البحث(١٣)

Recalibrating the Standard Progressive Matrices-Plus Using the Three-Parameter Logistic Model

إعداد :

Dr. Mohammed Mohammed Ateik AL-khadhera
King Saud University, department of psychology, college of education, Saudi Arabia.
Prof. Ismael Salamah Albursana
King Saud University, department of psychology, college of education, Saudi Arabia.
Prof. Salaheldin Farah Attallah Bakhietb
King Saud University, department of special education, college of education, Saudi Arabia.
Yousif Balil Bashir makib
King Saud University, department of special education college of education, Saudi Arabia.

,

Recalibrating the Standard Progressive Matrices-Plus Using the Three-Parameter Logistic Model

Dr. Mohammed Mohammed Ateik AL-khadhera
King Saud University, department of psychology, college of education, Saudi Arabia.
Prof. Ismael Salamah Albursana
King Saud University, department of psychology, college of education, Saudi Arabia.
Prof. Salaheldin Farah Attallah Bakhietb
King Saud University, department of special education, college of education, Saudi Arabia.
Yousif Balil Bashir makib
King Saud University, department of special education, college of education, Saudi Arabia.

Highlights

- recalibrating the Standard Progressive Matrices-Plus (SPM+) using the three-parameter logistic model (3PL)
- participants were (N = 2045) from Yemen
- analysis revealed a new calibration of the SPM+
- item difficulty ranged between 1.82 and 2.9 logits for 57 items
- items 58, 59, and 60 whose difficulty values were 99.99 were excluded

Abstract

The study aimed to recalibrate the Standard Progressive Matrices-plus (SPM+) using the three-parameter logistic model (3PL). Principal component-based exploratory factor analysis was performed on SPM+ scores of a sample of 2045 basic education students. Calibration was performed according to the 3PL model using BILOG-MG3. Analysis revealed a new calibration of the SPM+. Item difficulty ranged between 1.82 and 2.9 logits for 57 items. Items 58, 59, and 60 whose difficulty values were 99.99 were excluded. Item 33 had the highest difficulty index (2.9), while item 10 had the lowest difficulty index (–1.82). Items 3, 14, 18, and 46 retained their order in the calibration based on Raven's structure and the 3PL model. Items falling among these four items within the same dimension and among dimensions varied.

Keywords. calibration, Standard Progressive Matrices-Plus (SPM+), three-parameter logistic model (3PL), guessing

إعادة معايرة اختبار المصفوفات المتتابعة المعيارية المطور باستخدام النموذج اللوجستي ثلاثي المعلمات

د. محمد محمد عتيق الخضر

جامعة الملك سعود، قسم علم النفس، كلية التربية، المملكة العربية السعودية.

أ.د. إسماعيل سلامة البرصان

جامعة الملك سعود، قسم علم النفس، كلية التربية، المملكة العربية السعودية.

أ.د. صلاح الدين فرح عطا الله بخيت

جامعة الملك سعود، قسم التربية الخاصة، كلية التربية، المملكة العربية السعودية.

أ. يوسف بـــلال بشيـــر مكـــي

جامعة الملك سعود، قسم التربية الخاصة، كلية التربية، المملكة العربية السعودية.

• **المستخلص**:

هدفت الدراسة إلى إعادة معايرة اختبار المصفوفات المتتابعة المعيارية المطور (SPM+) باستخدام نموذج اللوجستي ثلاثي المعلمات .(3PL) تم إجراء تحليل استكشافي عاملي قائم على المكونات الرئيسية لدرجاتٍ SPM+ لعينة مكونة من ٢٠٤٥ طالبًا من طلاب التعليم الأساسي. تم تنفيذ المعايرة وفقا لنموذج PL3 باستخدام برنامج .BILOG-MG3 كشف التحليل عن معايرة جديدة لاختبار .SPM+ تراوحت صعوبة البنود بين -١.٨٢ و٢.٩ لوجيت لـ ٥٧ فقرة. تم استبعاد البنود ٥٨ و٥٩ و٦٠ بسبب قيم صعوبتها العالية (٩٩.٩٩). كان البند ٣٣ هو الأصعب (٢.٩)، بينما كان البند ١٠ هو الأسهل (-١.٨٢). حافظت البنود ٣ و١٤ و١٨ و٤٦ على ترتيبها في المعايرة بناءً على بنية ريفين ونموذج PL3 البنود الواقعة بين هذه البنود الأربع ضمن البعد نفسه وبين الأبعاد المختلفة تباينت في ترتيبها.

الكلمات المفتاحية. المعايرة، المصفوفات المتتابعة المعيارية المطور(SPM+) ، نموذج اللوجستي ثلاثي المعلمات(3PL) ، التخمين.

Introduction

Specialists in measurement seek to develop accurate and objective measures to reliably identify IQ as an index of ability that distinguishes individuals in terms of what they can perform in the testing situation and that can be used to predict what potentials an individual can reach if provided with education and training. Unless potentials are reliably identified, they will not develop into actual abilities in the future for lack of learning and training (Rabea, 2009). Fidelity and justice of diagnosis are therefore crucial to provide decision makers and practitioners with accurate information on students' performance. For this reason, the

measurement movement has witnessed tangible activity both internationally and regionally to adapt, validate, and/or standardize measures of intelligence that prove effective in assessing intelligence, such as Raven's Matrices.

Raven (1938, 1939, 1940) developed his test based on the theory of Spearman who was the pioneer of the two-factor theory of intelligence. It measures the two main components of Spearman's g: the ability to think clearly and make sense of complexity (known as educative ability) and the ability to store and reproduce information (known as reproductive ability). Spearman's research not only led him to develop the concept of the g factor of general intelligence, but also the s factor of specific intellectual abilities. Spearman contended that specific activities share the g factor but differ in the s factor. Differences in performance on IQ tests reside in these two types of factors (Carlson, Buskist, & Martin, 1997; Raven, 2000).

Regardless of its undeniable importance, Raven's Standard Progressive Matrices have received as much criticism as intelligence tests developed in the light of the classical theory of measurement, which are known to lack objectivity of measurement. According to proponents of recent trends in psychological measurement, for measurement of things and characteristics to be objective, the results obtained from measurement need to be independent of the specific tool used to obtain them. Such psychometric problems include the following:

1. Lack of linearity due to the presence of one measurement unit and variation in distance between every two successive scores, which results from the raw score of the individual's observed performance on test items.

2. The dependence of test item parameters (item difficulty and item discrimination) on the characteristics of the sample. Such parameters vary due to the varied abilities of the sample subjects. An item is easy for high ability subjects, but difficult for low ability ones. Relative homogeneity of sample subjects leads to lower discrimination coefficients in comparison with a heterogeneous sample of subjects. The psychometric characteristics of a test therefore differ by the mean and range of the ability of sample subjects. Results are therefore delimited to a population that is similar to the population from which subjects are sampled (Abo-Hashem, 2006).
3. The fact that the individual's total score is affected by test items, that is, it is high when item difficulty is low and vice versa.
4. The inability to compare individuals in measured performance and trait unless the same test items or equivalent items are used, which is a difficult requirement from a practical point of view. This reduces the value of results derived from the traditional theory of measurement.
5. The fact that test reliability is affected by the testing situation. The testing situation can differ because of circumstances of administration when using the test-retest method or equivalent forms. This, in turn, can affect the accuracy and objectivity of measurement. Besides, omitting or modifying any item leads to a change in individuals' scores, a change that is difficult to predict (Hambleton & Swaminathan, 2010).
6. Measures developed in the light of the traditional theory of measurement do not allow for comparison in longitudinal and cross-sectional studies. An individual's achievement or

ability cannot be measured across ages since an individual's measures are expressed by the average of the measurement sample.

7. They do not allow for the direct measurement of actual scores, as they relate to the total score and measurement error.
8. They do not provide sufficient information on the strength of an individual's performance when answering test items for absence of probable estimation of an individual's correct response.
9. The assumption that measurement errors for all subjects are equal, even though the performance of some subjects can be more consistent than the performance of others. This consistency also varies due to variation in individuals' ability level. It is natural that error increases in the case of a difficult test being administered to a group of low-ability subjects, and decreases if it is administered to high-ability subjects (Al-Feqi, 2013).

To overcome the shortcomings of traditional measures that give misleading and false results and lead to invalid predictions about individuals' performance and abilities away from justice, which is the ultimate aim of measurement (Al-Walili, 2005), alternative measurement models emerged in the early 1950s with the publication of Lord's Ph. D. Since its emergence, these alternative measures have undergone continuous validation to achieve objectivity of measurement. They therefore have become indispensable to psychological and educational test developers, as they allow for independence between item characteristics and respondents' abilities. Independence here means that respondents' ability is not affected by item characteristics (sample free), and that item characteristics are not affected by respondents' ability

(item free). Lord (1953), Brinbaum (1958) and Rasch (1960) developed Dichotomous Response Models for Item Response Theory (IRT). These are models that measure items whose scores range between two estimates (0-1).

IRT theory assumes that individuals' performance can be predicted or interpreted based on a characteristic of performance called a trait. That is, one or more traits underlie individuals' responses to test items. These are latent traits that cannot be estimated directly. Rather, they can be estimated through responses to items. Objective measurement is secured in IRT through the stability of item difficulty calibration regardless of how different test takers are. Objectivity is also achieved by freeing the individual's ability from item difficulty, estimating standard error for item statistics and individuals' ability, and providing a measurement unit on which item statistics and the individuals' ability are scaled - the Logit - to secure linearity of measurement (Al-Feqi, 2013).

Researchers therefore have attempted to modify classical IQ tests, including the various forms of Raven's test, based on the modern theory of measurement. For example, Chissom and Hoenes (1976) employed a Rasch model to compare the ability of the D-48 and IPAT culture fair intelligence tests to predict SAR achievement test scores. Using the Prox-method, Alaam (1985) used a Rasch model in examining the 22-item Mindfulness Test developed by Ramziyah Al-Ghareeb using a sample of male and female university students. Nenty (1986) explored cultural bias in Cattell's test by administering it to large samples of Americans, Nigerians, and Indians. The researcher used four different methods in analyzing test items and establishing

non-bias in its items. These four methods were Sheuneman's Modified Chi-Square, Rudner's Item Difficulty Coefficient, Convey's Item Difficulty Coefficient, and a 1P Rasch Model. El-Korashy (1995) used a Rasch Model to select items for an Arabic version of the Otis-Lennon Mental Ability Test. Zimowski and Wothke (1987) analyzed tests of spatial ability for their visuospatial and verbal reasoning components using a one-parameter latent trait model via the Bilog program. Nour El-Din (1995) explored the psychological dimensions of the Stanford-Binet Intelligence Scale: Fourth Edition by examining its items, establishing its reliability and validity, and extracting its various criteria for a sample of preschoolers.

In the study conducted by Van der Ven and Ellis (2000, cited in Eid, 2005), the Standard Progressive Matrices Test was administered to 901 students whose ages ranged between 12 and 15 years to separately analyze its five sub-tests. Three of the five tests were found to be consistent with a Rasch Test: A, C, and D. The other two sub-tests, B and E, were not. Al-Tantawi (2004) used a Rasch Model to standardize Raven's Progressive Matrices Test. The calibration sample consisted of 1411 elementary and intermediate school students whose ages ranged between 6 and 13. Results revealed that the order of items in the final version of the test was identical with that of the original test.

Bakhiet (2012) recalibrated and standardized the Standard Progressive Matrices using a Rasch Model. The study aimed to investigate the extent to which the items of the Standard Progressive Matrices Test fitted the One-Parameter Rasch Model, which is the basis of the modern theory of psychological measurement. Based on data, the researcher

developed new criteria for the test that can be used to interpret individuals' ability levels.

Based on what was mentioned above, it is clear that it is imperative to standardize intelligence tests developed in the light of the classical theory to fit them to recent trends and achieve the highest degree of objectivity. Such tests need to be recalibrated according to IRT. Hence, the researchers of the present study sought to recalibrate the Standard Progressive Matrices-Plus (SPM+) using the three-parameter logistic (3PL) Model. Hopefully this will eliminate the weaknesses in previous studies concerning the calibration of Raven's using a Rasch model, as test items are answered by selecting the correct answer from alternative options, which makes performance on it questionable because of guessing effects. This can adversely affect decisions made based on this performance. More specifically, the study addressed the question "What is the calibration of the items of Raven's SPM+ using the 3PL Model?"

Definition of Terms

Standard Progressive Matrices-Plus (SPM+)

The SPM+ is one of Raven's (2008) basic versions of Standard Progressive Matrices (SPM). Owing to the calibration process performed in Britain in 1998 and its recommendations, this version of the progressive matrices was modified by adding items with high difficulty parameters to strengthen the discrimination power of the test to overcome measurement errors resulting from respondents' familiarity with it. This coincided with the development of Mill Hill's Vocabulary Test as a supplement to the measurement of verbal ability (Raven & Court, 2002).

The Three-Parameter Logistic Model (3PL)

Lord (1980) developed this model adding a third parameter that he termed the Lower Asymptote Line or Guessing Parameter, which represents the probability of low-ability individuals' arriving at the correct answer of a multiple-choice item by guessing when ɵ = 0. That is, selection of the correct answer by guessing does not relate to ability level, and the theoretical range of c is $0 \le c \le 1$. This eliminated the effect of random guessing on performance. Its mathematical equation is as follows:

$$pi(\theta) = Ci + (1 - Ci)\frac{e^{Da(\theta - bi)}}{1 + e^{Da(\theta - bi)}}$$

Where Ci = the guessing parameter for the item i.

Calibration

Calibration means estimating the location of the item on the measured trait continuum using the 3PL Model (Alaam, 1987).

Method

The researchers used the analytical descriptive method to achieve the aims of the study and answer its question. This method is optimal for analyzing, describing, and comparing characteristics according to the 3PL Model to attain accurate and objective measurement.

Participants

Participants were male and female students (6-14 years) drawn by stratified random sampling from the population of Basic Education students in the City of Dhamar. The study instrument was administered to

participants (N = 2045) during the school year 2015-16 to obtain stable estimates of the parameters.

Instrument

The researchers used the Standard Progressive Matrices-Plus Test (Raven, 2008).

Procedures

The test was marked manually and data were scored on excel sheets. Data were then analyzed using the BILOG-MG3 program of the 2PL Model to extract participants' ability, item difficulty, and standard error of estimation.

Results

First: Testing assumptions:

Unidimensionality: The researchers used Principal Component Factor Analysis to test the unidimensionality assumption after checking the conditions of factor analysis.

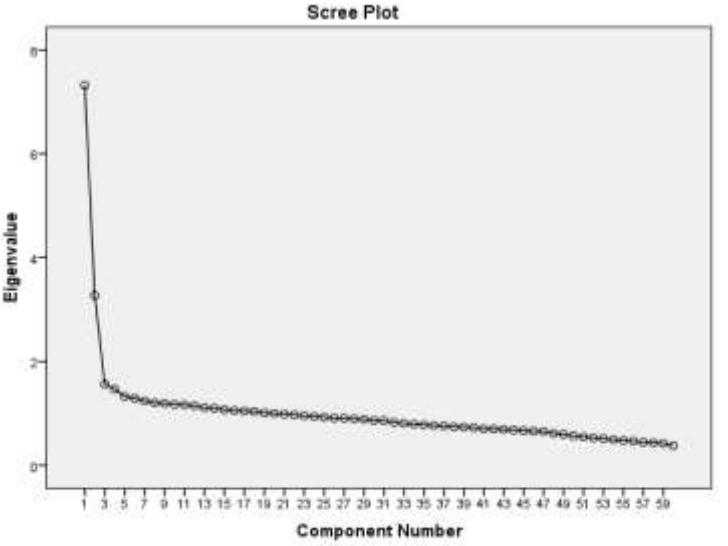
Table 1. Bartlett's Test of Sphericity and Kaiser-Meyer-Olkin (KMO) Test

KMO	Bartlett		
	χ^2	df	α
	18769.142	1770	0.000

Values of latent roots and the percentage of variance explaining factors whose latent roots were ≥ 1 were computed. Factor substantiality was identified by a value 0.3 as the minimum level for acceptance of item loadings based on Gilford's criterion. Nine factors were extracted that accounted for 51.28% of variance. Values of the latent root for the first factor and the second factor were found to be 7.32 and 3.27, respectively. Hence, the proportion of the first factor to the second factor exceeded 1:2 according to Lord's Index (Lord, 1980). The first factor explained 12.20% of the total (51.28%) variance, which represents 24% of the total

variance. This an index of unidimensionality based on Rechase's contention (cited in Embreston & Riese, 2000) that the first factor's explaining at least 20% of the total variance is an index of unidimensionality. The Scree Plot Test showed a steep decline between the value of the latent root for the first factor and the second factor, which also supports the scale's unidimensionality.

Figure 1. The Scree Plot Test



Local Independence: Meeting the unidimensionality assumption is enough to assert meeting the local independence assumption. This is evident in correlation coefficients among items that did not reach 1, which indicates a complete match in answering two items. This means that the answer to an item does not depend on the answer to another item (Hambleton & Swaminathan, 2010).

Item Characteristic Curve: This assumption refers to the presence of a characteristic curve for every item, which is

yielded by the BILOG-MG3 program analysis containing three aspects, one of which is the graphical analysis of the items. This graphical analysis shows a characteristic curve for every test item.

Speediness: Speediness is an implicit assumption which is also supported by the unidimensionality assumption. That factor analysis produced two factors, not one, is an index of meeting the speediness assumption. That is, speed is not a factor that affects the test results, which indicates that the test is a strength test, not a speed test (Hambleton & Swaminathan, 2010).

Second: The Appropriateness of items and participants for analysis based on IRT models

Absence of items to which all participants responded correctly (i.e., items below participants' ability) or wrongly (i.e., items above participants' ability) was verified. Data were also analyzed to exclude participants who were not appropriate for the calibration process. These are participants who failed to answer all test items correctly (i.e., they are below the test level) and participants who answered all test items correctly (i.e., they are above the test level) (Al-Anbaki, 2009). Hence, test data were confirmed to meet the assumptions of IRT, which supports the appropriateness of the data for analysis based on the 3-Dimension Model of IRT.

Raven's SPM+ has a total of 60 items presented in 5 sets (A–E), with 12 items per set. Items in each set relate to the same theme and range in difficulty from the least to the most difficult (Saheli, 2008). Furthermore, sets are arranged in order of increasing difficulty in accordance with their order in the test and they differ in answer modes. Similarly, the

cognitive mental operations measured range from the least to the most difficult. The test therefore starts with the identification of missing pieces to complete figures or patterns and ends with comparison and reasoning as the highest ability. Respondents are required to select a response to complete a missing pattern or space (set A), match figures (set B), represent a regular change in figure patterns (set C), reorder, change, or switch figures (D), and analyze or identify the relationship among pieces of the figure (Raven, 2008).

Discussion

Data analysis produced a new calibration of the SPM+. Item difficulty ranged between 1.82 and 2.9 logits for 57 items. Difficulty coefficients for items 58, 59, and 60 were 99.99. They are higher than participants' ability level and are therefore useless from a technical perspective. Accuracy and objectivity of item calibration based on the 3PL Model was supported, as the difference between the estimates of any two successive items was lower than the sum of their standard error. Item 33 had the highest difficulty index (2.9), while item 10 had the lowest difficulty index (–1.82). It can be observed that the most difficult items (58, 59, and 60) retained the same order in the two calibrations. Additionally, items 3, 14, 18, and 46 retained their order on the calibration based on Raven's structure and the 3PL model. Items falling among these four items within the same dimension and among dimensions varied.

Table 2. Item difficulty estimates and order according to the 3PL Model of the IRT

LOGIT	3PLM	CTT	LOGIT	3PLM	CTT	LOGIT	3PLM	CTT
2.23	40	41	1.4	17	21	-1.82	10	1
2.25	35	42	1.48	31	22	-1.25	11	2
2.25	36	43	1.5	20	23	-1.22	3	3
2.28	41	44	1.53	16	24	-1.15	1	4
2.29	30	45	1.54	26	25	-1.11	2	5
2.33	46	46	1.58	15	26	-0.62	12	6
2.34	54	47	1.58	29	27	-0.42	6	7
2.37	55	48	1.59	27	28	0.27	4	8
2.41	37	49	1.6	28	29	0.27	23	9
2.54	43	50	1.64	22	30	0.48	25	10
2.56	50	51	1.77	39	31	0.55	5	11
2.68	56	52	1.78	21	32	0.62	7	12
2.8	44	53	1.91	34	33	0.88	9	13
2.81	49	54	2.01	48	34	0.99	14	14
2.88	57	55	2.02	38	35	1.04	8	15
2.89	53	56	2.06	52	36	1.07	19	16
2.9	33	57	2.11	45	37	1.15	13	17
99.99**	58	58	2.16	51	38	1.32	18	18
99.99**	59	59	2.18	32	39	1.35	24	19
99.99**	60	60	2.22	47	40	1.38	42	20

First Dimension
Second Dimension
Third Dimension
Fourth Dimension
Fifth Dimension

Items retaining order
** Excluded items
New

It is clear that this study made use of linearity, which is the characteristic of the 3PL Model where there is one measurement unit for both item difficulty and the respondent's ability, the Logit. Analysis revealed a difference in the order of items before and after calibration. The

calibration of items differed within and among dimensions. Ten items retained their order in the first dimension (items 1-12), with one item having the same location (item 13). Two items moved higher and 8 items retained their order in the second dimension (items 13-24), with two items having the same location (items 14 and 18). One item moved lower in the first dimension and 3 items moved higher in the third dimension. In the third dimension (items 25-36), 5 items retained order within the dimension, whereas 2 items moved lower in the first dimension and 5 items moved higher to the fifth dimension. In the fourth dimension (items 37-48), 5 items retained order within the dimension, with one item having the same location (item 46), while 4 items moved lower, one of which moved to the second dimension and 3 to the third dimension, and 3 items moved higher in the fifth dimension. In the fifth dimension (items 49-60), 8 items retained their order within the dimension and 4 items moved lower, 3 of which moved to the fourth dimension and one to the third dimension. Thus, the order of items after calibration became more logical than it was before calibration. These results also support the contention that traditional intelligence tests show high psychometric characteristics when used with latent trait models.

The final version of the scale after calibration with the 3PL Model consisted of 57 items with an order of increasing difficulty after omitting three items as shown in table 2.

Recommendations

- Using the modern theory of testing, particularly the 3PL Model, in developing psychological tests based on the

psychometric characteristics extracted according to the 3PL Model.
- Calibrating Raven's SPM+ with age groups according to the age range specified for the test.
- Making available the programs required for using the modern theory of testing in analyzing tests such as Xcalibre & Rumm2030, Bilog-Mg3, and R-Studio.

References

- Abo-Hashem, A. M. (2006). Comparing the classical theory and Rasch model in selecting items of the Approach to Studying Inventory for university students. *Journal of the Faculty of Education, Zagazig University, 52*, 1–52. *(In Arabic)*
- Alaam, S. (1985). Analyzing mental tests data using Rasch Logarithmic model (an experimental study). *The Arabic Journal for Human Sciences, 5*(17), 100–122. *(In Arabic)*
- Alaam, S.(1987). A comparative critical study of latent trait and classical models in psychological and educational measurement. *The Arabic Journal for Human Sciences, 27*, 18–44. *(In Arabic)*
- Al-Anbaki, H. (2009). *Estimating cut-off scores in criterion referenced tests*. (Unpublished Ph.D. dissertation). Faculty of Education, Baghdad University. *(In Arabic)*
- Al-Feqi, E. (2013). *Psychological and educational evaluation and measurement* (2nd Ed.). Riyadh: Al-Alam Al-Arabi Bookstore. *(In Arabic)*
- Al-Tantawi, M. (2004). *A psychometric study on the standardization of Raven's progressive matrices using Rasch Model*. (Unpublished M.A. thesis). Girls' College, Ain Shams University. *(In Arabic)*
- Al-Walili, E. H. (2005). Equivalence of test scores in the light of the classical and modern test theories (a psychometric comparative study). *Journal of the Faculty of Education, Benha University, 63*(15), 51–149. *(In Arabic)*

- Bakhiet, S. (2012). Recalibrating and standardizing the standard progressive matrices test using Rasch Model. *Journal of the Faculty of Education, Khartoum University, 4*(6), 61–94. *(In Arabic)*
- Carlson, N. R, Buskist, W., & Martin, G. (1997). *Psychology, the science of behavior*. Allyn and Bacon.
- Chissom, B. & Hoenes, R. (1976). A comparison of the ability of the D-48 and IPAT culture fair intelligence test to predict SAR achievement test scores for 8th and 9th grade students. *Educational and Psychological Measurement, 36*, 561–564.
- Eid, K. (2005). Evaluating the structure of advanced progressive matrices test and its short version using factor analysis and Rasch Model. *Journal of Psychological and Educational Research, 3*, 256–283. *(In Arabic)*
- El-Korashy, A. (1995). Applying the Rasch model to the selection of items for a mental ability test. *Educational and Psychological Measurement, 55*(5), 753–763.
- Embreston, S. E. & Riese, S. P. (2000). *Item response theory for psychologists.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Hambleton, R., Swaminathan, H. (2010). *Item response theory: Principles and application*. Boston: Kluwer-Nigh off Publishing
- Lord, F. (1980). *Application of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Nenty, H. J. (1986). Cross-cultural bias analysis of Cattell's culture fair
  intelligence test. Paper presented at the annual meeting of the American educational research association, 70th, San Francisco, CA. April 16–20.
- Nour El-Din, A. (1995). *Some psychometric characteristics of Stanford Binet scale-revised among samples of preschoolers*. (Unpublished Ph.D. dissertation). Faculty of Education, Ain Shams University. *(In Arabic)*
- Rabea, M. S. (2009). *Personality inventory*. Amman: Dar Al-Masirah, Jordan.

- Raven, J. C. (1938). *Progressive Matrices*. H. K. Lewis and Co.
- Raven, J. C. (1939). The R.E.C.I. series of perceptual tests: An experimental survey. *British Journal of Medical Psychology, 18*(1), 16–34. https://doi.org/10.1111/j.2044-8341.1941.tb00316.x
- Raven, J. C. (1940). *Progressive Matrices; Instructions, Key and Norms*. H. K. Lewis and Co.
- Raven, J. (2008). *Standard Progressive Matrices-Plus version and Mill Hill Vocabulary Scale manual*. London: Pearson.
- Raven, J. Raven, J. C., & Court J. H. (2000). *Manual for Raven's progressive matrices and vocabulary scales*. Section 3.
- Saheli, N. (2008). *A preliminary standardization of Raven's progressive matrices on samples of students with special needs in Syria*. (Unpublished M.A. thesis). Faculty of Education, Damascus University. *(In Arabic)*
- Zimowski, M., & Wothke, W. (1987). Purification of spatial tests and reasoning components in spatial tests. Paper presented at the Annual meeting of the American Educational Research association, Washington, DC, April 20–24.

❋❋❋❋❋