Development and Validation of a Persian Verbal Analogies Test Using the Rasch Model

Mona Tabatabaee Yazdi

Development and Validation of a Persian Verbal Analogies Test Using the Rasch Model

Mona Tabatabaee Yazdi
Islamic Azad University

This study examines the validity of a 49-item verbal analogies test by the application of the Rasch model and concurrent validation procedure. The test was translated and modified by the researcher to be used for Iranian students. The test was administered to 186 Iranian undergraduates. The results indicated that the Rasch model fit the test after deleting eight items from the scale. Concurrent validation demonstrated that the test correlates significantly with other verbal and nonverbal fluid reasoning measures, and suggests that the instrument is potentially valid and useful as a measure of verbal analogical reasoning in the Persian language.

An analogy is a juxtaposition of two objects which are deemed to have similarities. Analogical reasoning (AR) rests on these similarities to draw further conclusions, thus, AR is defined as the ability to perceive and use these similarities. AR is an important aspect of human thought and has played a fundamental role in scientific, philosophical, and legal discoveries (Bartha, 2016). For some researchers AR is the essence of intelligence or the cornerstone of human cognition (Esher, Raven, & Earl, 1942; Spearman, 1927). The *Raven's matrices*, a very well-recognized test of intelligence, is based on analogies and designed to "[measure] capacity to form comparisons, reason by analogy, and develop a logical method of thinking" (Raven, 1938).

Analogical reasoning refers to the ability to perceive relationships between different phenomena (Gentner, 1983). It is one of the main cognitive skills that distinguishes humans from other species. Verbal reasoning tasks require applicants to comprehend, analyze, evaluate, and make inferences from concepts which are represented as words. Analogical reasoning contributes to creativity (Dunbar, 1997; Sternberg, 1988), fluid intelligence (Cattell, 1971; Gentner, 2010), and human innovation (Markman & Wood, 2009), and provides a way to measure important cognitive skills that reflect comprehension skills (Kuncel, Hezlett, & Ones, 2004). Besides, analogical reasoning has an important role in psychological theories of intelligence, cognition, and Piaget's (1949, 1950) theory of intelligence (Sternberg & Nigro, 1980).

Author info: Correspondence should be sent to: Mona Tabatabaee Yazdi, English Department, Islamic Azad University, Ostad Yusofi St. 91871-Mashhad, Iran. Email: tabatabaee.mona@gmail.com
North American Journal of Psychology, 2017, Vol. 19, No. 3, 561-572.

In view of that, Kuncel, Hezlett, and Ones (2001, 2004) asserted that analytical reasoning abilities are not only critical to success in college and graduate school, but also in the workplace. They (2004) argued that analogical reasoning is a measure of *g*, the general intelligence factor. They also reported that verbal analogies tests, such as the *Miller Analogies Test* (MAT), is a valid predictor of performance in both academic and work sites, mostly when compared with the *Graduate Record Examination* (GRE). Therefore, it is possible that students with poor analogical reasoning skills may miss a number of key concepts in their books and classes.

Furthermore, cognitive psychologists argued that thinking analogically provides students with practical benefits in problem solving and making explanations (Gentner, Holyoak, & Kokinov, 2001; Holyoak & Thagard, 1996). It is also helpful in finding an effective way to assess verbal reasoning, inferential ability, analytical intelligence (Kuncel, Hezlett, & Ones, 2004; Sternberg, 1977, 1988), and the acquisition of knowledge and skills (Kuncel, Hezlett, & Ones, 2004). Analogical reasoning also is used by scientists to find new solutions for unpredicted issues in their research (Dunbar & Blanchette, 2001).

At least two studies have reported that discovering underlying relationships between words that seem different but have something in common, is an effective way to promote innovation (Holyoak & Thagard, 1996; Schunn, Paulus, Cagan, & Wood, 2006). Moreover, helping students comprehend the semantics of the target language is one of the first important skills that an analogy provides students, by not only recognizing each word in the analogy, but more importantly the relationships between the words. Helping students to find out the semantic relations between the words of the target language, which could positively influence their decision-making skill, seems to be one of the direct responsibilities of teachers. This is the reason why the learning of verbal analogies is a crucial, more personalized and humanistic trend in language teaching (Plaister, 1981).

In this regard, Sternberg (1985) proposed that all of the information-processing elements (i.e. encoding, inference, mapping, application, comparison, & justification) used in inductive reasoning are essential for solving analogy problems. Studies concerning analogical reasoning showed that students with greater executive-function skills (i.e. skills such as planning, observing, task switching, and controlling attention (Diamond, 2002; Stuss, 2007)) and vocabulary knowledge in elementary school performed better on a verbal analogies tasks (Richland & Burchinal, 2013). Moreover, Ignoffo (1980) discussed the use of vocabulary-building analogies with English native speakers in a reading course, and stated that in working the analogy, the student is forced to attempt many decoding procedures including articulation, problem solving, and thinking.

Furthermore, Cornoldi, De Beni, and Passaglia (1996) concluded that skilled comprehenders have significantly better performance than poor comprehenders on vocabulary knowledge and verbal reasoning tests. More recently Tabatabaee Yazdi and Baghaei (in press) reported that verbal analogical reasoning is a valid predictor of reading comprehension in English as a foreign language. Thus, low reading comprehension in the foreign language may be because of insufficient development of the verbal reasoning abilities.

The present study seeks to validate a 49-item verbal analogies test using the Rasch model, which is a psychometric model for analyzing measurement instruments data, and has been used widely for examining questionnaires and construct validity in the social sciences (Baghaei, 2008). Moreover, to provide external validity evidence for the test and to propose that the test is potentially valid, a concurrent validation procedure was used by estimating the correlation of the test with a number of criterion measures.

METHOD

Measures

Verbal analogies. I constructed 49 four-option multiple choice verbal analogy items in the form of ‘A is to B as C is to D.’ Some of the items were translated and modified by the researcher from free verbal analogies practice tests available on the McGraw Hill website (http://www.mheducation.com/) and some were constructed by the researcher.

The test was translated into Persian through backward and forward procedures. That is, an expert back-translated the Persian version into English. Back-translation was done to confirm the accuracy of the translation. Next the Persian translations of some items were revised based on careful examination of the English back-translations and the original English items. Lastly, it was checked again by another expert for translation accuracy and some minor changes were implemented in the final version of the test. An English item follows:

Mason is to Wall as….

1. Artist is to Easel	2. Fisherman is to Trout
3. Author is to Book	4. Sculptor is to Mallet

Persian adaptation of Baddeley’s three-minute Grammatical Reasoning Test. Baddeley’s *Grammatical Reasoning Test* (1968) which is a test of fluid reasoning *(Gf)*, is frequently used in research on cognitive abilities. Baghaei, Khoshdel, and Tabatabaee Yazdi (2017) adapted the test to apply in the Persian language by using a different pair of verbs and geometrical shapes instead of letters. The sixty eight items of the test were classified under four classifications of ‘Affirmative Active,’ ‘Affirmative Passive,’ ‘Negative Active,’ and ‘Negative Passive.’

The test was shown to have high retest (.76) and Cronbach’s alpha reliability estimates (.91), had an excellent fit to a one-factor confirmatory factor model, and correlated acceptably with other measures of fluid intelligence. The mean of the test for the sample used in the study was 26.36 (*SD*=10.30).

Raven’s Advanced Progressive Matrices. The short form of *Raven’s Advanced Progressive Matrices* (APM; Raven, Court, & Raven, 1977) was used to measure fluid intelligence. The 12 items were selected by Arthur and Day (1994) on the basis of rigorous psychometric criteria from the 36-item APM with the aim of reducing administration time. The Cronbach’s alpha reliability of APM scores was .70 in the present study. Moreover, the study reported the mean and standard deviation of 6.01 and 2.51, respectively.

Pearson reading comprehension test. To measure participants L2 reading comprehension, the researcher used the reading comprehension section of one of the official past papers (July, 2011) of the *Pearson Test of English General*. The test consists of 24 three-option multiple-choice items based on four passages with lengths of 279, 299, 354, and 356 words. The study reported the Cronbach’s alpha reliability of .71 for the reading scores, and the mean and standard deviation of 13.07 and 3.97, respectively.

Participants

One hundred and eighty six participants (67.7% females; mean age = 26.39 years, *SD* = 6.18) from three different universities in Mashhad, Iran participated in this study. The students were undergraduates from different fields and social sciences and humanities. Their native language was Persian with English as a foreign language. The measures were administrated in regular class time to eight intact classes. Participation was voluntary and participants were provided with profiles of their cognitive abilities as a compensation for their cooperation. The research was approved by the ethics committees of the universities.

RESULTS

The data were analyzed using eRm package (Mair, Hatzinger, & Maier, 2016) in R version 3.3.1. (R Core Development Team, 2016). The Rasch model (Rasch 1960/1980) has been used extensively for analyzing questionnaires and construct validity in the social sciences (Baghaei, 2008). Fit of the data to a latent trait model, like the Rasch model, is evidence that a construct is underlying the covariances among the items and causes the item responses; hence the test is valid (Baghaei & Tabatabaee Yazi, 2016; Borsboom, 2008).

In this study the Rasch unidimensional model was applied to validate a 49-item verbal analogies test for application in the Iranian context. Andersen's (1973) likelihood ratio (LR) test with the median of raw scores as a partitioning criterion showed that the test with 49 items does not fit the Rasch model: $\chi 2$=161.487, *df*=48, *p*<.00. The Andersen's chi square by degrees of freedom (Baghaei, Yanagida, & Heene, 2017a; Baghaei, Yanagida, & Heene, 2017b) was 3.64, which was larger than the maximum cut off of 1.40 set by Baghaei et al. (2017b) for a test with 50 items.

The graphical model check revealed that 8 items lie far from the 45 degree line, thus not fitting the Rasch model. Therefore, these eight items which are misfitting items (Kubinger, 2005) were deleted. After deleting the eight misfitting items the Rasch model was fitted again. Andersen's LR test with the median of scores as a partitioning criterion showed that the 41 remaining items did fit the Rasch model: $\chi 2$= 57.40, *df =40*, *p* = 0.037. The Andersen's chi square by degrees of freedom was 1.43 which is smaller than the criteria set by Baghaei et al. (2017b) for a test with 40 items. Principal components analysis of standardized residuals, which is a method of overall fit and unidimensionality (Baghaei & Cassady, 2014; Linacre, 2009), showed that the first contrast in the residuals had a strength of 2.4 which is lower than the cut-off value of 2.5, hence, supporting unidimensionality (Linacre, 2009).

Table 1 shows the difficulty of the items (under the heading of "Estimate"), and their standard errors, and the fit indices for each item. Infit indices relate to unpredicted patterns of observations by persons on items that are unfairly targeted on them.

The evaluation of the items showed an item difficulty range of −1.37 to 1.59 logits, with obvious gaps in the item difficulty hierarchy signifying the fact that some part of the construct domain has not been covered by the test (Baghaei, 2008). Person estimates ranged from −1.51 to 4.91, with a separation reliability of .89. In Rasch measurement the person separation index, which is a summary of the genuine separation as a ratio to separation, including measurement error, is used instead of reliability indices. Separation reliability indicates how well the person parameters are discriminated on the measured variable. A high separation reliability index shows that there is a strong possibility that persons with high ability estimates have higher ability estimates than persons/items with low estimates (Linacre, 2009).

Moreover, infit mean square statistics were considered to check the quality of the items and their contribution to the measurement scale. They specify how well the items signify the single underlying construct intended to be measured. "High infit mean squares indicate that the items are mis-performing for the people on whom the items are targeted. This is a bigger threat to validity, but more difficult to diagnose than high outfit" (Linacre, 2009, p. 596). Outfit values are sensitive to outliers and can deviate from the expected values by a couple of lucky guesses or careless misses; that is why infit values are more reliable in examining item quality (Linacre, 2009, p. 596). Infit mean-square values in the range of 0.70 to 1.30 are considered (Linacre, 1999; Wright & Linacre, 1994). The ideal value is 1. Values greater than 1.3 indicate deviant response patterns that alter the measurement and are considered a sign of multidimensionality or construct irrelevant variance (Baghaei, 2008); values smaller than 0.70 are nonthreatening. They provide researchers with redundancy and duplication of information.

Figure 1 is the item-person map (Wright map) of the data. A person-item map shows the location of item parameters as well as the distribution of person parameters. It is useful to compare the range and position of the item measure distribution (lower panel) to the range and position of the person measure distribution (upper panel). Items should ideally be located along the whole scale to meaningfully measure the 'ability' of all persons. (Bond & Fox, 2007).

Concurrent Validation

To provide external validity evidence for the Persian verbal analogies test in the present study a number of criterion measures were administered to subsamples of the participants and the coefficient of correlation between the criteria and the verbal analogies test was computed. The *Persian Verbal Analogies Test* had a correlation of r=.53 (n=84, p<.01) with the short form of the Raven's APM, a correlation of r=.50 (n=46, p<.01) with Baddeley's G*rammatical Reasoning Test* (1968), and a correlation of r=.44 (n=84, p<.01) with the *Pearson Reading Comprehension* Test (July, 2011).

TABLE 1. Item Measures and Fit Statistics for the Verbal Analogy Scale

Item	Estimate	Error	Infit MNSQ	Infit *t*
1	-0.34	0.18	1.05	0.58
2	1.20	0.16	0.87	-1.84

3	0.34	0.17	1.14	-1.86
5	-0.06	0.18	0.91	-1.01
7	-0.34	0.18	1.07	0.77
8	0.78	0.16	0.94	-0.77
9	-0.34	0.18	0.96	-0.37
10	-1.27	0.23	0.92	-0.45
11	-0.41	0.19	1.00	0.05
12	-0.65	0.20	1.04	0.40
13	0.25	0.17	1.03	0.42
14	-0.65	0.20	0.90	-0.85
15	-1.01	0.22	1.01	0.18
17	0.19	0.17	0.96	-0.45
19	0.10	0.17	1.05	0.69
20	-0.34	0.18	1.05	0.60
21	-1.11	0.22	0.88	-0.78
22	-0.06	0.18	0.86	-1.61
24	0.92	0.16	0.94	-0.79
25	-1.59	0.26	1.00	0.09
27	-1.59	0.26	0.90	-0.47
29	0.10	0.17	0.87	-1.63
30	0.72	0.16	0.80	-3.00
31	-0.69	0.20	0.95	-0.37
32	0.58	0.16	1.03	0.45
33	0.66	0.16	0.90	-1.29
35	0.00	0.17	1.08	1.02
36	-0.45	0.19	0.97	-0.26
37	0.06	0.17	0.95	-0.53
38	0.00	0.17	0.99	-0.07
39	-0.03	0.18	0.91	-1.03
40	0.34	0.17	0.85	-2.00
41	-0.09	0.18	1.06	0.76
42	0.64	0.16	1.10	1.43
43	0.58	0.16	1.00	0.08
44	1.26	0.16	0.90	-1.32
45	1.37	0.16	0.81	-2.62
46	0.03	0.17	1.02	0.32
47	0.81	0.16	0.98	-0.26
48	-0.26	0.18	1.06	0.73
49	0.72	0.16	1.19	2.61

*MNSQ = mean-square

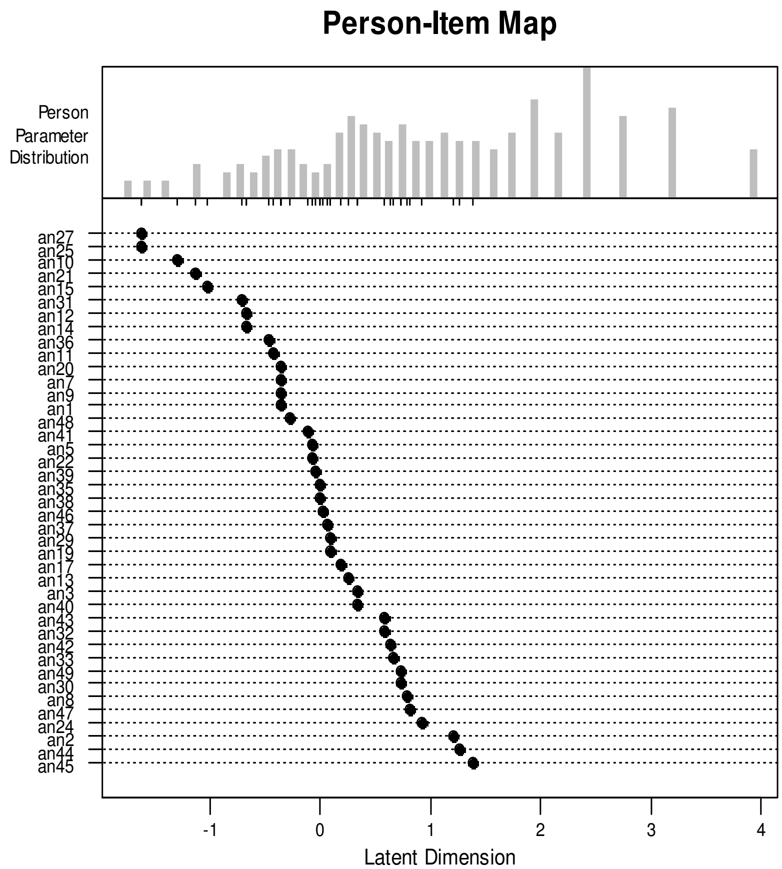

FIGURE 1 Person Item Map

DISCUSSION

The purpose of the present study was to validate a Persian verbal analogies test using the Rasch model and concurrent validation procedure. Due to the importance of verbal analogies in cognitive research and the lack of a valid verbal analogies measure in the Persian language an attempt was made to construct and validate a Persian verbal analogies test in this study. Findings of the study confirmed that the Rasch model fits the verbal analogies test developed in this study after deleting eight items. This supports the internal validity of the test[1].

---

[1] The verbal analogies test validated in this study is available from the author on email request.

Since all the items were multiple-choice the hypothesis that the cause of misfit is due to different test methods can be ruled out (Baghaei, Hohensinn, & Kubinger, 2014). An explanation for the misfit of the items could be that the vocabularies used in the item stems and the options were from different fields in sciences and humanities. To correctly answer the items, examinees needed to have content knowledge of the vocabularies and then figure out the abstract relationships among the vocabularies. The kinds of these relationships were also diverse and varied from linguistic to mathematical relations. These variations in content and relationship types could have introduced some multi-dimensionality in to the test and caused the misfit. If this is the case, the multidimensional Rasch model (Adams, Wilson, & Wang, 1997; Baghaei, 2012) could be used to account for the data. That is, items belonging to different content areas or items sharing the same relationship types are modeled to be on their own separate dimensions (Baghaei, 2013; Baghaei & Aryadout, 2015; Baghaei & Ravand, 2016). This will improve model fit and enable test users to make more informed decisions about the examinees' cognitive strengths and weaknesses and, if need be, provide diagnostic feedback.

Moreover, the person-item map (Figure 1) illustrates that the test covers a limited range of the ability scale. In this study, the person-item map showed that the items are mainly clustered to the center and the left of the scale, suggesting that persons with moderate and low verbal analogical skill are likely to answer most of the items correctly. Very few questions target examinees with very high verbal analogical ability. That is, the test contains few challenging items for participants with a higher level of verbal analogical skill.

Concurrent validation showed that the test highly correlates with a verbal fluid reasoning test, i.e, the Persian adaptation of Baddeley's *Grammatical Reasoning Test*, a nonverbal fluid reasoning test, i.e., the short form of Raven's APM, and a reading comprehension test in English as a foreign language. These correlations support the concurrent validity of the measure against other valid criteria.

The overall findings illustrated that the forty-one multiple-choice item verbal analogies test is an effective unidimensional representation of verbal analogies skill in this study's sample.

One limitation of the study is that the participants were all undergraduate Iranian university students. Future research could validate the instrument developed in this study with speakers from other Persian speaking countries and other age groups and educational levels.

REFERENCES

Adams, R., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*, 1–23.

Andersen, E.B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*, 123-140.

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement, 54*, 394-403.

Baddeley, A. D. (1968). A three minute reasoning test based on grammatical transformation. *Psychonomic Science 10*, 341-42.

Baghaei, P., Khoshdel, F., & Tabatabaee Yazdi, M. (2017). *The Persian adaptation of Baddeley's three-minute grammatical reasoning test*. Manuscript submitted for publication.

Baghaei, P., & Cassady, J. (2014). Validation of the Persian translation of the Cognitive Test Anxiety Scale. *Sage Open, 4*, 1-11.

Baghaei, P., Yanagida, T., & Heene, M. (2017a). Development of a descriptive fit statistic for the Rasch model. *North American Journal of Psychology*, *19*, 155-168.

Baghaei, P., Yanagida, T., & Heene, M. (in press). A new fit statistic for the dichotomous Rasch model: Modified cut-off values. *Rasch Measurement Transactions, 31,* 1621-1624.

Baghaei, P., & Aryadoust, V. (2015). Modeling local item dependence due to common test format with a multidimensional Rasch model. *International Journal of Testing*, *15*, 71–87.

Baghaei, P., & Ravand, H. (2016). Modeling local item dependence in cloze and reading comprehension test items using testlet response theory. *Psicológica, 37*, 85-104.

Baghaei, P., Hohensinn, C., & Kubinger, K. D. (2014). The Persian adaptation of the foreign language reading anxiety scale: A psychometric analysis. *Psychological Reports*, *114*, 315–325.

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education, 28*, 1087-1103.

Baghaei, P. (2012). The application of multidimensional Rasch models in large scale assessment and validation: An empirical example. *Electronic Journal of Research in Educational Psychology*, *10*, 233–252.

Baghaei, P., & Tabatabaee Yazdi, M. (2016). The logic of latent variable analysis as validity evidence in psychological measurement. *The Open Psychology Journal. 9*, 168-175.

Bartha, P. (2016). Analogy and Analogical Reasoning. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from: https://plato. stanford.edu/archives/win2016/entries/reasoning-analogy/.

Bond, T.G., & Fox Ch.M. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. 2nd Edition. Lawrence Erlbaum Associates.

Borsboom, D. (2008). Latent variable theory. *Measurement, 6*, 25-53.

Cattell, R. B. (1973). *Measuring intelligence with the Culture Fair tests.* Champaign, IL: Institute for Personality and Ability Testing.

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Process and intervention* (pp. 113–136). Mahwah, NJ: Lawrence Erlbaum.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). London, England: Oxford University Press.

Dunbar, K. & Blanchette, I. (2001). The invivo/invitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences, 5*, 334--339.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & S. Vaid (Eds.), *Conceptual structures and processes: Emergence, discovery and change* (pp. 461–493). Washington, DC: American Psychological Association.

Esher, F. J. S., Raven, J. C., & Earl, C. J. C. (1942). Discussion on testing intellectual capacity in adults. *Proceedings of the Royal Society of Medicine, 35*, 779-785.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34* (5), 752–775.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D., Holyoak, K. J., & Kokinov, B. K. (2001). *The analogical mind: Perspectives from cognitive science.* Cambridge, MA: MIT Press.

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Ignoffo, M., F. (1980). The thread of thought: Analogies as vocabulary building method. *Journal of Reading,* 23, 520.

Kubinger K.D. (2005). “Psychological test calibration using the Rasch Model—Some critical suggestions on traditional Approaches.” *International Journal of Testing*, *5*(4), 377–394. doi:10.1207/s15327574ijt0504_3.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148–161.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127* (1), 162–181.

Linacre, J. M. (1999). Understanding Rasch measurement: estimation methods for Rasch. Measures. *Journal of Outcome Measurement, 3*, 381-405.

Linacre, J. M. (2009). *A user's guide to WINSTEPS*. Chicago, IL: Winsteps.

Mair, P., Hatzinger, R., & Mair, M. J. (2016). *eRm: extended Rasch modeling* [Computer software]. R package version 0.15-7. http://CRAN.R-project.org/package=eRm.

Markman, A. B., & Wood, K. (Eds.). (2009). *Tools for innovation: The science behind the practical methods that drive new ideas*. New York, NY: Oxford University Press.

Piaget, J. (1949). *Traitd de loique*. Paris: Colin.
Piaget, J. (1950). *The psychology of intelligence*. New York: Harcourt, Brace.
Plaister, T. (1981). Teaching vocabulary, listening comprehension, and reasoning by means of analogies. *Foreign Language Annals. 14* (1), 25-29.
R Core Team (2016). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (Copenhagen: Danish Institute for Educational Research, 1960. ed.). Expanded edition, Chicago: University of Chicago Press, 1980.
Raven, J. C., Court, J. H., & Raven, J. (1977). *Raven's Progressive Matrices and Vocabulary Scales*. New York: Psychological Corporation.
Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London: Lewis.
Richland, L., E. & Burchinal, M., R. (2013). Early Executive Function Predicts Reasoning Development. *Psychological Science, 24*(1) 87–92.
Schunn, C. D., Paulus, P. B., Cagan, J., & Wood, K. (2006). *Final report from the NSF innovation and discovery workshop: the scientific basis of individual and team innovation and discovery*. Washington (DC): National Science Foundation.
Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
Sternberg, R. J. (Ed.). (1988). *The nature of creativity: Contemporary psychological perspectives.* New York, NY: Cambridge University Press.
Sternberg, R. J., and Nigro, G. (1980). Developmental Patterns in the Solution of Verbal Analogies. *Child Development, 51*, 27-38.
Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review. 84*, 353-378.
Stuss, D. T. (2007). New approaches to prefrontal lobe testing. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 292–305). New York, NY: Guilford Press.
Tabatabaee Yazdi, M., & Baghaei, P. (in press). Reading Comprehension in English as a Foreign Language and some Cattell-Horn-Carroll Cognitive Ability Factors. *The Reading Matrix: An International Online Journal.*
Wright*, B. D., &* Linacre*, J. M. (*1994*). Reasonable mean-square fit values. Rasch Measurement Transactions, 8,* 370-371.