Journal of Human Sciences
Oum El Bouaghi University
ISSN 1112-9255/E-ISSN 2588-2414
Volume 08 Number 03 December 2021

OEB Univ. Publish.

Assessment of Psychometric for Raven Coloured Progressive Matrices (CPM) Using Rasch Model

التقييـــم السيكومتري لاختبـار رافن للمصفوفات المتتابعة الملون باستخدام نموذج راش

Rachid ziad[1]

[1]ziad.rachid@univ-tebessa.dz, Larbi Tebessi University-Tebessa-Algeria

Date of receipt: 2021-08-21 Date of revision: 22/09/2021 Date of acceptation: 2021-10-22

Abstract

The study aimed at the Psychometric Assessment of Raven Coloured Progressive Matrices (MPC) of intelligence on primary school students in the Algerian environment using the Rasch model. A sample of 427 was selected in the Convenience Sampling method, 2015/2016. The results showed the deletion of 06 items from 36 items of the overall test, which were shown to meet Rasch model assumptions as items fit, Unidimensionality, local independence, items polarity, the reliability coefficient of persons and items and separation indicators, and Differential Item Functioning (DIF) of gender. Thus, the validity of the remaining items assessed to measure the intelligence of primary school pupils.

Keywords: Psychometric; Raven Coloured Progressive Matrices (MPC); intellectual quotient (IQ); Rasch model.

الملخص

هدفت الدراسة إلى التقييم السيكومتري لاختبـار رافن للمصفوفات المتتابعة الملون للذكـاء على تلامذة المرحلة الابتدائية في البيئة الجزائرية باستخدام نموذج راش. اختيرت عينة مكونة من 427 بطريقة المعاينة الميسرة، السنة الدراسية 2016/2015. أظهرت النتائج حذف 06 مفردات من 36 مفردة مكونة للاختبار الكلي، تبين أنها تفي بافتراضات نموذج راش كملائمة المفردات، وتحقق أحادية البعد، والاستقلال المحلي، وقطبية المفردات، ومعاملات ثبات الأفراد والمفردات ومؤشرات الفصل، والأداء التفاضلي للمفردات بين الجنسين. وبالتالي صلاحية المفردات المتبقية التي تم تقييمها من أجل قياس ذكاء تلامذة المرحلة الابتدائية.

الكلمات المفتاحيـــة: السيكـومتري؛ اختبـار رافـن للمصفوفات المتتابعـة الملون؛ نسبة الذكـاء؛ نموذج راش.

***المؤلف المراسل: الاسم الكامل، الإيميل:**

الدكتور زيـــاد رشيـــد ziadpsy@gmail.com

1. Introduction:

Intelligence is one of the closest psychological concepts associated with the educational process and the school achievement of primary school students, and in general, knowing the nature of intelligence and its association with academic achievement enables the teacher to understand the factors associated with success in school life as well as in public life and

helps to perform his educational mission in a more sophisticated and effective manner.

Intelligence (IQ) measurement was undoubtedly the most important thing that concerned scientists at the beginning of the movement of measuring individual differences, so the history of Intelligence measurement is in line with the date of measuring these differences, and therefore in line with the history of experimentation in psychology. Intelligence can also be measured through performance work that depends in essence on the skill of the individual's cognitive organization and good reception, the skill of his expression and his use of elements of external reality and its reconfiguration in complex configurations. (Safwat Abu Faraj, 2007)

Since psychologists have different views on the nature of intelligence and its definitions, there are also multiple theories suggested to show its nature, and this study will rely on Spearman general intelligence theory, which defines intelligence as a general mental capacity behind different behaviors as a form of mental energy generated by the brain.

Raven Coloured Progressive Matrices were tested within the two-factor intelligence model proposed by Spearman theory, which was considered one of the most prominent theories that tried to explain the mental activity, and Spearman was considered the main mentor of the mental measurement movement in Britain for nearly half a century. Raven Coloured Progressive Matrices test is one of the most popular in most countries of the world, and one of the most reflective and realistic translations of the theory data of Spearman two-factor of intelligence, who tried to measure mental activity without relying on acquired skills.

2. Research Problem:

The test of Raven Coloured Progressive Matrices is one of the three matrix tests prepared by the English psychologist j. Raven, where the Matrices appeared in 1938, and he and his students continued to develop these Matrices for more than 30 years until his death in 1970. Raven Matrices are non-verbal intelligence tests and are largely cross-cultural and rely mainly on collective application and can be applied individually in certain circumstances. Raven drew his basic idea of building Matrices from Spearman, who used geometrically painted panels and asked the tester to describe the rule controlling their relationship. Raven's first experimental image was nine shapes spearman-like, and instead of asking the tester to mention the rule explaining the relationships between the shapes, Raven was asked to know the missing part of the shapes.

The purpose of the Matrices is to measure the ability to develop relationships and linkages. Raven designed the Progressive Matrices test to measure The General Factor of Spearman, where many scientists confirmed that Raven's tests are a good measure of the general factor, and are one of the most famous and widespread tests free from the impact of culture.

Raven tests Standardized of intelligence measurement in the Arab environment, such as the Abu Hattab study, 1979 on the Saudi environment, the Al Thani study, 2001 on the Qatari environment, the study of al-Khatib and al-Mutawakel, 2001 on the Sudanese environment, and the study of Majali, 2005 on the Jordanian environment, in Kuwait, Abdul Khaleq and Raven legalized kuwaiti students, 2006, while in Algeria the test was Standardized in 1961 by Guemonprez as part of the psychotechnics selection of Algerian Muslim candidates for the professional formation center for adults; it was in the colonial phase of Algeria. Jaballah and Belbekkai, 2017 tried to adapt the test in the Algerian environment to primary school students with the supply and provide Algerian interpretation criteria for the test.

Globally, it has been Standardized in most countries using a classical theory of measurement, while many studies have chosen to developed Raven tests for modern measurement theory and item response theory, similar to van Ellis,2000, which aimed to use Rasch analysis to test Raven Coloured Progressive Matrices by applying the Raven test separately to the five subgroups and investigating whether it was a one-dimensional test and each subgroup measured one dimension and one general factor.

The Mona Rabie Mitard study (2000), which aimed to re-Calibrating re-grade Raven Progressive Matrices test items according to its difficulty using the one-parameter Rasch model, and then tested the extent to which the person ability to test the total and sub-testing of his or her was equivalent, consisted of a sample of 1,411 primary and middle school students between the ages of 6 and 13 from Cairo and beyond. The results showed the unidimensionality and measurement independence provided by Rasch model in the calibration of each test group in its final form, and misfitting items were deleted. While the study of Ali Abbas Shanan Zamli, which used the latent trait theory according to Rasch model in the development of raven test for advanced Progressive Matrices on middle school students in Baghdad city, and the calibration of raven test for advanced Progressive Matrices, after applying to a group of 500 primary school students in Baghdad, the researcher used Rascal software, after verifying the assumptions of the Rasch model as one-dimensional based on exploratory factor analysis, Guttman-Kaiser criterion, and to verify the

matching of The model based on chi-square goodness of fit matching value, removal two items of the test for misfitting at a statistical significance level (0.05), and the value of the test discrimination coefficients approached (1) which was (0.85), and achieved measurement independence to achieve objectivity in measurement. Matore et al, 2018 study aimed at assessing the psychometric characteristics of Raven's advanced Progressive Matrices test in the Malaysian environment on multi-technology school students using the Rasch model. They used the survey method and the Convenience Sampling method to select 150 students. The results showed that 23 of the 36 items that made up the test were misfitting to the Rasch model. The remaining 13 persons met all conditions such as one-dimensional, local independence, items polarity, reliability and separation index. Thus, it showed its suitability to evaluate intelligence in the future. Previous studies show that their results are conflicting, particularly in studies that used the two different measurement theories to analyze their classical measurement theory and the item response theory to the development of this test.

Accordingly, the problem of the study is focused on the urgent need to provide a test that measures intelligence ability and its redevelopment in the Algerian environment, so that it can be used to measure the intelligence of primary school students for the purpose to classifying and diagnosing them in the future. Because of the suitability of the Raven test of colored progressive matrices to measure this ability, and its freedom from the influence of bias cultural factors, this research focused on the Calibrating grading of this test using the test item response theory and the Rasch model.

2.1. Research Question:

1. How difficult is the items of Raven Coloured Progressive Matrices testing to be Calibrating using the one-parameter Logistic Rasch model
2. How identical are responses to Raven Coloured Progressive Matrices test items for the one-parameter Logistic Rasch Model?
3. Does items Raven Coloured Progressive Matrices test show Differential Item Functioning (DIF) depending on the sex variable?

3. Objectives:

The study aims to verify the assumptions of Rasch model of Raven Progressive Matrices test data, and Calibrating the difficulty of his items, and check the Differential Item Functioning (DIF) of his items as well depending on gender.

4. The importance of the study:

1. This study is one of the first few attempts (for the researchers) to Calibrating a general intelligence test, using the item response theory and Rasch model in the Algerian environment.

2. Raven Coloured Progressive Matrices test is one of the best available tests to measure general intelligence, and one of the best tests that are free from cultural limitations to measure intelligence.

5. Procedural definitions of the concepts of the study:

1. Intelligence according to raven: Raven concept of intelligence and a statement of meaning is derived from the two-factor of intelligence theory of Spearman, who used paintings with geometric shapes and the testers describe the rule governing the relationship between these shapes. Raven's tests of Coloured Progressive Matrices measure:

5.1. The general mental ability (Capacity) of the person during the testing, when given as power tests without specifying the time of the answer, thus measuring the accuracy of observation and clear ordered thinking that does not depend on the previous information gained by the person, where Raven asserts that successive matrices tests test the person energy at the moment of the test to understand meaningless forms, he must observe them and understand the relationships between them, understand the shapes, and complete each system of relationships presented, thereby developing an organized method of inference, preferably use the test to measure mental ability in human, genetic and clinical studies.

5.2. Mental competence (Proficiency) when given as a speed test in which the answer time is determined between 30-40 minutes, thus measuring a person ability to make quick and accurate judgements according to the requirements of the situation, is therefore used to distinguish between individuals with a speed of reflection and individuals who have slow thinking, preferably using Raven tests to measure mental competence in the field of choice and educational and professional guidance. (Raven et al, 1998)

6.One-parameter Logistic Rasch Model (G.Rasch):

called One-parameter, a special case of the two-parameter Birnbaum model. This model is concerned with measuring all the test items that are supposed to discriminate between people but only varies in difficulty, which is given by the following equation:

$$P_i = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

Where **i** items is equal (**1, 2, ...n**) $Pi(\theta)$ ؛The probability that the respondent who has the ability θ to answer the item **i**؛ β_n the ability

parameter of the person **i**؛ **δi** the difficulty of the item **i**. (Haidar, 2013, p286)

7.Method and Tools:

7.1. Research method:

The researcher followed descriptive statistical analytical method in the calibration of Raven Coloured Progressive Matrices test items.

7.2. Sampling and Participants:

The current study Population represents primary school students in different states in the Algerian environment, and is one of the primaries of the following states: Constantine, El Oued, M'sila, And Chlef who study during the 2015/16 school year. The sample was made up of 427 pupils, 290 females and 137 male primary school pupils. The ages of the sample members were limited to (11.22-6.33) years and the mean and standard deviation of the age (9.44±1.44) years for the members of the total sample. The mean and standard deviation of the age of male sample members (9.41±1.42) years (9.48±1.48) were female sample years. The study sample members were selected in a convenience sampling.

7.3. Instrument of Research:

Raven Coloured Progressive Matrices Test (MPC) first appeared in 1947 and been modified in 1956, where the English scientist Raven spent about 30 years preparing and develop the Test form. The test applies to ages between (5.6- 11.6) years of age and mentally retarded, as well as older persons between the ages of 65 and 85. The correlation between Raven Coloured Progressive Matrices with the WISC Children's Wechsler Verbal Section ranged from (0.31-0.84) to the performing section between (0.05-0.74) and between Raven Coloured Progressive Matrices and Stanford-Binet test ranged from (0.05-0.74) to Stanford-Binet (0.32-0.68) Between Raven Coloured Progressive Matrices and the test of the shapes included, they ranged from (0.04-0.58) to Raven Coloured Progressive Matrices and Goodenough-Harris Drawing a man Test (GHDAMT), ranging from the value of the link coefficient (0.48). Correlation coefficients between Raven Coloured Progressive Matrices and educational achievement ranged from (0.35-0.39) to correlation coefficients between Raven Coloured Progressive Matrices and teachers' estimates of student intelligence (0.34-0.49). Raven's Coloured Progressive Matrices test is well-Reliability, tracking previous studies that have confirmed this using different Reliability methods, indicating that this test is a good measurement tool. His Reliability coefficients ranged from studies by Kazem et al. (A and B) 2008, Kluever, 2006, and Uno Et al, 2005, Weichbold et Herka, 2003, Zamzmi 1999,

Raven, 1990, Sharkey 1987, Valencia, 1984, Carlson and Jensen, 1982, Evans, 1980, test-retest reliability between (0.62-0.99) and mean of (0.76), and in the Split-Half Reliability ranged from (0.44 - 0.99) to a mean of (0.88) and correlation coefficients between the subtests ranged from (0.55-0.82). While Abdul Fattah al-Quraishi in 1987 Standardized the test on a sample of Kuwaiti children, the correlations ranged from testing to some of the subtests of the Wechsler test and Mazes Porteous, and the Seguin Form Boord between (0.22-0.45) Correlations between the subtests of the scale were also calculated and ranged from (0.71-0.46) to the calculation of correlation coefficients between the scale subtests and the overall score and ranged from (0.91-0.75).

7.4. Statistical Analysis:

In this study, the Statistical Package of Social Sciences (Spss v21) and The Rasch Model was applied in calibration and assess the reliability and validity of the test using WINSTEPS® Version 4.0.0 (Linacre, 2018). In this study, the joint maximum likelihood estimation was used.

8. RESULTS AND DISCUSSION:

1. How difficult is the items of Raven Coloured Progressive Matrices testing to be Calibrating using the one-parameter Rasch model?

To answer this question, we will therefore discuss the items fit for the Rasch model, polarity of items, indexes of Reliability, and separation of persons and items.

Assessment of the fit of items: The program deletes the responses of persons who have answered correctly to all test items as well as persons who have not answered any correct answer, and the program deletes the items to which all students have correctly answered, or which no student has answered correctly, as the results of the analysis showed that four persons were deleted by the program, while all the items of the Raven Coloured Progressive Matrices test **(**36 items) succeeded in analyzing the responses of sample members (427) pupils.

After excluding the person who did not fit to the model, those with limits of fit above (2) or less than (-2), the analysis was re-analyzed to detect non-conforming items, where the difficulty parameter for each item was estimated, the standard error in measuring the difficulty parameter, and the internal items matching statistical values (The Standardized Information Weighted Fit Statistics for Items Infit ZSTD), expressed by statistics of The Mean Square Infit Statistics, MNSQ. have also been estimated The

Standardized Information Weighted Fit Statistics for Items Outfit ZSTD. expressed in the mean Square Outfit Statistics, MNSQ, for each of the difficulty parameters estimated in the logit unit, and table (01) shows the results.

Table 01. Explains the difficulty of items and its standard errors arranged downwards according to the difficulty of items.

Outfit		Infit						
ZSTD	MNSQ	ZSTD	MNSQ	Model S. E	Measure	PTBIS-CORR	Raw Score	Item
−0.1	0.99	0.3	1.01	0.11	1.81	0.40	190	**AB12**
−0.6	0.97	−0.4	0.98	0.11	1.79	0.42	192	**B12**
1.2	1.07	0.6	1.02	0.11	1.22	0.36	241	**A12**
−1.0	0.94	−1.3	0.95	0.11	1.02	0.43	258	**B10**
−1.8	0.88	−1.8	0.92	0.11	0.90	0.45	267	**AB11**
−0.4	0.97	−1.2	0.94	0.11	0.74	0.41	280	**AB10**
−0.6	0.95	−0.3	0.98	0.11	0.72	0.38	282	**B2**
1.2	1.10	1.7	1.08	0.12	0.65	0.27	286	**AB7**
−0.2	0.98	−0.4	0.98	0.12	0.61	0.37	290	**AB9**
0.9	1.08	2.0	1.11	0.12	0.47	0.24	300	**AB4**
−0.6	0.94	−0.9	0.95	0.12	0.45	0.39	302	**B9**
−0.1	0.99	1.0	1.01	0.12	0.44	0.34	302	**AB8**
1.5	1.16	1.9	1.11	0.12	0.35	0.22	307	**AB5**
−0.8	0.92	0.3	1.02	0.12	0.24	0.33	316	**B3**
1.2	1.13	1.5	1.10	0.12	0.21	0.24	318	**AB6**
−1.7	0.81	−1.0	0.93	0.13	0.09	0.23	326	**B7**
0.4	1.05	0.6	1.04	0.13	0.00	0.40	331	**B4**
−0.5	0.93	0.1	1.00	0.13	−0.10	0.27	337	**B5**
2.5	1.36	1.8	1.14	0.13	−0.12	0.14	338	**B1**
0.3	1.04	0.3	1.02	0.14	−0.14	0.28	339	**B6**

−0.6	0.90	−1.0	0.91	0.14	−0.31	0.36	348	**A10**
−0.9	0.85	−1.2	0.89	0.15	−0.48	0.38	356	**A9**
1.2	1.20	−0.1	0.99	0.15	−0.50	0.26	356	**A8**
−0.2	0.95	0.3	1.02	0.16	−0.72	0.23	366	**AB1**
−0.3	0.92	0.0	0.99	0.18	−1.03	0.24	377	**A4**
−1.4	0.69	−0.8	0.88	0.19	−1.13	0.35	380	**A5**
0.3	1.06	0.0	0.99	0.20	−1.32	0.21	385	**A3**
−0.4	0.88	0.0	0.99	0.20	−1.32	0.21	385	**A7**
1.1	1.37	0.3	1.05	0.25	−1.87	0.06	396	**A2**
−0.5	0.70	0.1	1.00	0.36	−2.66	0.13	405	**A1**
0.00	0.99	0.00	1.00	0.15	0.00		318.5	**Mean**
1.00	0.15	1.00	0.06	0.05	1.00		54.00	**SD**

Source: Winstpes v4. 0. 0

Table 01. shows statistical indicators after deleting the 14 misfitting persons, and the 6 misfitting items: A6-A11-B8-B11-AB2-AB3, delete two items from each sub-test The component of the test as a whole, for the purpose to ensure that the data resulting from the sample members responses to the test, verifying the objectivity of the test in its final form (30) items, and verifying assumptions, have been re-analyzed to estimates free of items difficulty and person's ability, and table 2 shows the results of the analysis of liberal values from the abilities of person's.

Table 02. Results of analysis of values free from the abilities of person's n=413.

Outfit		**Infit**					
ZSTD	**MNSQ**	**ZSTD**	**MNSQ**	**Model S. E**	**Measure**	**Raw Score**	**Statistics**
0.1	0.99	0.1	1.00	0.53	1.64	23.1	**Mean**
0.8	0.46	0.7	0.14	0.15	0.93	3.8	**SD**
2.9	3.49	2.0	1.35	1.03	3.77	29.0	**highest value**
1.6−	0.27	2.2−	0.67	0.40	0.46−	12.0	**lowest value**

Source: Winstpes v4. 0. 0

Table (02) shows that final estimates free of abilities of person's ranged from (12-29), average ability distribution (1.64) logit, and standard deviation of (0.9) 3) Logit, the standard error of the computational medium of ability estimates was (0.53), a value that is close to what the model assumes, indicating the accuracy of locating persons on the abilities range, and table (03) shows this:

Table 03. Results of analysis free from items difficulty 30 items.

Outfit		Infit					
ZSTD	MNSQ	ZSTD	MNSQ	Model S. E	Measure	Raw Score	Statistics
0.0	0.99	0.0	1.00	0.15	0.00	318.5	**Mean**
1.0	0.15	1.0	0.06	0.05	1.00	54.9	**SD**
2.5	1.37	2.0	1.14	0.36	1.81	405.0	**highest value**
1.8-	0.69	1.8-	0.88	0.11	2.66-	190.0	**lowest value**

Source: Winstpes v4. 0. 0

Table 03 indicates that the average distribution of difficulty was (zero) logit, standard deviation 1.0 logit, and estimates free from items difficulty ranged from -2.66 to 1.81, while the standard error of the computational medium For difficulty estimates, it was 0.15, a rather low value, indicating the accuracy of the difficulty estimates, and the parameters of the Raven Coloured Progressive Matrices test were estimated in its final form 30 items, using the unconditional maximum likelihood method Estimation (UCON), to estimate the most accurate abilities and difficulty, and to reduce the error in estimating the difficulty of items.

Evaluation of Validity: Table 01 shows that the values of statistical fit indexes for the Mean Square Infit Statistics, MNSQ and Mean Square Outfit Statistics, MNSQ of Raven Coloured Progressive Matrices test were close to the ideal position assumed by the model of one value, which ranged from 0.69 to 1.37, with a mean of 1 and a standard deviation of 0.06 for indexes of the Mean Square Infit Statistics, MNSQ, a mean calculation of 0.99 and a standard deviation of 0.15 for indexes of the Mean Square Outfit Statistics, MNSQ. Thus, in the area of product measurement identified by (Wright and Linacre, 1994), they are valued at between 0.70 and 1.30 or between 0.50 and 1.5. These indexes relate to the Validity of the Calibration of test items in their definition of the variable in question. It also relates to the Validity of

the Ability of persons to connect to this variable, which is based on the Validity of persons' responses to the test.

Polarity Items: The results of the analysis showed that point measure correlation PT-MEA CORR values ranged from 0.06 to item A2, and 0.45 item AB11, and therefore all the positive values, indicating that Raven Coloured Progressive Matrices (MPC) measured a single hypothetical composition.

Person-Item Map: To verify the structural Validity of the test, the question that must be answered: Does Raven Coloured Progressive Matrices (MPC) show a serial hierarchy or is it separate? The researchers used Wright's map after the final calibration, and figure (01) shows the results.

Source: Winstpes v4.0.0

Figure 01. Wright map of the Calibrating of Persons' Ability Measures and the item difficulty of Raven test.

Wright's map shows that the items designed for testing tend to be sequential, as the increased difficulty of items hierarchically supports other

research evidence that general abilities (intelligence) form hierarchy. However, there is some overlap in the difficulty thresholds for some items, which indicate that they are not separate, and may be partly due to some items that were not accurate for measuring the ability of intelligence, generally although Raven test items have shown to be a natural hierarchical structure.

Person and item reliability and separation index: Using the Rasch model, the liberal values of both items difficulty and person abilities were obtained, then two types of transactions: Persons Reliability, and items Reliability, Reliability according to item response theory (IRT), means accuracy in estimating the location of both persons and items on the attribute connection, and the accuracy of items in the definition of this caller can be determined by calculating the item separation index (GI). Which is defined as the ratio between the standard deviation of progressive values free from items and the mean standard error of these values.

Table 04. Reliability and separation coefficients and their standard errors of persons and items according to Rasch model.

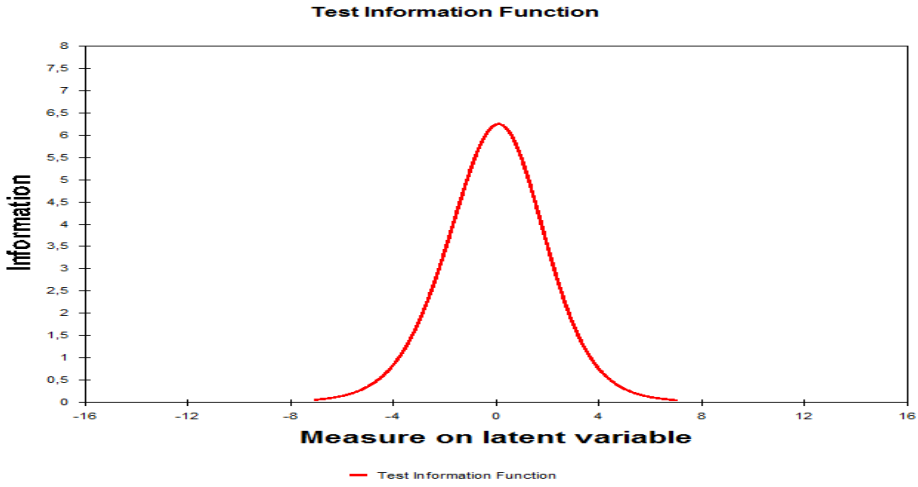
Root-Mean-Square Error (RMSE)		Standard Errors .SD		Separation GP		Reliability	
REAL RMSE	REAL RMSE	Item	Person	Item	Person	Item	Person
0.16	0.56	0.99	0.74	6.33	1.32	0.98	0.64
MODEL RMSE	MODEL RMSE						
0.15	0.55	0.99	0.75	6.39	1.38	0.98	0.65
SEM =2.11				(kr-20)=0.70			

Source: Winstpes v4. 0. 0

Table 04 shows that the value of the items Reliability index was 0.98, and the separation index for the test items was 6.39, which is greater than 2. The person's Reliability index was 0.65, and the value of 1.38 for the person separation index (GP), which is close to the preferred value of the two types. The number of distinct Strata Statistics for both items and persons is determined by the following mathematical formula: H=(4*G+1) /3, where H symbolizes the number of Strata Statistics, and G symbolizes the separation index and compensation in the previous equation for H=(4*1.38+1) /3; As for items H=(4*6.39+1) /3, the number of Strata Statistics of persons amounted to two (2) and nine Strata (9) of items. While the values of the Root Mean Square Error (RMSE) for the true Reliability of persons and items of the separation index for persons and items (REAL RMSE) i.e. the real error variance represents the worst case of reliability or the lowest reliability value of the test, where we recorded the value of 0.56-0.55 respectively, the value of the Mean standard error box root for the model for

persons, is equivalent to the standard error value of measurement in the classical measurement theory of 2.11. (Linacre, 2012, p.316)

Another index of test reliability using the Test Information Function (TIF) is a concept from item response theory (IRT) that helps determine the amount of information provided by item about a person by determining the shortest height of the curve representing that the item information function on the at what ranges of ability, so that any test items that measures the variable to be measured (ability) can be better determined at specific ability levels. The test information function is an amount that is inversely proportional to the standard measurement error.



Source: Winstpes v4. 0. 0

Figure 02. Test Information function (TIF) of Raven Coloured Progressive Matrices after the Calibrating.

Figure (02) shows that the value of the information function for the overall test is as great as possible at the (0) logit ability level, meaning that the test gives more effective information to middle-class persons, while the values of the functions' information provided by the test were as low as possible at high and low ability levels. It was also found that the value of the function's information gradually increases by increasing the ability and reaching its maximum value when the value of the ability is zero logit, or close to zero, reaching (0.03) logit, the value of the standard measurement error (0.4) and the maximum quantity of information (06.24), at the Raw Score (15), i.e. the quantity of information increases by the decrease of the standard error, and thus the value of the test Reliability factor increases.

2. How identical are responses to Raven Coloured Progressive Matrices test items for the One-parameter Logistic Rasch Model?

To give more credibility to the current test, the researchers used the following methods to verify the compatibility of responses to Raven Coloured Progressive Matrices of the Rasch model, which is also an evidence of Validity:

Rasch factor analysis using the Principal Component Analysis method of Residuals (PCAR):

Table .5 Standardized residual variance

Variance	Eigenvalue	Observed (%)	Expected (%)
Total raw variance in observations	37.6195	100 %	100 %
Raw variance explained by measures	7.6195	20.3 %	20.5%
Raw variance explained by persons	3.2114	8.5 %	8.6 %
Raw variance explained by items	4.4081	11.7%	11.8%
Raw unexplained variance (total)	30.0000	79.7%	79.5%
Unexplained variance in 1st contrast	2.3408	6.2%	7.8%

Source: Winstpes v4.0.0

Table 05 shows three criterions, proposed by Linacre, 2008, the first criterion, which is as a rule if the raw variance explained by measures is greater than or equal to 50%, but Lineker has defined the range from 20% to 80% as an acceptable range and is a strong indicator of one-dimensional tool based on the dimensionality and item fit statistics. This was achieved as the contrast index interpreted by raw variance explained by measures, Rasch estimates of the observed value of 20.3%, and the difference between it and the Expected value of the model was only 0.2%. The second criterion, which determines the value of the unexplained variance in 1st contrast explained by the second factor (first in the residual), is less than 10%, which was achieved at 6.2%. The third criterion, which determines the Eigenvalue value of the Unexplained Variance in 1st contrast by the second factor, should not exceed or be below value 3, which was achieved at 2.3408. By verifying these criterions, they can be considered as one-dimensional evidence of Raven's Colored Progressive Matrices Test. (Linacre, 2008, p.272)

Local independence: The values of the Statistical Index (Q3) (Yen, 1984) for the Standardized residual link values of the tested items showed that

they ranged from -0.17 to 0.46, with a mean of 0.24, as suggested by (Chen and Thissen, 1997), ranged from 0.20 and -0.20. Both suggested (Klooster et al, 2008; Davidson et al. 2c004) value 0.5. Some of them used the value of 0.7 as a cutoff (González-de Paz et al., 2014). Overall, these values for the most part did not exceed the level of violation of 0.30 according to (Zenisky and Hammbleton and Sireci, 2002). From there, we can judge that the tested items do not have items that are affected by other items of the same test, which means that the curves of the characteristics of the items do not overlap as well, which is referred to as a local independence (Linacre, 2012, p.405).

3. Does items Raven Coloured Progressive Matrices test show Differential Item Functioning (DIF) depending on the sex variable?

To verify this question, three methods have been used to detect the Differential Item Functioning (DIF) of the item, the difference between the two difficulties (DIF contrast), the Mantel-Haensel method, and the Rasch-welch method, and the results are shown in table 06:

Table 06: Results of Differential Item Functioning (DIF) of Raven test items by gender (GDIF).

Mantel-Haensel		Rasch-welch			DIF contrast	Focal Group	Reference group	Item
P-value	Chi-squared	P-value	DF	t- Test				
0.85	0.033	0.90	280	0.12−	0.09−	2.60−	2.69−	**A1**
0.69	0.160	0.30	211	1.04	0.61	2.30−	1.69−	**A2**
0.55	0.348	0.45	235	0.76	0.34	1.55−	1.21−	**A3**
0.40	0.710	0.20	308	1.29−	0.47−	0.74−	1.21−	**A4**
0.18	1.835	**0.04**	343	2.11−	0.80−	0.67−	1.46−	**A5**
0.66	0.199	0.62	285	0.50−	0.21−	1.19−	1.39−	**A7**
0.91	0.013	0.95	256	0.06	0.02	0.52−	0.50−	**A8**
0.09	2.856	**0.05**	306	1.97−	0.60−	0.10−	0.70−	**A9**
0.17	1.873	**0.05**	299	1.97−	0.57−	0.05	0.52−	**A10**
0.72	0.129	0.86	249	0.18	0.04	1.18	1.22	**A12**
0.02	5.056	**0.02**	197	2.42	0.96	1.42−	0.45−	**AB1**
0.67	0.183	0.45	247	0.75	0.19	0.34	0.53	**AB4**
0.25	1.294	0.06	233	1.90	0.51	0.00	0.51	**AB5**
0.24	1.364	0.10	234	1.67	0.46	0.10−	0.35	**AB6**

0.00	8.470	**0.00**	220	3.35	0.88	0.03	0.91	**AB7**
0.60	0.269	0.49	261	0.69-	0.17-	0.55	0.38	**AB8**
0.59	0.282	0.72	250	0.36	0.09	0.55	0.64	**AB9**
0.61	0.265	0.62	246	0.50	0.12	0.65	0.78	**AB10**
0.62	0.249	0.57	254	0.58-	0.14-	0.99	0.85	**AB11**
0.35	0.885	0.25	242	1.14-	0.27-	2.00	1.73	**AB12**
0.33	0.934	0.52	269	0.65-	0.18-	0.00	0.18-	**B1**
0.05	3.890	**0.03**	266	2.18-	0.52-	1.07	0.54	**B2**
0.86	0.030	1.00	255	0.00	0.00	0.24	0.24	**B3**
0.55	0.348	0.40	244	0.84	0.24	0.16-	0.08	**B4**
0.12	2.387	0.09	228	1.71	0.51	0.46-	0.05	**B5**
0.30	1.052	0.21	236	1.25	0.37	0.39-	0.02-	**B6**
0.48	0.489	0.37	269	0.90-	0.24-	0.25	0.00	**B7**
0.65	0.201	0.72	258	0.36-	0.09-	0.51	0.42	**B9**
0.16	1.946	0.14	256	1.48-	0.35-	1.25	0.91	**B10**
0.45	0.558	0.29	243	1.06-	0.25-	1.96	1.71	**B12**

Source: Winstpes v4.0.0

Table 06 shows that the statistical indicators of Differential Item Functioning of Raven's test items through the values of the non-parametric statistical chi-square for Mantel-Hansel, and the corresponding statistical significance level are lower or equal to 0.05, and that there are two items that show a differentially functioning AB1,AB7, and the same two items also showed differential functioning using the parameter statistics values of Rasch-Welch's T-Welsh test and the corresponding statistical significance level was lower or equal to 0.05, while the latter showed four items that also showed differentially functioning A5, B2, A9, A10, approximately 13% of the items of this test, and by adopting contrast size values based on the size of the difference between the difficulty of the item in the two groups, the reference (female), and the target group or focal (male), most values are lower than the specific value of DIF and limited between -0.5 and 0.5 log except the following items A2 , A5, A9, A10,

AB1, AB5, AB7, B2, B5. Test (T) values are limited to approximately 2± to logit. An extreme item may be deleted, reviewed as a whole, replaced in any weak part of it, may expand or clarify test instructions, or the original item may remain after reassessing the test specifications themselves. The result could be either change to test specifications or clarification of specific inferences that could be reached from test scores.

9. CONCLUSIONS:

The results of the pilot study showed that six items from the IQ test for Raven Coloured Progressive Matrices, misfitting to the Rasch model, and freeing persons were obtained. One of the characteristics of the items, these items enjoyed an acceptable degree of Validity and Reliability, and its Differential Item Functioning (DIF) according to the sex items was non-acceptable, and the number 30 items which can be used to measure the intelligence of primary school students in the Algerian environment to diagnosis and classify in the future.

9.Research proposal:

1.we suggest further studies on this test using other item response theory models (two parameter logistic model, three parameter logistic model).

2.Exploitation of the current test developed in Algerian environment for the purpose of selection and diagnosis in the educational field.

10. References:

Abdul Rahman, Issawi. (2003). Psychological and mental tests and measures, Cairo: Knowledge Facility.

Ali Abbas, Shanan Al-Zamili. (2012). Use the Latent Trait Theory according to Rasch model in the development of raven advanced Progressive Matrices testing on middle school students in Baghdad. Master degree, University of Baghdad, Ibn Rashid Faculty.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. Applied psychological measurement, 41(3), 178-194.

Haidar, Jacobite. (2013). Evaluation and measurement in the educational and psychological sciences - practical vision. Publisher: Morteza Center for Social Development.

Linacre, J. M. (2008). A user's guide to WINSTEPS Rasch-Model computer program, v. 3.65. 0. Chicago, Illinois: Winsteps. com.

Linacre, John, M. (2012). A User's Guide to Winsteps Ministep Rasch-Model Computer Programs. Winsteps.com.

Matore, M. E. M. E., Maat, S. M., Affandi, H. M., & Khairani, A. Z. (2018). Assessment of psychometric properties for Raven Advanced Progressive Matrices in measuring intellectual quotient (IQ) using Rasch model. Asian Journal of Scientific Research, 11(3), 393-400.

Metird, Mona Rabie. (2000). A Psychometric study on the development Raven Coloured Progressive Matrices Test using the Rasch model, the master's thesis of the Faculty of Girls, Ain Shams University of Egypt.

Raven, J. C., & John Hugh Court. (1998). Raven's progressive matrices and vocabulary scales (Vol. 759). Oxford: Oxford psychologists Press.

Safwat, Faraj. (2007). Psychometrics, I6, Cairo: Anglo-Egyptian Library.

Van der Ven, A.H.G.S, & Ellis (2000): A Rasch Analysis of Raven Standard Progressive Matrices Personality and individual Differences NICI Nijmegen Institute For Cognition And Information University of Nijemegen 29 (1).