International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 15, Issue 4, 2021

Standard Progressive Matrices (SPM): Validity and Reliability

Haifa Al-Bokai[a], Ali A. Al-Subaihi[b], [a]Associate professor in Educational and psychological Measurement and Evaluation, Princess Rahmah University College, Al-Balqaa' Applied University, Jordan, [b]Full Professor in Psychological Measurment, College of Education, Taibah University, Madinah Monawwarh, Saudi Arabia, Email: [a]hayfaelbokai@gmail.com, [b]alialsubaihi@yahoo.com

The aim of this research was to study the validity and reliability of Standard Progressive Matrices (SPM) in Jordan. To achieve this goal, the researchers applied the test on a stratified random sample, which consisted of 414 subjects chosen from a selection of secondary schools in Amman, whose ages ranged from fifteen to seventeen years old. The research reached a number of results, the most important for the validity section were: (1) good difficulty and discrimination indexes for the items; (2) significant correlation with the achievement test; (3) a significant difference between the average scores of the normal students and the gifted students when used Contrasted Groups method (4) a significant correlation with the SCAT; (5) a significant difference between the highest 15 percent of scores and the lowest 15 percent of scores when selected from the extremes of the distribution; (6) Confirmatory factor analysis (CFA), which insured the theoretical building for the items according to Spearman's theory of intelligence. Alongside, the results for the reliability section were: (1) internal consistency – with an alpha coefficient of 0.92; (2) split halves – 0.81; (3) test-retest – 0.69 – which showed significant correlation; (4) correlation between each item individually with the total score of the tests, which showed significant correlation for 59 items. The study recommend developing norms for the test, then to use it as a primary step to select the talent.

Key words: *SPM, Secondary School, Confermatory Factor Analysis (CFA), Jordan.*

Introduction

Humans differ from one another in their ability to understand complex ideas, adapt effectively to the environment around them, learn through experience, engage in diverse forms of reasoning, and overcome obstacles through thinking. Although individual differences between individuals can be significant, they cannot be consistent over time. The mental performance of a person will differ on different occasions, in different fields, even if it is judged by different standards. The concept of intelligence is an attempt to represent and organise this complex group of phenomena, which scientists sought to measure through various individual and group intelligence tests. None of us deny the advantages that group intelligence tests are characterised by – that they can be applied to large numbers of subjects at the same time, do not require any special skill in scoring, have an objective score, and can be standardised to large numbers of individuals. This has made group intelligence tests more popular than individual ones especially if their use included scanning or initial filtering. The Standard Progressive Matrices-Raven is one of the most popular group tests for measuring the general intelligence factor in the concept of Spearman. It is also considered one of the most common intelligence tests in research studies and used for broad purposes in educational fields, such as the initial detection of talented people, and their selection, in order to prepare them in special educational programs (Makhail, 2006; Albokai, 2017).

The current research is one of the thousands of studies in Jordan that have used the SPM to study psychometric characteristics, in preparation for standardisation and to develop norms for ages fifteen to seventeen, to be used for many purposes, such as screening for gifted students and measuring the general intelligence of students.

Research Problem:

Intelligence tests are among the tests that have occupied an important part of the modern measurement movement and they still represent one of the most important features of this movement to this day. Many researchers seek to use these tests in their different environments after studying psychometric properties and to develop norms for their local environments, so that they can use them in the many purposes they were designed for. The Standard Progressive Matrices is one of the most popular global intelligence tests that has aroused the appetite of researchers in many parts of the world to study its psychometric properties, and develop norms for the local environment. There are many reasons this test is in the forefront: (1) the number of researches using it, (2) it is one of the culture-free tests; (3) it measures the general intelligence or Spearman's *g* factor; (4) it covers a wide age range (6-80) years; (5) it is one of the group intelligence tests; and (6) for its wide applications in educational, professional and clinical fields.

This test has been standardised in many countries of the Arab and Western world, including the Gulf states, Syria, Jordan, America, Australia, Romania, India and Spain.

Although there are Jordanian studies that have standardised this test, they are either old studies and they used Advanced Progressive Matrices (APM) (Alyan & Al-Samadi, 1989), or they are recent, but they did not use confirmatory factor analysis and the age group is different from the one used in the current research (Al Zugailat, 2009; Al-Majali, 2005). This is what prompted the researchers to study the validity and reliability of the Standard Progressive Matrices (SPM). To the best of the researchers' knowledge, no Jordanian or Arab studies had used confirmatory factor analysis as a prelude to standardise it so as to be used in the initial detection of gifted people in order to administer specific tests for the gifted after having administered SPM (Al-Bokai, et.al, 2014, Carol, et.al, 2010).

Accordingly, we formulated the main question for research – What are the psychometric properties of the Standard Progressive Matrices (SPM)? To answer it, the following sub-questions will be answered:

1. What are the difficulty indexes for the SPM items?
2. What are the discrimination indexes for the SPM items?
3. What are the validation indexes for the SPM? This question is divided into five sub-questions as follows:

    (a.) What are the predictive validity indicators for the SPM using the school achievement criterion?
    (b.) What are the concurrent validity indicators for the SPM using the School and College Ability Test (SCAT)?
    (c.) What are the construct validity indicators for the SPM using the contrasted groups between the average scores of the gifted and non-gifted students?
    (d.) What are the discrimination power indicators for the SPM using the extreme differences between the highest 15% of the test scores and the lowest 15%?
    (e.) What are the factorial validity indicators for the SPM using the confirmatory factor analysis?

4. What are the reliability indexes for the SPM? This question is divided into three sub-questions as follows:
    (a.) What are the internal consistency coefficients for the SPM using Alpha coefficient and split-halves?
    (b.) What is the reliability coefficient for the SPM using the test-retest method?
    (c.) Is there a statistically significant correlation between the test's sub-scores and the total score for the SPM?

The current research seeks to achieve a major and fundamental goal that is embodied in verifying the validity and reliability of the Standard Progressive Matrices SPM on a sample of subjects, male and female, in Jordan, whose ages ranged from (15-17) years.

Previous studies

In this section, studies on psychometric properties will be presented, and only the most recent ones, because although there are numerous studies of SPM, they are not directly related to the objectives of the current research. These studies will be displayed in chronological order from newest to oldest.

The study of Abu Ghaly and Abu Mustafa (2014) in Gaza: The two researchers applied the SPM on a sample of (200) individuals whose ages ranged from (8-18) and were general education students, in order to verify the validity and reliability of the test. Using the test-retest (0.88), internal consistency with an Alpha coefficient (0.88), split-half coefficient (0.88) for the reliability. They verified the validity by using: (a) factorial validity, which results showed that one factor was extracted (66.5%) from the total variance of the matrix by using exploratory factor analysis, and (b) concurrent validity with the illustrated intelligence test (0.79).

Abul-Qasim study (2014) in Sudan: The researcher applied the SPM on a sample of (401) subjects, in order to verify the validity and reliability of the test. He first studied the difficulty indexes for the test items which ranged from (0.11-1.00), and for the reliability, used the internal consistency with an Alpha coefficient (0.85), the split-half coefficient (0.68), correlation coefficients for items with the total score of the test showed statistically significant for the (60) items. He verified the validity by using: (a) the age difference criterion – the results showed statistically significant differences between age groups, and (b) construct validity – in which the researcher studied the differences between the five groups of the SPM, the results showing statistically significant differences between each group and the next one.

Khalifa's study (2011) in Khartoum: The researcher applied the SPM on a sample of (5659) individuals from primary and secondary schools and college students, in order to verify the validity and reliability of the test. For the reliability, he used the internal consistency with the Alpha coefficient ranging from (0.90-0.96), the split-half coefficient ranged from (0.76-0.91), and correlation coefficients for items with the total score of the test showed statistically significant for (56) items out of (60) items. He verified the validity by using: (a) factorial validity, the results showing that one factor was extracted (69.9%) from the total variance of the matrix by using exploratory factor analysis, and concurrent validity between the SPM and chronological age variable with a statistically significant correlation between them.

Abdulgadr's study (2009) in Libya: The researcher applied the SPM on a sample of (280) subjects ages ranged from (8-21) years, in order to verify the validity and reliability of the test. He first studied the difficulty and discrimination indexes for the test items. As the results of the

items' difficulty show there are (80-100%) of the students answering eleven (11) items correctly - seven of the items being from group A, which were easy, and (21-79%) of the students answered correctly (42) of the items, which were of moderate difficulty. Less than 20 percent of them answered seven (7) items, which were very difficult. As for the discrimination coefficients, the results show that there are correlations between the 60 items. For the reliability, test-retest correlation coefficient was (0.90), internal consistency had an Alpha coefficient of (0.94), and split-half coefficient of (0.96). He verified the validity by using (a) factorial validity, the results showing that one factor was extracted (69.41%) from the total variance of the matrix by using exploratory factor analysis, and (b) for correlations between sub-groups, the result shows statistically significant correlations between them.

The study of Yahya, Ibrahim, and Jalal (2007) in Oman: The researchers applied the SPM on a sample of (544) subjects from primary, secondary schools and college students, in order to verify the validity and reliability of the test. For the reliability, the test-retest correlation coefficient ranged from (0.37-0.90), the internal consistency had an Alpha coefficient more than (0.90), correlation coefficients for items with the total score of the test showing as statistically significant, ranging from (0.75-0.90). They verified the validity by using: (1) concurrent validity with the academic achievement score and the language intelligence test, the results showing that there was no statistically significant correlation between the SPM and academic achievement, which ranged between (0.10 - 0.39) due to the weak stability of the academic achievement tests performed by teachers, while the correlation coefficient with the linguistic intelligence test ranged between (0.27 - 0.56); (2.) discriminant validity through (a) the age difference criterion, the results showing statistically significant differences between age groups in favour of older age ones, and (b) gender differences, the results showing statistically significant differences between males and females.

The study of Rahmah (2004) in Syria: The researcher applied the SPM on a sample of (289) individuals from primary to secondary schools with ages ranging from (8-18), in order to verify the validity and reliability of the test. She first studied the difficulty indexes for the test items, which the results showed the difficulty coefficients of less than 0.50 increase from one group to another, so the first group is the easiest and the second group is more difficult and so on, until we reach to the last one, which is the most difficult one. For the reliability, she used the test-retest (0.69), internal consistency with Alpha coefficient (0.87), split-half coefficient (0.88), correlation coefficients for items with the total score of the test showing statistically significant for (49) items out of (60) items. She verified the validity by using: (a) factorial validity, the results showing that one factor was extracted (68.9%) from the total variance of the matrix by using exploratory factor analysis, concurrent validity with the Mill-Hill Vocabulary Scale and Culture Fair Intelligence Test with a significant correlation between the SPM and them, the validity of the opposing groups between normal and gifted with significant differences between the two groups in favour of the gifted one.

Al-Nefaie's study in (2001) in the Kingdom of Saudi Arabia: The researcher applied the SPM on a sample of (120) subjects, in order to verify the validity and reliability of the test. He first studied the difficulty and discrimination indexes for the test items, with the results showing that the average of difficulty coefficients was (0.51) and discrimination coefficients were (0.49). For the reliability, he used the test-retest (0.90), internal consistency with an Alpha coefficient of (0.88), split-half coefficient (0.77). He verified the validity by using: (a) factorial validity, the results showing that one factor was extracted (65%) from the total variance of the matrix using exploratory factor analysis, while concurrent validity with Advanced Progressive Matrices (APM) showed significant correlation (0.69).

Research instrument:

This research used Standard Progressive Matrices (SPM) which is considered one of the Culture Free Tests, based on excluding the language factor in measuring intelligence. Consequently, it was found to be a valid tool for measuring the intelligence of individuals from different cultures and civilisations (Mikhail, 2006a). SPM is the result of the work of John Raven and his assistant Penrose, who was first published in 1938 in England (Raven, Raven & Court, 1998), with the aim of measuring Spearman's *g* factor or general intelligence, which is summarised in the "ability to perceive relationships and belongings" (Mikhail, 2006a, p. 491). In the eyes of English psychologists, this test is considered one of the best tools available to measure the general factor, which has been widely spread unrivaled in Britain and around the world (Makhail, 2015). According to the test manual (1998), the number of studies conducted about this test since its publication in 1938, up until the 1998, exceeded 2,500 studies, considered one of the most common intelligence tests.

The test itself consists of groups of successive shapes, each associated with a relationship, and presented with one blank cell. The subject has to choose from the given alternatives the one that is consistent with other forms after discovering the relationship between them, in order to fill the empty cell. The difficulty of the test increases gradually, as it begins with the easy items that depend on accuracy in distinguishing, and ends with the difficult items that depend on the perception of abstract relationships. This test is one of the maximum performance tests in which it does not need to specify the time given, it can be applied individually or grouply, and its application requires only some simple verbal instructions (Muhammad, 2017; Eid, 2005; Anastasi & Urbina, 1997).

SPM is the appropriate image for normal individuals whose age ranges between (6-80) years, and it is the main image of the test (Anastasi & Urbina, 1997; Abdalgadr, 2009). SPM consists of 60 items distributed into five subgroups, A, B, C, D and E, and each group consists of twelve (12) items arranged according to the principle of gradation in difficulty for each group and the five groups. The test items are a set of sequential patterns that are related to each other in a relationship. A parallel version appeared in 1998, in which the discrimination power increased for the higher category of individuals (Mikhael, 2015).

This test is usually administered as a strength test with no time limits, it is also administered in certain cases as a speed test. It can be given individually or in groups. Very simple oral instructions are required.

There are two other forms of Raven Tests: Coloured Progressive Matrices (CPM) and Advanced Progressive Matrices (APM).

Methodology:

This research followed the descriptive approach, using survey studies, considered one of the most appropriate approaches in terms of administering the test to the sample, analysing the data, and interpreting it in light of previous studies and the theoretical structure of the test.

The research population consisted of all secondary school students, male and female, in the city of Amman, with ages ranging between (15-17) years for the 2018-2019 academic year.

The sample consisted of (414) subjects drawn by the stratified random sampling method, considered one of the most suitable types of samples for the current research because the population is composed of two classes (males and females), with ages ranging between (15-17) years, drawn from a number of public schools affiliated to the city of Amman.

Research Finding and discussion:

Before starting to review the research results regarding the validity and reliability of the SPM, it must be noted that the SPM was used after translating its manual from English into Arabic. Since the SPM is a culture-free test, there was no need to modify its items, so it was administered as is.

The SPM was administered to a pilot sample of (54) male and female students from the tenth to the twelfth grade of secondary school, with ages ranging from (15-17) in the city of Amman. This was done to ensure the clarity of the instructions, to know the administration difficulties and attempt to overcome them as much as possible, and to calculate the time required by the subject to answer the test. The initial results showed the clarity of the test for the students, their understanding of what was required of them. There was therefore no need to make any modifications to the items.

In this section, the results of the research questions will be presented and discussed with consideration to previous studies.

1. What are the difficulty indexes for the SPM items?

It must be noted that the difficulty indexes are the percentage of correct answers for this item. The researchers extracted the difficulty coefficients for each of the (60) items for the SPM as shown in the following table.

Table (1) Difficulty indexes for each item of SPM

Group A	Difficulty index	Group B	Difficulty index	Group C	Difficulty index	Group D	Difficulty index	Group E	Difficulty index
A1	0.97	B1	0.96	C1	0.89	D1	0.93	E1	0.69
A2	0.98	B2	0.95	C2	0.84	D2	0.89	E2	0.77
A3	0.94	B3	0.93	C3	0.86	D3	0.84	E3	0.59
A4	0.97	B4	0.73	C4	0.70	D4	0.78	E4	0.58
A5	0.94	B5	0.79	C5	0.84	D5	0.87	E5	0.49
A6	0.97	B6	0.68	C6	0.72	D6	0.82	E6	0.41
A7	0.87	B7	0.66	C7	0.77	D7	0.60	E7	0.40
A8	0.86	B8	0.78	C8	0.61	D8	0.69	E8	0.25
A9	0.94	B9	0.75	C9	0.75	D9	0.57	E9	0.15
A10	0.84	B10	0.83	C10	0.41	D10	0.5	E10	0.11
A11	0.68	B11	0.69	C11	0.37	D11	0.21	E11	0.07
A12	0.61	B12	0.59	C12	0.10	D12	0.10	E12	0.07

It is clear from the previous table that the difficulty indexes in group (A) range from (0.61-0.98), in group (B) from (0.59-0.96), in group (C) from (0.1-0.89), in group (D) from (0.1-0.93) and in group (E) from (0.07 - 0.69), while the mean difficulty indexes for all items was (0.66).

When analysing the previous indexes, it becomes clear that the first items in each group were easy items, and their difficulty increases within one group and the difficulty of items increases from one group to another. Group (A) is easier than group (B), group (B) is easier than group (C), group (C) is easier than group (D), and group (E) is the most difficult group. That is, the first items in any group were easier than the last items in the group that preceded it. This is consistent with the construct structure of the test items, which assumes that the difficulty of the items increases gradually in one group and between groups. It begins with the easy items that require simple perceptual pairing. It depends mainly on accuracy in distinguishing and ends with difficult items that require the perception of abstract relationships related to shape, direction, or a number, and which change in complex ways (Mikhail, 2006a referred to in Thorndike and Hagen, 1989). The results of the current research are consistent with the results of these studies by Abu Al-Qasim (2014), Abdel-Qader (2009), Rahma (2004) and Al-Nefaiei (2001).

2. What are the discrimination indexes for the SPM items?

It must be noted here that the discrimination indexes is the difference between the percentage of students who answered the item correctly from the higher category, and the percentage of

students who answered the item correctly from the lower category (Al-Manisel & Al-Atoum, 2010).

The researchers extracted the discriminant coefficients for each of the sixty items for the SPM, as shown in the following table.

Table (2) Discrimination indexes for each item of SPM

Group A	discrimination index	Group B	discrimination index	Group C	discrimination index	Group D	discrimination index	Group E	discrimination index
A1	0.11	B1	0.11	C1	0.35	D1	0.20	E1	0.60
A2	0.09	B2	0.11	C2	0.42	D2	0.34	E2	0.58
A3	0.12	B3	0.21	C3	0.42	D3	0.45	E3	0.63
A4	0.10	B4	0.57	C4	0.53	D4	0.48	E4	0.71
A5	0.15	B5	0.49	C5	0.40	D5	0.38	E5	0.68
A6	0.08	B6	0.63	C6	0.38	D6	0.41	E6	0.51
A7	0.28	B7	0.61	C7	0.56	D7	0.56	E7	0.63
A8	0.24	B8	0.58	C8	0.60	D8	0.56	E8	0.45
A9	0.18	B9	0.60	C9	0.53	D9	0.67	E9	0.18
A10	0.40	B10	0.43	C10	0.53	D10	0.61	E10	0.17
A11	0.49	B11	0.59	C11	0.38	D11	0.40	E11	0.04
A12	0.25	B12	0.57	C12	0.15	D12	0.16	E12	0.07

It is clear from the above table that the discrimination indexes in group (A) range from (0.1-0.49), in group (B) from (0.6-0.63), in group (C) from (0.4-0.56), in group (D) from (0.2- 0.67) and in group (E) from (0.07 - 0.71), while the mean discrimination indexes for all items was (0.39). According to Hopkins (1998), items with a discrimination index of 0.40 and above have excellent discrimination, items whose indexes range between (0.30 - 0.39) have a good discrimination, items whose indexes range between (0.10 - 0.29) have a medium discrimination, and items whose indexes range between (0.01 - 0.10) have a weak discrimination. A negative index is an indication of a wrong or ambiguous item (Abdalgadr, 2009).

Returning to the previous table, we notice that there are no negative items, but there are (12) items that are weakly distinguished, and these are either very difficult items as most of them were in the last items of each group, or very easy items, as most of them were in the first items of each group (as noted in the table) which indicates that very difficult or very easy items exhibit little variation and do not fulfill the differential function (Mikhael, 2006). This confirms the theoretical assumption of building SPM by making the very easy items at the beginning of each group, and the difficult items at the end of each group. With reference to the previous table, there are (14) items of moderate distinction, (5) items of good distinction, and (29) items of excellent distinction. The results of this research are consistent with those of Abdul Qadir (2009) and Al-Nefai'i (2001).

3. What are the validation indexes for the SPM?

A number of methods have been used to verify the validity, criterion validity with two criteria, school achievement test and School and College Ability Tests (SCAT), contrasted groups, extreme groups and factorial validity.

In order to answer this question, the sub-questions related to the methods used to verify validity must be answered, which is next:

(a.) What are the predictive validity indicators for the SPM using the school achievement criterion?

Validity was studied in terms of an external criterion, in this case academic achievement, which refers to the students' scores in the first semester of the 2018/2019 academic year.

To answer this question, the researchers studied the correlation between SPMs' scores and school achievement scores for a sample of (206) students, by the Pearson method. The following table shows the result of the correlation.

Table (3) Correlation coefficient between school achievement and SPM

Test	No.	Mean	S.D	***Pearson***	P. value
School achievement	206	75.08	10.81	0.49**	0.00
SPM	206	39.75	9.72		

It is noticed from the above table that the correlation coefficients between the scores' of SPM and academic achievement are statistically significant at ($\alpha = 0.01$), which indicates the validity of the SPM with the school achievement criterion. This result is consistent with the findings of Rahma (2004), but it differs with those findings from Yahya, Ibrahim, and Jalal (2007).

(b.) What are the concurrent validity indicators for the SPM using a School and College Ability Test (SCAT)?

A School and College Ability Test (SCAT) was used to study its correlation with the SPM in a previous study (Al-Bokai, et.al, 2014), with (204) subjects. The results showed a statistically significant correlation at 0.01 for the verbal, quantitative and total score of the SCAT with the total score of SPM, as shown in the following table:

Table (4) Pearson corrolation between SCAT & SPM

SCAT(204)	Total Score	Verbal section	Quantitative section
SPM(204)	0.56**	0.45**	0.55**

(Albokai, et.al, 2014)

(c.) What are the construct validity indicators for the SPM using the contrasted groups between the mean scores of the gifted and non-gifted students?

To verify the validity in terms of contrasted groups, the researchers used the (T)Student test for independent samples to know if there is a significant difference between the mean scores of the gifted and non-gifted students. The gifted students were selected from Abdullah II School for Excellence in the city of Salt. As shown in the following table.

Table (5) T. students between gifted and non-gifted groups

Students	**No.**	**Mean**	**S.D**	**T**	**P. value**
gifted	85	49.81	6.325	9.07	0.00
Non-gifted	206	39.54	9.60		

It is noticed from this table that there are statistically significant differences between the mean scores of normal students and the mean scores of gifted students at SPM at a level of significance ($\alpha = 0.00$) in favour of gifted students, which indicates the validity of the SPM in terms of contrasted groups. This result is consistent with the findings of Rahma (2004).

(d.) What are the discrimination power indicators for the SPM using the extreme differences between the highest 15% of the test scores and the lowest 15%?

To verify the validity in terms of extreme groups, the researchers used the (T)Student test for independent samples to discover if there was a significant difference between the mean scores of students who obtained the highest 15% and the mean scores of students who scored the lowest 15%, as shown in the following table.

Table (6) T. student between highest and lowest groups

Group	No.	Mean	S.D	T	P. value
Highest 15%	79	51.21	1.99	35.93	0.00
Lowest 15%	62	19.11	7.62		

It is noticed from the above table that there is a statistically significant difference between the mean scores of the highest 15% of students and the mean scores of the lowest scores of 15% in favour of the highest group, which indicates the validity of the SPM in terms of extreme groups.

(e.) What are the factorial validity indicators for the SPM using the confirmatory factor analysis?

Factorial validity was verified using confirmatory factor analysis (CFA), which the researchers set out from the theoretical structure of SPM using CFA to ensure the validity of its theoretical structure and to ensure the data matching it.

Before performing the CFA, the researchers made sure of the suitability of the sample size for analysis, which was based on the approval of specialists who advised that the ratio between the number of individuals to the number of items not be less than (5) individuals for each item (Tighza, 2012). Applying this to the tool used in the current research: the appropriate number is then (5 x 60 = 300) individuals, while the number of the sample approved by the researchers is (414) subjects. Thus the sample size is sufficient for CFA.

The researchers used the one-dimensional confirmation factor model using AMOS software, in which the researchers assume that a particular concept involves one factor that the items have sufficiently common in this concept; the relationship area between the items represents the theoretical significance of the concept (Tighza, 2012).

The results are as shown in table (7), the saturation of each of the (60) items of the SPM.

Table (7) the saturations of each of the sixty (60) items of the SPM

A	.537	d10	.503	ee9	.121	b12	.471
B	.870	d9	.492	ee10	.178	d12	.114
C	.945	d8	.538	ee11	.097	c2	.630
D	.900	d7	.506	ee12	.114	c3	.742
E	.712	d6	.457	b1	.137	c4	.553
a1	.740	d5	.744	b2	.347	c5	.636
a2	.884	d4	.592	b3	.582	c6	.455
a3	.691	d3	.458	b4	.593	c7	.469
a4	.754	d2	.712	b5	.378	c8	.499
a5	.662	d1	.307	b6	.591	c9	.589
a6	.664	ee2	.664	b7	.580	c10	.346
a7	.490	ee3	.627	b8	.673	c11	.294
a8	.141	ee4	.411	b9	.626	c12	.132
a9	.579	ee5	.614	b10	.622	c1	.650
a10	.473	ee6	.549	b11	.555		
a11	.273	ee7	.499	ee1	.627		
d11	.317	ee8	.426	a12	.155		

Figure (1) also shows the theoretical model for the SPM, which includes the five groups and their items as follows:

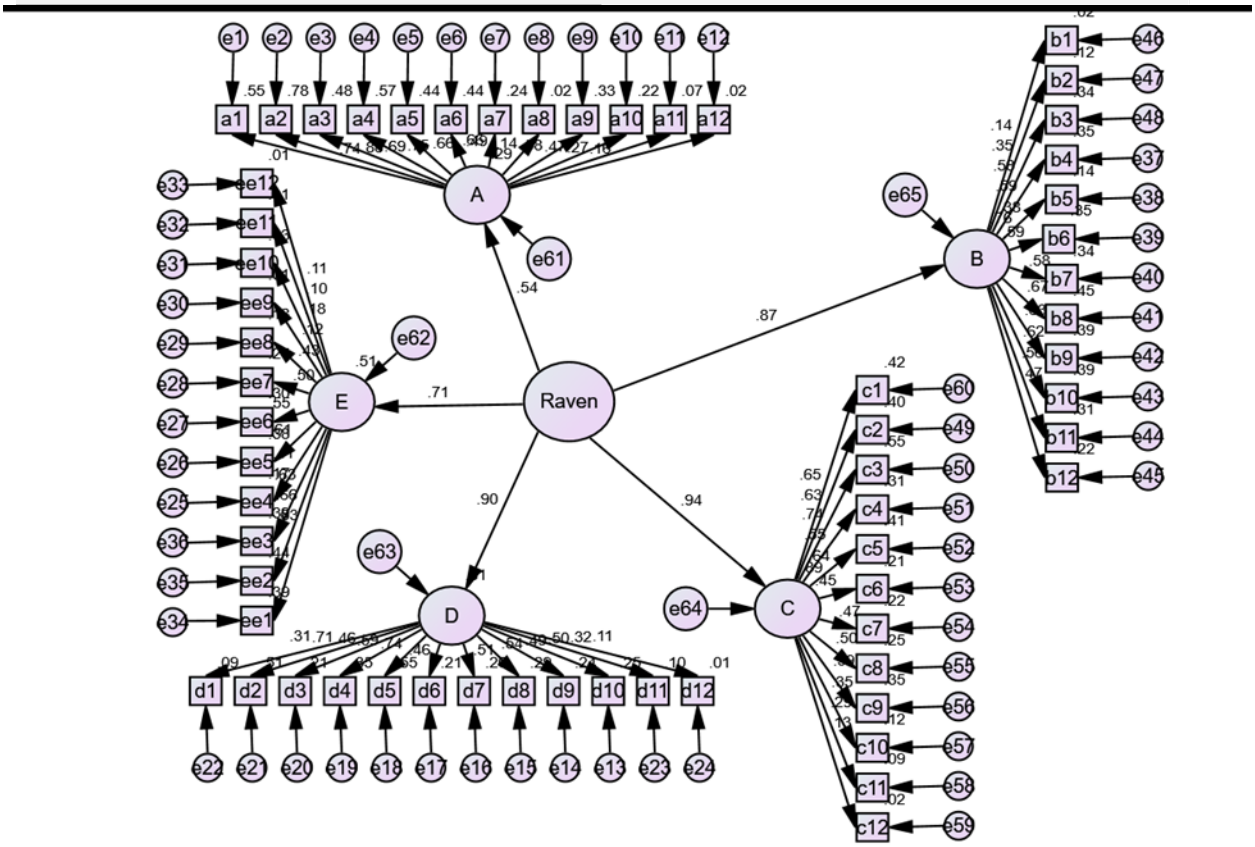

Figure (1) the theoretical model for the SPM

To test the confirmatory factor model, the confirmatory indicators must be estimated based on the approved classifications and the most commonly reported indicators. This is according to the study of Hooper, Coughlan, and Mullen (2008), which relied on numerous studies of conformity indicators for CFA, such as the studies of Hu and Bentler (1999), Boomsma (2000), Mac Donald and Ho (2002) and Kline (2016), all of which confirmed the effectiveness of five indicators more than others. These five indicators are: Chi-square Mean/ Degree Freedom (CMIN/DF), Comparative Fit Index (CFI), The Root Mean Square Error of Approximation (RMSEA), Standardised Root Mean Square Residual (SRMR) and Parsimony-Adjusted Normed Fit Index (PNFI). Specifically, we will mention the following indicators in Table (8)

Table (8) The theoretical indicators and the calculated value in the current study

Indicator name	**The value to judge the quality of the model**	
	Theoretical value	**Calculated value**
CMIN/DF	Less than 3	1.77
CFI	more than 0.90	0.82
(RMSEA)	Less than 0.08	0.04
SRMR	Less than 0.08	0.06
PNFI	more than 0.6	0.63

By reading the above table data, it becomes evident that all the values are in agreement, with the exception of the CFI value, which did not reach the theoretical limit. The practice in factor analysis is the scarcity of achieving all indicators, taking into account that constructive validity is only one of the types of validity. It can be said that the theoretical structure of the test was verified according to the CFA, which confirmed the existence of one factor of intelligence according to Spearman's theory, that of the G factor.

The researchers did not find any Arabian studies that verified the theoretical structure of SPM using confirmatory factor analysis, rather, all Arabian studies discovered the theoretical structure of the SPM using the exploratory factor analysis (Abu Ghaly and Abu Mustafa, 2014; Abdul Qadir, 2009; Al-Nafiei, 2001; Rahma, 2004; Khalifa, 2011). Unlike foreign studies, the researchers found one study (Lynna, Allikb & Irwingc, 2004) that has results that agree with the results of the current research.

4. What are the relability indexes for the SPM?

To answer this question, the sub-questions regarding the methods used to verify reliability must be answered, which is next:

(a.) What are the internal consistency coefficients for the SPM using Alpha coefficient and split-halves?

The test reliability was calculated by Guttman split-half Coefficient and Cronbach alpha coefficient, as shown in the following table.

Table (9) Cronbach alpha coefficient & split-half coefficient

	SPM items (60)
Guttman split-half coefficient	0.813
Cronbach alpha coefficient	0.926

According to the rule that classifies the internal consistency of the scale according to the Cronbach alpha value, which states that consistency is acceptable at (α= 0.70) or less, it is good at α value of between 0.80 and 0.90, and it is excellent at (α>0.90) (Nunnally & Bernstein, 1994). The internal consistency of the scale used in this research is "excellent" for the alpha coefficient and "good" for the split-half coefficient. These results agree with all the results of previous studies presented in this research.

(b.) What is the reliability coefficient for the SPM using test-retest method?

Reliability was calculated by the test-retest method by re-administring the test to the students after a period of time ranging from 7-12 days. The number of students was (289). The researchers used the Pearson correlation coefficient, as shown in the following table:

Table (10) Pearson coefficient for test-retest reliability

	Mean	S.D	Person cofficient
First administered	40.67	10.08	0.69**[1]
Second administered	42.73	10.69	

As shown in the above table, there is a significant correlation at 0.01, which indicates the stability of the test over time. This result is consistent with the results of Rahma (2004) and Abdelkader (2009).

(c.) Is there a statistically significant correlation between the test's sub-scores and the total score for the SPM?

The following table shows the correlation coefficients of the items with the total score:

Table (11) Pearson corrolation between sub-scores and the total score

Group A	Pearson coefficient	Group B	Pearson coefficient	Group C	Pearson coefficient	Group D	Pearson coefficient	Group E	Pearson coefficient
A1	0.423**	B1	0.125*	C1	0.605**	D1	0.251**	E1	0.529**
A2	0.426**	B2	0.363**	C2	0.558**	D2	0.632**	E2	0.598**
A3	0.356**	B3	0.534**	C3	0.664**	D3	0.412**	E3	0.523**
A4	0.353**	B4	0.557**	C4	0.491**	D4	0.551**	E4	0.338**
A5	0.395**	B5	0.360**	C5	0.566**	D5	0.646**	E5	0.427**
A6	0.328**	B6	0.512**	C6	0.447**	D6	0.391**	E6	0.393**
A7	0.440**	B7	0.518**	C7	0.425**	D7	0.478**	E7	0.416**
A8	0.312**	B8	0.565**	C8	0.498**	D8	0.498**	E8	0.350**
A9	0.480**	B9	0.561**	C9	0.539**	D9	0.505**	E9	0.136**
A10	0.502**	B10	0.544**	C10	0.371**	D10	0.499**	E10	0.169**
A11	0.424**	B11	0.503**	C11	0.298**	D11	0.351**	*E11*	*0.072*
A12	0.201**	B12	0.460**	C12	0.145**	D12	0.164**	E12	0.137**

As can be noted from the above table, there are 59 items out of 60 items that were statistically significant to the total score, which indicates the internal consistency of the test. This result differs from the result of Rahma's study (2004), which showed that 49 out of 60 items were related; she attributed it to the weak correlation of easy and difficult items with the total score, and her sample ages were younger than the age of the sample used in the current research.

** at 0.01 [1]

Conclusion:

In conclusion, it can be said that the Standard Progressive Matrixes preserved the culture-free property, and has good psychometric properties, which have been confirmed by studying its validity and relability in many ways, as is shown in the findings section in this research.

Recommendations:

The current research recommends the following:

1. Standardise the Standard Progressive Matrixes SPM on a sample from the school secondary level, because it is important for identifying talented students and selecting them to be placed in suitable educational programs.
2. Extensive studies should be carried out on the diagnostic capacity of the SPM.
3. The use of the SPM in the diagnosis of mental disabilities.
4. The use of the SPM as a first stage in the detection of autism.

REFERENCES

Abdalgadr, A. (2009). Standardisation of Raven's Standard Progressive Matrices Test for a Libyan Sample. Unpublished doctoral theses. University of Salford, Salford, UK.

Abu Al-Qasim, M. (2014). Standardisation SPM of John Raven to measure general Intelligence, on a sample of students who are gifted, intellectually talented, and have learning difficulties, in the state of Khartoum. Unpublished PhD thesis. Umm Durman Islamic University. College of Arts. Sudan.

Abu Ghali, E. & Abu Mustafa, N. (2014). Standardisation SPM of Raven For the age group (8-18) years for general education students in the governorates of Gaza. Taibah University Journal for Educational Sciences. Volume 9, Issue 1, pp. (90-108).

Al-Bokai, H. (2017). A Gifted Education Model for Jordan: A study of Johns Hopkins University Center for Talented Youth. Dirasat, Educational Sciences, Volume 44, No 4, Supplement 1, pp.413-427. Jordan University.

Al-Bokai, H., Al-Omari, J. and Al-Rahmaneh, A. (2014). Development of the School and College Ability Tests advanced form in Jordan. The Educational Journal, Ain Shams University, Issue 38, Part 1, Egypt.

Alian, Kh. & Smadi, J. (1989). Mental performance standards for Jordanian individuals who Their ages more than 11 years old in Advanced Progressive Matrices (APM). Dirasat. The University of Jordan, Issue 15 (8), pp. 107-132.

Al-Majali, M. (2005). Derivation of the Psychometric Properties of SPM, Standard Level for the age group 12-17 years, unpublished master's thesis, Muta University.

Al-Manizel, A. & Al-Atoum, Y. (2010). Research methods in educational and psychological sciences. Sharjah: University Library.

Al-Nefaie, A. (2001). Standardisation of the Advanced Progressive Matrices (APM) for intermediate and secondary stages' students in Makkah Al-Mukarramah region. unpublished master's thesis, faculty of Education, Umm Al Qura University.

Anastasi, A. &Urbina, S. (1997): Psychological Testing. Seventh Edition, Prentice- Hill, Inc, U.S.A.

Eid, Gh. (2005). Evaluate the structure the Advanced Progressive Matrices (APM) and the Short Form of it using factor analysis and the Rush model. Journal of Psychological and Educational Research, 3rd Issue, pp. 256-283.

Elst, W., Ouwehand, C., Rijn, P., Lee, N., Boxtel, M., and Jolles, J. (2013). The Shortened Raven Standard Progressive Matrices: Item Response Theory–Based Psychometric Analyses and Normative Data. Assessment 20(1) 48 –59. SAGE publication.

Hooper, D., Coughlan, J. and Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. The Electronic Journal of Business Research Methods. Volume 6 Issue 1 2008, pp. 53 - 60, available online at www.ejbrm.com.

Khalifa, H. (2011). Standardisation of the SPM in the state of Khartoum. Psychological Studies - Al-Basrah Center for Research, Consulting, and Educational Services - Issue 5, Pp. 41-76, Algeria.

Kline, R. B. (2016). Methodology in the social sciences. Principles and practice of structural equation modeling (4th ed.). Guilford Press.

Lynna, R., Allikb, J. & Irwingc, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. Intelligence 32 (2004) 411 – 424.

Michael, M. (2006). Measurement and evaluation in modern education. Fifth edition. Damascus University Publications.

Michael, M. (2006a). Psychometric measurement. First part. Damascus University Publications.

Michael, M. (2015). Psychological and Educational Measurement and Evaluation of the normal and special needs. Al-Easar Al-Elmi House for publishing and distribution. Jordan.

Mills, C., Ablard, K., Brody, L. (1993). The raven's progressive matrices: Its usefulness for identifying gifted/talented students. Roeper Review, V.15-Issue 3: Longitudinal Studies in Gifted Education.

Muhammad, A. (2017). Children's performance scores on the SPM to measure intelligence. A field study of the El Fasher and Nyalas' cities. Journal of Educational and Psychological Sciences, Palestine. Volume One- Issue 1, pp. 127-143.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York, NY: McGraw-Hill

Rahma, A. (2004). The effectiveness of using time series analysis and regression analysis in studying intelligence in individuals from seven to eighteen years old. Unpublished PhD thesis. Faculty of Education. university of Damascus.

Rasmussen, D. (2010). A neural modelling approach to investigating general intelligence. Unpublished Master thesis. Waterloo, Ontario, Canada.

Raven, J., Prieler, J., & Benesch, M. (June 15, 2005). A Replication and Extension of the Item-Analysis of the Standard Progressive Matrices Plus, together with a Comparison of the Results of Applying Three Variants of Item Response Theory. WebPsychEmpiricist. Retrieved June 15, from http://wpe.info/papers_table.htmlCorres.

Raven, J., Raven, J. C., & Court, I. H. (1998). Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview, Oxford Psychologists Press Ltd.

Tigza, B. (2012). Exploratory and confirmatory factor analysis. Dar Al-Maseera, Jordan.

WebPsychEmpiricist (2017): Retrieved September 27, from http://wpe.info/papers_table.html.

Yahya, A., Ibrahim, A. & Jalal, S. (2007). Standardisation of the SPM In the Omani environment. Childhood studies. January 2007. pp. 35-46.

Zgilat, A. (2009). Standardisation of the SPM - Standard Level for Jordanian universities' students, unpublished master's thesis, Muta University.