Personality and Individual Differences 55 (2013) 915–920

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Assessing mental flexibility with a new word recognition test

Cynthia T. Matthew *, Steven E. Stemler

Wesleyan University, 217 High Street, Middletown, CT 06459, USA

ARTICLE INFO

Article history:
Received 2 March 2013
Received in revised form 7 July 2013
Accepted 17 July 2013
Available online 8 August 2013

Keywords:
Mental flexibility
Cognitive ability
Test development
Behavioral flexibility

ABSTRACT

This paper presents the development and preliminary evaluation of a new word recognition test (WRT) designed to measure individual differences in *mental flexibility*, defined as the ability to solve novel problems in unfamiliar settings. Conceptually designed to simulate problem solving in real world performance situations, the test was developed to recruit fluid and reproductive abilities and the interplay between convergent and divergent thinking. It is based on a framework that integrates and extends previous theoretical and methodological approaches to the study of cognitive ability and creative cognition. The WRT was administered with various cognitive ability and criterion measures to an undergraduate student sample ($n = 266$). Results provide preliminary evidence of construct validity. WRT scores correlated as expected with reference measures of cognitive ability, creative performance, and college performance (GPA). Regression analyses showed the WRT explained an additional 4.5% of variance in college performance over and above traditional cognitive ability measures that take up to five times as long to administer. Results suggest further study is warranted given the potential for its contribution to basic research and applied use.

1. Introduction

The capacity to respond effectively to novel problems and unfamiliar settings is essential to successful performance in a broad range of human activities. The challenge from conceptual and psychometric perspectives is how best to define and measure this capacity. Cognitive ability tests that assess flexible thinking are based on various conceptualizations of fluid intelligence and creative thinking and tend to be lengthy with small to moderate predictive power (Lang, Kersting, Hulsheger, & Lang, 2010; Schmidt & Hunter, 1998). Instruments that are better able to predict individual differences in flexible performance may have practical utility and theoretical value. This paper presents the development and preliminary evaluation of the word recognition test (WRT), a newly designed measure of mental flexibility based on a framework that integrates and extends previous theoretical and methodological approaches to the study of cognitive ability and creative cognition.

The ability to deal with novelty and to adapt ones thinking to new cognitive problems without relying extensively on an explicit base of declarative knowledge has been labeled alternatively *fluid intelligence* (Cattell, 1963, 1971), *analytic intelligence* (Carpenter, Just, & Shell, 1990), and *eductive ability* (Raven, 1952, 2009; Spearman, 1923, 1927.) Measures have traditionally relied on various types of pattern recognition tests (series, classification, matrices and conditions) most notably Raven's advanced progressive matrices (RAPM) (Raven, Raven, & Court, 2003). In contrast, the ability to acquire knowledge, conceptualized as *crystallized intelligence* (Cattell, 1963, 1971) and *reproductive ability* (Raven, 2009; Spearman, 1927) have relied on vocabulary tests like the Mill Hill (Raven, Raven, & Court, 1985).

The study of creative cognition, which is closely related to fluid ability, emphasizes the process of forming novel cognitive structures through unusual associations of elements, combining/reorganizing existing elements in knowledge structures, reframing, and perceiving or utilizing visual imagery (Getzels, 1975; Guilford, 1967; Mednick, 1962; Mumford & Gustafson, 2001).

Some specific cognitive abilities associated with creative cognition include *divergent thinking* or the capacity to generate a variety ("flexibility") and number ("fluency") of ideas (Guilford, 1950, 1967), and *convergent thinking* or the evaluative capacity to identify the best combination of ideas or knowledge elements to produce a "best-fit" or "correct" response (Cropley, 2006). Both divergent and convergent processes ultimately involve forming and modifying perceived patterns. Witkin (1975) field dependence–independence model of cognitive style has also been found to be closely related to fluid ability, and, by extension, creative cognition. The theory distinguishes global thinking (field dependence), the tendency to perceive things as they exist as a whole, from analytic thinking (field independence), and the tendency to impose structure on what is perceived. The embedded figures test (EFT: Witkin, 1950; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962), designed to measure cognitive style, is a figural pattern recognition

* Corresponding author. Tel.: +1 860 304 7152; fax: +1 860 685 2761.
E-mail addresses: cmatthew@wesleyan.edu, ctmatthew@creekrowpartners.com (C.T. Matthew), sstemler@wesleyan.edu (S.E. Stemler).

0191-8869/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.paid.2013.07.464

test, methodologically quite similar to the RPM. The perceptual theory of top-down and bottom-up information processing (Myers, 2012) may be a useful overarching framework to describe both cognitive processing concepts. Convergent and field dependent thinking can be considered more broadly top-down approaches, while divergent and field independent thinking, bottom-up.

Drawing from previous research (Matthew, Beckmann, & Sternberg, 2008; Matthew & Stemler, in press), *mental flexibility* is conceptualized here as the capacity to effectively apply acquired knowledge to new problems through a dynamic process of switching back and forth between convergent and divergent thinking. Traditional tests of fluid ability, made up of figural pattern recognition tasks, assess cognitive ability independent of declarative knowledge. Analysis of the RAPM reveals two distinctive processes: (1) figural, perceptual or gestalt (pattern recognition), and (2) analytic or analogical (capacity to induce various rules) (Carpenter et al., 1990; Mackintosh & Bennett, 2005). In order to better measure the capacity to apply knowledge to solve novel problems in unfamiliar real-world settings, a test design that simultaneously recruits fluid and crystallized abilities and the interplay of bottom-up and top-down information processing strategies seems warranted. Such a measure does not appear in extant literature. Toward this end the WRT was developed.

The WRT is similar to figural tests of pattern recognition commonly used to measure fluid ability and cognitive style; however, it relies on acquired knowledge in the verbal domain. Sentences are presented, in which known words of various lengths are rearranged. Identification of the correct words in each sentence requires simultaneously switching one's attention back and forth between letter, word and sentence to generate possible word alternatives (divergent/analytic/bottom-up) and correctly identify words in each sentence (convergent/global/top down). In this way the WRT is expected to engage alternating top-down/bottom-up processes and ultimately capture the dynamic inter-linkage of fluid and reproductive capacity (Spearman, 1923, 1927), thereby providing a more efficient measure of mental flexibility. Word recognition tests have been developed in the past to measure various conceptualized components of verbal intelligence (word comprehension – reproductive) and verbal fluency (word production – fluid) but not their interplay (Beauducel & Kersting, 2002; French, Ekstrom, & Price, 1963; Thurstone, 1938; Thurstone & Thurstone, 1941).

Based on the foregoing theoretical discussion, we expected that WRT should correlate with measures of general cognitive ability but not too highly. WRT should also correlate with measures of field dependence–independence, as the task engages global (word, sentence) and analytic (letter) cognitive processing. Finally we expect the WRT will correlate with and predict creative and academic performance over and above reference measures. Accordingly, we tested the following hypotheses to assess convergent/discriminant validity (H1), predictive validity (H2), and incremental validity (H3) of the WRT:

H1: WRT scores will correlate positively but not too highly with fluid, reproductive/crystallized ability and field dependence–independence test scores.

H2: WRT scores will correlate positively with creative performance and college performance (GPA).

H3: WRT scores will predict college performance (GPA) over and above traditional measures of fluid and crystallized/reproductive ability.

2. Methods and materials

2.1. Participants

Participants were recruited through fliers and e-mail announcements from three universities in Connecticut, USA during the 2007–2008 academic year. They were told that the purpose of the study was to explore 'mental flexibility' and "how we think outside the box" and were offered $30 for their participation.

Data was gathered from 299 undergraduate student volunteers (mean age = 20 years, SD = 2 years). Approximately 74% of participants were female and 25% were male. The vast majority of participants were native English-speaking (94%, $n = 281$); 6% were non-native English-speaking. In terms of ethnic background, 6% ($n = 19$) were African American, 6% ($n = 17$) were Asian American, 4% ($n = 12$) were Hispanic American, 77% ($n = 229$) were European American, and 7% ($n = 21$) reported 'other'. The average number of years of college completed by the participants was 2.4 years (SD = 1.2).

2.2. Procedure

Data collection was part of a larger study aimed at examining pattern recognition as a basic cognitive process that gives rise to mental flexibility across task domains and required developing new assessment instruments (Matthew & Stemler, in press). Participants engaged in a single, 3 h, paper and pencil testing session conducted at their home university. Testing sessions were broken up into three sections with two breaks; two versions were administered to counterbalance the effect of order on test performance. After participants provided informed consent, a battery of tests was administered by experimenters using a standardized script. Upon completion of the testing session, each participant was paid and provided with a debriefing handout.

2.3. Measures

2.3.1. Word recognition test (WRT; Matthew & Stemler, in press)

WRT design was inspired by jumbled word text circulated on the internet in 2003 as follows:

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe. (*According to a researcher (sic) at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place. The rest can be a total mess and you can still read it without problem. This is because the human mind does not read every letter by itself but the word as a whole*).

While the statement has nothing to do with research conducted at Cambridge University, in a review of relevant research on letter ordering and reading Davis (2003) notes elements of truth in this meme: people can recognize words with their letters rearranged, provided the first and last letter remain unchanged, and identify words of different sizes and degrees of rearrangement with variable degrees of difficulty.

The test developed for this preliminary study was aimed at the English-speaking college student population. Sentences developed by Davis were selected and modified based on their relevance to college student participants. New sentences were written by the senior author, reviewed for relevance by her college student assistant, and rearranged according to Davis's findings. A sample sentence of the WRT is shown in Figure 1.

The WRT requires participants to write down as many words as can be deciphered in a timed session. Various sentences were piloted with a small convenience sample of 15 participants to obtain sentences with variable reported difficulty and time required to complete. Four final sentences were selected in which word length ranged from 4 to 12 letters and the sentence length ratio (mixed up words/total words in sentence) ranged from 6/13

to 12/19. Four minutes was estimated to be a length of time that might avoid a ceiling or floor effect based on the pilot experiment. A scoring rubric was developed that assigned up to five points for correct words based on word length, in which longer words were assigned higher scores. For example, one point was assigned to correct words with 4–5 letters, and five points assigned to words with 12 letters. An aggregate score for the test was obtained by adding together the total score for correct rearranged words in each sentence. The reliability estimates for this test administration was Cronbach's $\alpha = .75$.

2.3.2. Raven's advanced progressive matrices (RAPM; Raven, Raven, & Court, 1998)

A short form consisting of 18 alternate items from Set II of this 35 item test of fluid intelligence/general intelligence was utilized. The prompts are visual patterns, each with a part missing, and respondents select the correct part to complete the pattern from options presented. Participants were given 20 min to complete the short form. Full-test split-half reliability coefficients vary between .83 and .87. The Cronbach's coefficient alpha estimate of reliability was $\alpha = .74$.

2.3.3. Mill–Hill vocabulary scale (MH; Raven et al., 1985)

A short form of the test of crystallized/reproductive ability containing 33 items (set A) was administered untimed. Participants are presented with a word and instructed to select, from among four options, the closest synonym to the word presented. Full test split-half reliability coefficients are .90 and test–retest reliabilities, .87–.95. Cronbach's alpha estimate of reliability for this test administration was $\alpha = .70$.

2.3.4. Group embedded figures test (GEFT; Witkin, Oltman, Raskin, & Karp, 1971, 2002)

This timed 25-item test of cognitive style measures global/analytic thinking. Participants must locate a previously seen simple figure embedded within a larger complex one. The reported reliability estimate for this measure was Spearman Brown coefficient = .87. The reliability estimate for this test administration was Cronbach's $\alpha = .90$.

2.3.5. Abbreviated Torrance test for adults (ATTA; Goff & Torrance, 2002)

This timed test of creative performance is designed to measure creative thinking ability. Participants are given 3 min to complete each of three activities, (one verbal and two figural). Responses were scored by scholastic testing service (STS) on four norm-referenced abilities and 15 criterion-referenced creativity indicators. The scores are combined to form a "creativity index". A reported reliability estimate for total creativity indicators is KR21 coefficient = .69. The reliability estimate for this test administration was Cronbach's $\alpha = .63$.

2.3.6. Scholastic assessment test (SAT; College Board)

Participants reported their SAT and American college test (ACT) scores in a background survey (described below). ACT scores were transformed to SAT equivalents using a conversion table and formula provided by The Princeton Review (2008), in which ACT total scores were assigned equivalent SAT total scores (verbal, math and writing) and then divided by a total score of 2400. SAT scores were similarly divided by total score.

Prompt: Rhesareecrs pfreer atsinsstas who are cscnteiounois and rbelliae with sontrg iertesnt in the flied.

Answer: Researchers prefer assistants who are conscientious and reliable with strong interest in the field.

Fig. 1. Sample word recognition test sentence.

2.3.7. Background survey

Participants were asked to report various demographic information including sex, ethnicity, financial aid status, academic approval for extended testing time, academic year, major, SAT scores, ACT scores, and college grade point average (GPA).

3. Results

3.1. Preliminary analyses

Data for two participants were removed because duplicate case numbers were assigned to their responses. Distributional properties were acceptable for all measures. The WRT data was somewhat negatively skewed, suggesting the test may have been a bit too easy. Females and males performed equally well. A univariate analysis of variance test of WRT scores by ethnicity showed a significant main effect (F (5, 285) = 2.27, $p < .05$, $\eta^2 = .039$). However, post hoc tests of ethnicity deviation contrast did not reach statistical significance.

Because the WRT depends on knowledge of the English language and is timed, independent samples t-tests were undertaken to assess mean differences in test performance in the following two subsamples: (1) native English speaking as compared to non-native English speaking, and (2) no academic approval for extended testing time due to special learning needs as compared to those with approval. In both cases there were significant mean differences; mean scores for English speaking participants were significantly higher than for non-English speaking participants (t (2, 284) = 2.14, $p < .03$, $d = .25$); mean scores for students who are not eligible for extended testing time were significantly higher than for those who were eligible for extra testing time (t (2, 281) = 3.06, $p < .01$, $d = .36$). Accordingly test analyses were conducted on the participant subsample ($n = 266$) who had indicated on the demographic survey that they were English speaking and did not require extended testing time. Descriptive statistics of measures are displayed in Tables 1a and b.

3.2. Data analyses

Sample sizes in analyses vary because of missing data due to skipped tests or unreported SAT subtest scores. Hypothesis tests were conducted utilizing correlation analyses and hierarchical regression analyses.[1]

3.3. Hypothesis tests of validity

3.3.1. Convergent/discriminant

Correlations among all measures are displayed in Table 2.

Results of the correlation analysis presented in Table 2 support construct validity expressed in H1,which states WRT scores will correlate positively but not too highly with fluid, reproductive/crystallized ability and field dependence–independence test scores.

WRT scores correlate positively with RAPM scores (fluid ability), SAT scores (crystallized ability), GEFT scores (field-dependence/independence), and MH scores (reproductive ability).

[1] We did not obtain valid results with confirmatory SEM analyses due to identification and sample size issues.

Table 1a
Descriptive statistics. (a) Test means and standard deviations.

Variable	N	Mean	SE	SD
WRT	252	67.70	0.63	10.06
RAPM	252	11.69	0.20	3.13
MH	252	17.75	0.21	3.35
GEFT	253	11.64	0.31	5.01
ATTA-CI	222	74.03	0.62	9.26
SAT/ACT	211	0.76	0.01	0.10
SAT V	146	606.27	7.50	90.57
SAT M	156	614.46	7.15	89.25
SAT W	93	615.16	11.63	107.65
GPA	215	3.31	0.028	0.41

Note: Analyses conducted on final subsample of $n = 266$. Sample sizes vary due to unsystematic missing data.

Table 1b
(b) Independent sample tests of mean differences: Native vs. non-native Eng.; Approval vs. no-approval extended test time.

WRT	*T*	DF	Sig. (2-tailed)
Eng language	2.14	284	0.034
X-time	3.06	281	0.002

Note: Analyses conducted on full sample. Sample sizes differ due to unsystematic missing data.

3.3.2. Predictive and incremental

The results presented in Table 2 provide support the predictive validity expressed in H2, which states WRT scores will correlate positively with creative performance. Support was demonstrated for the incremental validity prediction specified in H3, which states WRT scores will predict college performance over and above fluid and crystallized/reproductive ability test scores. Results of a hierarchical regression analysis with GPA as the dependent variable on predictor variables RAPM, entered first, MH entered second, and WRT entered last were significant (F (3, 196) = 7.155, $p = .001$, $R^2 = .099$, $f^2 = .11$). However, MH and WRT variables were the only significant contributors to explained variance upon the addition of WRT in the final step; the RAPM variable was no longer statistically significant. Consistent with H3, WRT made a significant contribution to explained variance over and above RAPM and MH ($F\Delta$ (1, 191) = 9.699, $p = .002$, $R^2\Delta = .045$, $f^2 = .045$). The WRT partial correlation in the third step of the model suggests reasonable contribution in support of incremental validity (Hunsley & Meyer, 2003). Model and coefficient summaries of regression results detailed in Tables 3a and b.

For exploratory purposes, a second hierarchical regression analysis was undertaken utilizing SAT as a predictor variable in place of MH, as an alternative measure of crystallized intelligence. Results of a hierarchical regression analysis with GPA as the dependent variable on predictor variables RAPM, entered first, SAT entered second, and WRT entered last approached but did not reach the accepted level for statistical significance (F (3, 179) = 2.241, $p = .085$, $R^2 = .051$); however, the WRT was the only predictor in the regression that approached statistical significance ($\beta = .006$; t (1, 179) = 1.815, $p = .07$, $R^2\Delta = .018$).

4. Discussion

Results provides preliminary evidence supporting the criterion-related, predictive, and incremental validity of the WRT. As expected, WRT scores correlated positively with commonly accepted reference measures of fluid intelligence (RAPM: Raven's advanced progressive matrices), crystallized intelligence (SAT), reproductive ability (MH: Mill–Hill), and field dependence–independence (GEFT: group embedded figures test.) Notably, correlation coefficients with college performance (GPA) exceeded those with the other reference measures; indeed, the WRT explained an additional 4.5% variance in college performance over and above RAPM and MH taken together.

WRT scores correlated modestly with creative performance (ATTA: abbreviated Torrance test of creative thinking) – more highly than cognitive ability reference measures. The low correlations may be attributable to low reliability of the creative thinking test. A stronger correlation was expected given fluid intelligence is among the factors theoretically associated with creative performance. However, the correlation is consistent with Torrance (1974) reported correlations using full-length traditional measures (i.e., median correlation of .06 with figural and .21 with verbal measures).

WRT scores were moderately correlated with SAT across subtests (verbal, math and writing) as expected, suggesting criterion-related validity with academic knowledge and skills. However, SAT correlation coefficients were notably higher with the RAPM and MH tests. This may provide clues as to the differential validity of the WRT. The SAT involves solving well-defined problems that require application of academic knowledge and skills. The cognitive processes measured by the RAPM and MH tests may be well-matched to solving such problems.

4.1. Limitations

There are two types of methodological limitations to this study. First, in regard to measures: (1) abbreviated forms of the RAPM and MH were utilized. Although reliability estimates, which were comparable with other measures in the study, should assure reasonably valid results, this should be confirmed in future testing; (2)

Table 2
Correlations among study measures.

Measures	1	2	3	4	5	6	7	8	9	10
1. WRT	**.75**									
2. RAPM	.15*	**.74**								
3. MH	.27**	.32**	**.70**							
4. GEFT	.17*	.49**	.38**	**.90**						
5. ATTA-CI	.13*	.10	.03	.11	**.63**					
6. SAT/ACT	.26**	.40**	.59**	.42**	.14	–				
7. SAT V	.22**	.27**	.56**	.22**	.09	.88**	–			
8. SAT M	.27**	.40**	.31**	.37**	.19*	.84**	.59**	–		
9. SAT W	.31**	.37**	.50**	.38**	.23*	.90**	.83**	.71**	–	
10. GPA	.26**	.08	.21**	.13*	.19**	.19**	.17	.05	.31*	–

Note. Reliability estimates of internal consistency for each measure are indicated in bold in the diagonal.
* $p < .05$.
** $p < .01$, two-tailed.

Table 3a
Summary of hierarchical regression analysis for variables predicting GPA. (a) Model summary.

Model	R	R^2	Change statistics	
			$R^2\Delta$	$F\Delta$
1	.085	.007	.007	1.44ns
2	.233	.054	.047	9.76***
3	.314	.099	.045	9.70***

*** $p < .00$, two-tailed.

Table 3b
(b) Coefficient summary.

Predictor variable	GPA				
	B	SE	t	$R^2\Delta$	Partial R
Step 1:					
RAPM	.12	.010	1.202	.007	.085
Step 2:					
RAPM	.002	.010	.235	.007	.017
MH	.030	.010	3.124**	.047**	.217
Step 3:					
RAPM	.000	.010	.022	.007	.002
MH	.023	.010	2.406*	.047**	.169
WRT	.009	.003	3.114**	.045**	.217

Note. Unstandardized regression coefficients are reported. ***$p < .00$, two-tailed.
* $p < .05$.
** $p < .01$.

SAT/ACT and GPA scores were self reported. Although a meta-analysis by Kuncel, Crede, and Thomas (2005) found a mean correlation of .90 between self-report and actual college GPA, future testing would ideally utilize SAT and GPA scores reported directly from educational institutions. Second, the data did not meet the conditions for latent structural equation model confirmatory factor analysis, a better and more comprehensive technique for assessing construct validity.

4.2. Future research

Next steps should be aimed at further construct validation and psychometric development. Additional items could be added to the current test designed for English-speaking college students to evaluate optimal test length and timing. To examine the potential for generalizing the test to other populations, alternate forms could be developed and evaluated with appropriate criterion-measures. Construct validity needs to be further assessed to include other convergent measures such as divergent thinking, discriminant measures such as cognitive and behavioral rigidity, and personality correlates such as openness to experience, traditionalism, and conscientiousness. In addition, future studies should focus on investigating the potential trainability and adverse impact of the test.

4.3. Implications

Overall, the WRT introduces many advantages to the measurement of mental flexibility. From a theoretical perspective, the WRT supports the definition of mental flexibility as the capacity to switch back and forth between divergent and convergent thought processes. Pragmatically speaking, the WRT is short (i.e., four minutes) and engaging, which potentially can reduce motivational differences among test takers and maximizes efficiency of administration. Test construction is quite simple; by using population-relevant words, the test can be readily adapted to different occupational, educational and cultural groups in different languages and alternate forms easily developed.

Existing cognitive ability tests arguably leave substantial unexplained variance (estimated between 50% and 89%) in predicting educational and occupational performance (Lang et al., 2010; Schmidt & Hunter, 1998). The WRT potentially may emerge a stronger predictor of real world performance than existing measures.

5. Conclusion

A potentially useful model for assessing individual differences in mental flexibility may be one that effectively recruits fluid and reproductive abilities and the interplay between convergent and divergent thinking. Should further development of the WRT continue to corroborate its construct validity; it may have the potential to enhance our understanding of individual differences in mental flexibility and provide a useful instrument that predicts behavioral flexibility in applied settings.

Acknowledgement

This study was part of a larger research project sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences (Grant #W91WAW-07-C-00).

References

Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin model of intelligence structure (BIS). *European Journal of Psychological Assessment, 18*, 97–112.
Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review, 97*(3), 404–431.
Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.
Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.
Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal, 18*(3), 391–404.
Davis, M. (2003). Problems of reading jumbled text. MRC Cognition and Brain Sciences Unit, http://www.mrc-cbu.cam.ac.uk/people/matt.davis/Cmabrigde/.
French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.
Getzels, J. W. (1975). Problem finding and the inventiveness of solutions. *Journal of Creative Behavior, 9*, 12–18.
Goff, K., & Torrance, E. P. (2002). *Abbreviated Torrance test for adults*. Illinois: Scholastic Testing Service.
Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444–454.
Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*(4), 446–455.
Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63–82.
Lang, J. W. B., Kersting, M., Hulsheger, U., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology, 63*, 595–640.
Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's matrices measure? An analysis in terms of sex differences. *Intelligence, 33*, 663–674.
Matthew, C. T., Beckmann, J., & Sternberg, R. J. (2008). Development of a test battery to assess mental flexibility based on Sternberg's theory of successful intelligence, (Technical Report #1222). *U.S. Army Research Institute for the Behavioral and Social Sciences*. Alexandria, VA.
Matthew, C. T., & Stemler, S.E. (in press). *Exploring pattern recognition as a predictor of mental flexibility*. Final report submitted to the U.S. Army Research Institute for the Behavioral and Social Sciences for Project Award #W91WAW-07-C-00.
Mednick, S. A. (1962). The associative basis for the creative process. *Psychological Review, 69*, 200–232.
Mumford, M. D., & Gustafson, S. B. (2001). Creative thought: cognition and problem solving in dynamic systems. In M. Runco (Ed.), *Creativity research handbook*. Cresskill, NY: Hampton.
Myers, D. (2012). *Psychology* (10th ed.). New York: Worth.
Raven, J. C. (1952). *Human nature: Its development, variations and assessment*. London: H.K. Lewis & Co., LTD..
Raven, J. (2009). The Raven progressive matrices and measuring aptitude constructs. *The International Journal of Educational and Psychological Assessment, 2*, 2–38.

Raven, J. C., Raven, J., & Court, J. H. (1985). *Mill-Hill vocabulary scale, junior multiple choice*. New York: The Psychological Corporation, Harcourt Brace Jovanovich, Inc..
Raven, J. C., Raven, J., & Court, J. H. (1998). *Manual for Raven's advanced progressive matrices*. Oxford, England: Oxford Psychologists Press.
Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.
Spearman, C. (1923). *The nature of intelligence and the principles of cognition*. London: Macmillan.
Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies on intelligence*. Chicago: University of Chicago Press.
Torrance, E. P. (1974). *Torrance tests of creative thinking*. Lexington, MA: Personnel Press.
Witkin, H. A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality, 19*, 1–15.
Witkin, H. A. (1975). Some implications of research on cognitive style for problems of education. In J. M. Whitehead (Ed.), *Personality and learning*. London: Hodder and Stoughton.
Witkin, H. A., Dyk, R., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation*. New York: John Wiley & Sons.
Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the embedded figures test*. Palo Alto, CA: Consulting Psychologists Press, Inc..
Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (2002). *Group embedded figures test manual*. Menlo Park, CA: Mindgarden, Inc..