness of the WAIS–III Digit Span subtest to predict the everyday attention of 75 participants with heterogeneous neurological conditions who were administered the Digit Span subtest as well as the ecologically valid Test of Everyday Attention. In addition, the more visually oriented Picture Completion subtest along with the verbally loaded National Adult Reading Test were administered. Analysis indicated that, although Digit Span was a weak but statistically significant predictor of attentional ability (accounting for 12.7% of the unique variance), Picture Completion was a somewhat stronger predictor (accounting for 19% of the unique variance). The weak association of Digit Span and the Test of Everyday Attention, along with the finding that Picture Completion was a better predictor of performance on the Test of Everyday Attention, question the clinical utility of using Digit Span as a measure of everyday attention.

Digit span tasks have traditionally been considered to measure a client's attentional abilities (Lezak, 1995; Kaufman & Lichtenberger, 1999; Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Sattler, 2001). It has further been assumed that digit span tasks measure not only attention within the context of formal assessment, but also generalize to aspects of a client's everyday life. Given the recent upsurge in documenting that cognitive measures do indeed relate to a person's actual life (Acker, 1990; Wilson, 1993; Sbordone & Long, 1996; Groth-Marnat & Teal, 2000; Higginson, Arnett, & Voss, 2000), along with debates related to the wisdom of conducting subtest interpretation for tests such as the Wechsler intelligence and Wechsler memory scales (Lezak, 1988, 1995; McDermott, Fantuzo, & Glutting, 1990; Macmann & Barnett, 1997), it seems essential to investigate formally the extent to which a procedure such as Digit Span actually relates to everyday attention.

Performance on digit span tasks requires a person to respond effectively to and encode the digits and then accurately recall, sequence, and vocalize the information. If the person is given digits backwards, they must also perform the more difficult functions of holding the numbers longer and transforming them to a different sequence prior to restating them. The digits backwards procedure therefore requires greater mental control than digits forward in that a person is required to manipulate information retrieved from memory (Sohlberg & Mateer, 1989). An important feature seems to be

[1]Please send all correspondence to Gary Groth-Marnat, Hollister Ranch, Parcel 98, Gaviota, CA 93117 or e-mail (garygm@education.ucsb.edu).

encoding (for digits backwards) which is a more complex function than simple alertness and attention (Mirsky, Anthony, Duncan, Ahern, & Kellam, 1991). A somewhat different model considers that digit backwards has a "visual/auditory spanning" component in that, for digits backwards, subjects scan mental images of the digits before transforming them into their reverse order (Shum, McFarland, & Bain, 1990; Reynolds, 1997). Whereas digit span has usually been associated with working memory, its attentional component is usually given greater emphasis. This is partially supported in that even patients with amnesic brain injuries have performed quite well on digit span tasks (Butters & Cermak, 1980).

Interpretive sources usually suggest that high scores reflect "good auditory short-term memory and excellent attention" whereas low scores suggest "difficulty concentrating" (Lezak, 1995; Groth-Marnat, 1999, p. 176). Such interpretations imply that low scorers would be expected to have attentional difficulties significant enough to interfere with processing such information as instructions, phone numbers, or appointments. However, a number of criticisms have been leveled at such interpretive inferences. The test situation requires higher structure than more unstructured real life situations, and these real life contexts may result in more competing demands on the client (see Ponsford, 2000). In addition, attention is often considered to involve a complex set of abilities (visuospatial orienting, shifting of attention, alertness, vigilance, sustained attention, selective focused attention, working memory) which the relatively simple digit span test may not adequately or fully assess (Posner & Petersen, 1990). Finally, controversy has arisen related to the ability of a Wechsler subtest such as Digit Span to measure specific abilities due to it (and other Wechsler subtests) demanding too many heterogeneous abilities or skills (low subtest specificity) and not having sufficient reliability and validity (McDermott, *et al.*, 1990; Macmann & Barnett, 1997).

Assessing the extent the WAIS–III Digit Span subtest can actually assess relevant aspects of a client's everyday attention would allow clinicians to evaluate the extent to which they can rely on interpretations based on Digit Span. It was thus hypothesized that Digit Span would be a good to moderate predictor of scores on a test specifically designed to assess everyday aspects of a client's attention (Test of Everyday Attention). In contrast, a primarily visual-recognition test (WAIS–III Picture Completion) would have low to moderate associations with scores on the Test of Everyday Attention, and the purely verbal-recognition functions of the National Adult Reading Test further suggest scores would have little or no association with those on the Test of Everyday Attention.

Method

Participants

The sample was comprised of 75 participants (38 men and 37 women, *M* age = 65.1, *SD* = 16.9, range = 22–88) with heterogeneous neuropsychological complaints who were being seen at a neuropsychologically oriented outpatient clinic and a retirement home. These participants were likely to have a wide range of attentional difficulties. Exclusion from the study occurred if they suffered from visual or hearing deficits that would have prevented them from undertaking the tests or if they had less than a Year 8 education. Neuropsychological complaints involved cognitive slowing associated with aging (36), memory problems (14), head injury (6), Alzheimer's disease (6), stroke (5), Parkinson disease (2), ataxia (2), multiple sclerosis (2), neurotoxicity from spider venom (1), and brain tumor excision (1).

Materials

The Test of Everyday Attention (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996) is a test of attention comprised of eight subtests, each of which was designed to be ecologically valid (Crawford, Sommerville, & Robertson, 1997). For example, the Map Search subtest requires participants to search for symbols on a map in much the same way one might expect a tourist to do on a trip to a new city. In the Elevator Counting with Distraction subtest, participants are required to pretend that they are in an elevator where the floor indicator is not working such that they have to count a series of low tones and ignore high tones to establish on which floor they are. Elevator Counting was excluded based on findings suggesting an inadequate range given its low ceiling. A composite score for the remaining seven subtests ("Total Attention") was calculated by summing the age-scaled scores (range = 1–19; *M* = 10, *SD* = 3) and dividing by the number of subtests attempted.

Test-retest reliability coefficients for the subtests of the Test of Everyday Attention ranged from .59 to .90 (*M* = .75) for a control group and .41 to .90 (*M* = .78) for stroke patients. The Test of Everyday Attention has good face validity and significantly discriminated between stroke and control subjects, head-injured participants and controls, and between minimal and mild signs of Alzheimer's dementia. The Test of Everyday Attention demonstrated reasonable convergent validity with other reported measures of attention, with validity coefficients ranging from .42 to .63 (Robertson, *et al.*, 1996).

The Digit Span subtest from the Wechsler Adult Intelligence Scale–III is considered a test of immediate rote recall and attention and concentration which is sensitive to establishing and sustaining a focus of attention (Groth-Marnat, *et al.*, 2000). The average split-half reliability coefficient across all age groups for Digit Span was .90, with an average Standard Error of Measurement of .94. The test-retest stability coefficient (14–84 days, *M* = 34.6 days) averaged over age groups was .83 (Wechsler, 1997). It is a fair measure of general ability, with a *g* loading of .57 (Kaufman & Lichtenberger, 1999). Digit Span has the highest specificity of all WAIS–III subtests: .50/.10 (reliable unique variance/subtest error variance; Kaufman & Lichtenberger, 1999). Digit Span demonstrated moderate criterion validity when correlated with the Stanford-Binet–IV composite score ($r = .48$) and Stanford-Binet–IV short-term memory ($r = .52$; Wechsler, 1997). Shum, *et al.* (1990) investigated the construct validity of eight tests of attention, including Digit Span Forwards and Backwards, and reported that they loaded on the construct of immediate attention span taken from the current model of attention (Posner & Petersen, 1990).

Scores standardized for age were used to compute Digit Span. It was also considered necessary to calculate Digit Span Backwards and Forwards performance separately, given that these two components are dissimilar (Griffen & Heffernan, 1983; Reynolds, 1997). Age scaled scores are not available for Digit Span Backwards and Digit Span Forwards; therefore, the Cumulative Percentages table for Longest Digits Forwards and Backwards was used (WAIS–III manual) to compute a transformed percentile score that accounted for age (subtraction of the age percentage score from 100 producing Longest Backwards Transformed and Longest Forward Transformed).

The Picture Completion subtest from the Wechsler Adult Intelligence Scale–III is considered to measure distinguishing relevant from irrelevant information, visual alertness, and long-term visual memory (Groth-Marnat, 1999). The average split-half reliability coefficient for Picture Completion was .83, with an average Standard Error of Measurement of 1.25. The test-retest stability coefficient averaged over age groups for Picture Completion was .66 (Wechsler, 1997). It is a fair measure of general ability, with a *g* loading of .64 (Kaufman & Lichtenberger, 1999). Picture Completion shows ample specificity, .35/.17 (reliable unique variance/subtest error variance; Kaufman & Lichtenberger, 1999). Horn (1985) conceptualized Picture Completion as measuring Broad Visualisation. It demonstrates convergent validity with Block Design, which falls into the same category, with a mean correlation of .51 between the two subtests (range = .40–.61; Wechsler, 1997). Scores standardised for age were used in this study.

The National Adult Reading Test ("Reading" score) is a recognition-based vocabulary test in which the client must recognize phonetically irregular words that require pronunciation (Nelson, 1982). Vocabulary is considered to correlate with overall ability (Crawford, 1992; Lezak, 1995; Kaufman & Lichtenberger, 2002) which is supported by the test having a high correlation with general intelligence. The split-half reliability was .90, test-retest reliability .98, and interrater reliability .96 (Crawford, 1992; Crawford, *et al.*, 1997). It has a higher correlation with the WAIS–R Verbal IQ ($r = .85$) than the WAIS–R Performance IQ ($r = .33$; Crawford, 1992), demonstrating empirically its far greater association with verbal abilities.

Procedure

Participants were tested individually and data were collected at one session in a quiet room in their own home. Demographic details relating to age, sex, and nature of neurological condition were recorded before testing on each of the four measures. Each measure was administered according to the standard procedures outlined in the manual. The order of administration of the tests was counterbalanced to control for the possibility of participants' fatigue.

Results

Table 1 provides the means and standard deviations for all variables. A bivariate correlation matrix was generated to examine whether there were

TABLE 1
Descriptive Statistics For Age, Digit Span, Picture Completion, Test of Everyday Memory, and National Adult Reading Test

Variable	*M*	*SD*	Range
Age	65.1	16.9	22–88
Digit Span	10.7	3.2	3–23
Picture Completion	10.2	3.0	4–18
Test of Everyday Memory	9.1	2.3	3.70–13.90
National Adult Reading Test	107.1	6.4	90–120

significant linear relationships (see Table 2). Digit Span and its two components (Backwards and Forwards) had weak to moderate significant correlations with Total Attention, the composite score, with a range of .33 to .46. These correlations supported the hypothesis that Digit Span would predict scores on the Test of Everyday Attention. It should be noted, however, that even though Digit Span Forwards raw scores were slightly better ($r = .46$) at predicting Total Attention than Digit Span Backwards raw scores ($r = .39$), this difference disappeared when transformed scores were used (Longest Backwards Transformed $r = .33$ versus Longest Forwards Transformed $r = .36$). This is contrary to conceptualizations and suggestions that Digit Span Backwards might be a better predictor of attention than Digit Span Forwards since the backwards task is more difficult and complex (Griffen & Heffernan, 1993; Reynolds, 1997; Groth-Marnat, *et al.*, 2000). Surprisingly, Picture Completion exhibited a stronger correlation ($r = .51$) than Digit Span, which did not support the hypothesis that Picture Completion would predict Total Attention less well than Digit Span. As expected, the National Adult Reading Test ("Reading") provided nonsignificant correlations with the Total Attention score, which supported the hypothesis that the "Reading" measure would be a nonsignificant predictor of Test of Everyday Attention performance. Digit Span Forwards had a moderate significant correlation with Reading scores ($r = .40$). However, when age-corrected subtest scores were used to form a composite (Digits Forwards and Backwards) Digit Span score, the correlations became nonsignificant ($r = .20$, see Table 2). Specific correlations between the predictor variables and the subtests of the Test of Everyday Attention were not undertaken since power was insufficient.

TABLE 2

Bivariate Correlation Matrix For Digit Span and Hypothesized Variables ($N = 75$)

Test	1	2	3	4	5	6	7	8
1. TEA Total								
2. Digit Span	.36*							
3. Digit Span Backwards–raw score	.39*	.51*						
4. Longest backwards span transformed	.33*							
5. Digit Span Forwards-raw score	.46*	.64*	.42*					
6. Longest forwards span transformed	.36*							
7. Picture Completion	.51*	.20	.17		.11			
8. National Adult Reading Test	.17	.20	.24		.40*		.20	

Note.—Empty cells represent inapplicable correlations. *$p = .01$ (2 tailed). Note that only the .01 level was accepted as significant due to power considerations.

To assess further how well Digit Span (and Picture Completion and "Reading" scores) could predict performance on the Test of Everyday Attention, hierarchical multiple regression was performed. Age was not included in the regression since these were already controlled for by using age-corrected scores for all variables. An examination of the scatter plots of predicted values against residuals did not indicate strong evidence for violations of the regression assumptions of linearity, normality, and homoscedasticity (Tabachnick & Fidell, 1989). There were no influential cases as all of Cook's distance values were less than 1, and multicolinearity was not a problem as the tolerance values for the predictors were relatively high. It was found that Digit Span was a significant predictor of Total Attention ($R^2 = .127$, $p < .01$), accounting for 12.7% of the variance. Picture Completion was entered on Step 2, as it was expected to be the second best predictor of Test of Everyday Attention performance, and was a significant predictor (R^2 change = .19, $p < .01$), accounting for 19% of the variance in Total Attention after controlling for Digit Span. This is consistent with the bivariate correlations which also surprisingly indicated that Picture Completion was a better predictor of Total Attention than Digit Span. "Reading" (National Adult Reading Test) was entered on Step 3 as it was expected to be the poorest predictor of performance on the Test of Everyday Attention and was not a significant predictor (R^2 change = .01, $p > .05$). This indicates that "Reading" scores accounted for none of the unique variance in Total Attention after controlling for Digit Span and Picture Completion. The three predictors combined accounted for approximately 32% of the variance in Total Attention score ($R^2 = .32$), leaving 68% of the variance unexplained.

Discussion

The present study examined the assumption that the WAIS–III Digit Span subtest predicts everyday attention. The primary hypothesis was based on the long history of clinical usage of Digit Span as a measure of attention. The hypothesis that Digit Span would be a moderately high predictor of scores on an ecologically valid test of attention (Test of Everyday Attention) was given weak support in that Digit Span accounted for only 12.7% of the unique variance in the Test of Everyday Attention. The further hypothesis was that the WAIS–III Picture Completion subtest would predict Test of Everyday Attention scores less well, was not supported as results actually demonstrated a stronger relationship with Total Attention than Digit Span (19% of the variance was accounted for by Picture Completion). The final hypothesis, that performance on the National Adult Reading Test would not predict scores on the Test of Everyday Attention, was supported. Thus, Digit Span scores as a predictor on an ecologically valid test of attention was not as strong as expected. It is further conceded that, although statistical significance was attained on a number of aspects, this significance might not translate into clinical significance for generating diagnostic decisions about a person's everyday attention. This is most apparent considering that, when using Digit Span alone, 87.3% of the variance was left unaccounted for.

One potential limitation of the study involves the computation of a Total Attention score. As the Test of Everyday Attention does not provide a composite score, it was deemed necessary to compute a score that incorporated as many of the aspects of everyday attention as possible to gain an overall rating of everyday attention. The Total Attention score included a range of attentional behaviors including divided and sustained as well as auditory and visual attention. Other authors have used the Test of Everyday Attention in a similar way (Higginson, *et al.*, 2000); however, this may not reflect the intentions of the test designers. The Test of Everyday Attention is presented as providing information about particular areas of attention (auditory, visual, divided, sustained attention, etc.), and each subtest is considered an index of performance in that area. This information can be additionally viewed in light of normative data about subtest differences (Crawford, *et al.*, 1997). Had a larger sample been used, a more confident analysis might have been possible of each subtest's relationship with Digit Span. However, power was not sufficient in this study to analyze the number of variables a design like this would have generated.

In considering the above limitations, some conclusions can be made from the study. Firstly, given that Digit Span was not as strongly predictive of performance on the Test of Everyday Attention as expected, reliance on it as an accurate predictor of everyday attentional capacity may not be warranted. This is particularly highlighted given that Picture Completion was an equal or better predictor of Test of Everyday Attention performance. Alternatively, it might be possible that attention is a more important component of performance on Picture Completion than previously thought. Concentration and alertness to the environment are cited as potentially influential factors, rather than unique abilities measured by this subtest (Kaufman & Lichtenberger, 1999). Perhaps other WAIS–III subtests are equally capable of predicting Test of Everyday Attention performance or everyday attention. Researchers might consider this possibility and investigate the relationship between other subtests and Test of Everyday Attention performance. Further research might also use a larger sample to investigate relationships between Digit Span and the specific subtests of the Test of Everyday Attention. Secondly, although it has been suggested that the summation of Backwards and Forwards Digit Span may obscure clinically important information (Griffen & Heffernan, 1983; Banken, 1985; Reynolds, 1997; Groth-Marnat, *et al.*, 2000), there was no difference between the predictive effectiveness of Digit Span Backwards and Forwards on the Test of Everyday Attention.

Finally, it should be stressed that in actual clinical practice it would not be optimal to use a single subtest such as Digit Span to assess attention. Instead, additional tests, along with clinical interviews with the patient and his family would be crucial (Van Zomeren & Brouwer, 1994; Lemsky, 2000). Even attention tests such as the Test of Everyday Attention, designed to be ecologically valid, are still given in a formal assessment setting, and a number of researchers maintain that sampling of behavior in such a structured environment does not adequately sample attention (Ponsford, 2000). For example, it has been argued that attention over time is inadequately sampled by Digit Span and that some measure of continuous performance might be more diagnostically useful (Rapport, Webster, & Dutra, 1994). It has therefore been proposed that a systemic approach should be taken in measuring attention since social, occupational, and physical environmental factors are likely to affect a person's attentional behavior differentially (Heinrichs, 1990; Ponsford, 2000).

In summary, results suggest caution in using Digit Span as an accurate predictor of everyday attention. It further highlights the importance of assessing established psychological measures for their ecological validity rather than simply comparing measures with other tests, models, or theories of attention.

REFERENCES

ACKER, M. (1990) A review of the ecological validity of neuropsychological tests. In D. E.

Tupper & K. D. Cicerone (Eds.), *The neuropsychology of everyday life: assessment and basic competencies.* Boston, MA: Kluwer Academic. Pp. 19-55.

BANKEN, J. A. (1985) Clinical utility of considering digits forward and digits backward as separate components of the Wechsler Adult Intelligence Scale–Revised. *Journal of Clinical Psychology*, 41, 686-691.

BUTTERS, N., & CERMAK, L. S. (1980) *Alcoholic Korsakoff's syndromes.* New York: Academic Press.

CRAWFORD, J. R. (1992) Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford & D. M. Parker (Eds.), *A handbook of neuropsychological assessment.* Hove, UK: Erlbaum. Pp. 21-49.

CRAWFORD, J. R., SOMMERVILLE, J., & ROBERTSON, I. H. (1997) Assessing the reliability and abnormality of subtest differences on the Test of Everyday Attention. *British Journal of Clinical Psychology*, 36, 609-617.

GRIFFEN, P. T., & HEFFERNAN, A. (19893) Digit Span, forward and backward: separate and unequal components of the WAIS Digit Span. *Perceptual and Motor Skills*, 56, 335-338.

GROTH-MARNAT, G. (1999) *Handbook of psychological assessment.* (3rd ed., Rev.) New York: Wiley.

GROTH-MARNAT, G., GALLAGHER, R. E., HALE, J. B., & KAPLAN, E. (2000) Wechsler intelligence scales. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice.* New York: Wiley. Pp. 129-194.

GROTH-MARNAT, G., & TEAL, M. (2000) Block design as a measure of everyday spatial ability: a study of ecological validity. *Perceptual and Motor Skills*, 90, 522-526.

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., & VÖSS, W. D. (2000) The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Neuropsychology*, 15, 185-204.

HORN, J. L. (1985) Remodeling old models of intelligence. In B. Wolman (Ed.), *Handbook of intelligence.* New York: Wiley. Pp. 267-300.

KAUFMAN, A. S., & LICHTENBERGER, E. O. (1999) *Essentials of WAIS–III assessment.* New York: Wiley.

KAUFMAN, A. S., & LICHTENBERGER, E. O. (2002) *Assessing adolescent and adult intelligence.* (2nd ed.) Boston, MA: Allyn & Bacon.

LEMSKY, C. M. (2000) Neuropsychological assessment and treatment planning. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice.* New York: Wiley. Pp. 535-574.

LEZAK, M. D. (1988) IQ: RIP. *Journal of Experimental and Clinical Neuropsychology*, 10, 351-361.

LEZAK, M. D. (1995) *Neuropsychological assessment.* (3rd ed.) New York: Oxford Univer. Press.

MACMANN, G. M., & BARNETT, D. W. (1997) Myth of the master detective: reliability of interpretations for Kaufman's "Intelligent Testing" approach to the WISC–III. *School Psychology Quarterly*, 12, 197-234.

MCDERMOTT, P. A., FANTUZO, J. W., & GLUTTING, J. J. (1990) Just say no to subtest analysis: a critique on Wechsler theory and practice. (Conference on Intelligence Theories and Practice, Memphis, Tennessee) *Journal of Psychoeducational Assessment*, 8, 290-302.

MIRSKY, A. F., ANTHONY, B. J., DUNCAN, C. C., AHERN, M. B., & KELLAM, S. G. (1991) Analysis of the elements of attention: a neuropsychological approach. *Neuropsychology Review*, 2, 109-145.

NELSON, H. E. (1982) *The National Adult Reading Test (NART): test manual.* Windsor, Berks., UK: NFER-Nelson.

PONSFORD, J. L. (2000) Attention. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice.* New York: Wiley. Pp. 355-400.

POSNER, M. I., & PETERSEN, S. E. (1990) The attention system of the human brain. *Annual Review of Neurosciences*, 13, 25-42.

RAPPORT, C. J., WEBSTER, J. S., & DUTRA, R. L. (1994) Digit Span performance and unilateral neglect. *Neuropsychologia*, 32, 517-525.