RAVEN'S PROGRESSIVE MATRICES TEST: SCALE CONSTRUCTION AND VERIFICATION OF "FLYNN EFFECT"

Lilia Rossi Casé; Rosa Neer; Susana Lopetegui

Abstract

In this paper, the scales of Raven's Progressive Matrices Test, General Scale and Advanced Scale, Series II, for the student population (third cycle of *EGB* and *Polimodal*) in the city of La Plata are presented. Considerations are made as regards both the increase in scores (Flynn effect) observed in relation to the previous scale (1964) and the different mean scores according to two age groups (13-16 and 17-18 years of age) and education mode. The findings enabled inferences related to the significance of the increase, particularly in the case of the higher scores in the population attending a special kind of educational institution.

Key Words: Progressive matrices test; percentile norms; flynn effect

Introduction

This paper deals with the updating of test norms or scales; in this case, those corresponding to Raven's Progressive Matrices Test, and the increase of the mean scores observed in the samples of the individuals examined, belonging to successive generations.

Raven's Progressive Matrices Test –widely acknowledged for its excellence as a measuring instrument– was first published in 1938 by its author, John C. Raven, a student of the English psychologist Charles Spearman, whose work was inspired in Galton and Wundt. In 1904, Charles Spearman stated his "eclectic two-factor theory" using factor analysis as method. Such theory states every single skill in man has a common factor, general to all of them (G factor) and a specific factor for each of them (S factor). Both these factors –in different proportion– are present in each skill.

J. Guilford, author of the Three Dimensional Structure of Intellect Model (SI), said: "No event in the history of mental tests ever bore greater significance than Spearman's statement of theory of the two factors". Further research acknowledged other factors called group factors.

John Raven constructed his test on this theoretical basis. There follows a brief description of its main characteristics:

- It is a test of intellectual capacity, of general mental ability.

- It is a factor test, it evaluates the component of the G factor, the eductive capacity, i.e. to make sense out of confusion, to shape variables, to go beyond that which is given or obvious.
- It is a non-verbal test.
- It is a multiple choice test.
- It is a test of multiple choice lacunar matrices. The individual's task is to complete series of drawings in which the last drawing is missing, selecting the appropriate one from a range of possible choices.

It has different presentations: General Scale, Special Coloured Scale (Booklet Form and Board Form), Advanced Scales Series I and II. The author also devised a Vocabulary verbal test two versions of which are available: one American and one British.

Since its creation, the test has been widely used in its country of origin as well as in other countries. The test has been revised several times, to obtain in most cases norms or scales suitable to different populations.

In the last version, important considerations have been made concerning general capacity, educative capacity, reproductive capacity, "G" factor, general intelligence, problem-solving and competences, in order to determine the psychological scope each of these terms has. This allows a more appropriate interpretation of test results in the clinical, educational and labour settings. The variations made among the norms corresponding to a certain percentile, which vary the demand depending on the scale used, follow this same pattern.

A special aspect to be taken into account is that the mean scores obtained in the intelligence tests regularly and remarkably increase all over the world. This phenomenon has been observed even since the appearance of the tests. In 1984 James Flynn summarised the research results on this subject; thus, "Flynn effect" has come to be the name given to the permanent increase of absolute values in all tests important for all ages.

The Flynn effect is quite noticeable in G- loaded tests, as is the case in the Progressive Matrices Test. The author carried out studies on the Raven in Holland and then in other countries and always noticed the same trend: the increase of mean scores over time. The score increase rate of G-loaded tests doubles that of the so-called "omnibus" tests, such as Wechsler Scales (WAIS and WISC).

One of the immediate consequences of this verification is that norms become increasingly more demanding, i.e., in order to obtain the same transformed score, direct scores should

be higher; therefore, more problems should be solved. The cause of this phenomenon is still unknown. There is not enough evidence to assert that said fact may show an actual increase in intelligence over time. Said increase may probably be linked to factors such as more appropriate feeding or nutrition which, in turn, lead to increased psychological development; the progressive acquisition of certain skills to perform satisfactorily in tests; an improvement in schooling or children's education; or an increased use of technology in culture, which provides more stimuli. Though probably none of them may be ruled out –for they all may contribute to the increase in mean scores– one should not overlook the last of these interpretations.

The analysis of the results rendered by the *Primer Operativo Nacional de Evaluación de Finalización del Nivel Secundario 1997* (First National Assessment Campaign of Completion of the Secondary Level 1997), allowed us to notice that in Maths 68% of the students satisfactorily solve those problems entailing the interpretation of information presented in graphs, whereas in those problems requiring the solution of equations and calculus of probability the percentage of students doing well drops to 53% and 40% respectively.

These issues, as well as the need to update the norm or scales of the tests used in the professional work, provided the grounds for carrying out the research the results of which are presented herein.

The norms or scales of Raven's test –expressed in percentiles– currently available in our setting correspond to the original results provided by the author and to the different national and regional groups, examined in different periods.(1)

In order to carry out the research, the authors of this paper took into account on the one hand the importance of Raven's Progressive Matrices Test as an instrument for psychological evaluation and its widespread use for a variety of purposes in different settings where psychology is applied and, on the other hand, the fact that the norms of psychological tests are in no way absolute, universal or permanent (Anastasi, 1989), but they merely represent the performance in the test of the subjects making up the normative group. Thus, in 1995 we began to carry out a study whose goals were:

1. To construct percentile norms for the city of La Plata which provide test users with updated scales, adapted to one's own population, using the General and Advanced Scales, Series I and II, corresponding to the year 1993. The former presents modifications in the placement of matrices and the arrangement of series and the latter is an original enlargement, introduced in the country for the first time.

2. To compare scores obtained by the examinees in the study carried out between 1995-1998 with those published for the city of La Plata in 1964. Obviously, the data of 1964 correspond only to the General Scale.

3. To compare the mean scores obtained in the General Scale and the Advanced Scale.

4. Compare the mean scores of subjects taking part in different educational modes, in order to observe significant differences among them.

The target population was made up by the overall student population attending different educational institutions in La Plata. Special attention was given to the group of students attending the *Bachillerato de Bellas Artes con Orientación Plástica* (Fine Arts Baccalaureate, with major in Painting) to detect possible differences in their performance in relation to other groups, given the perceptive nature of the test and the particular interest these students have in creative works, which use combinations of colours, shapes and designs.

Participants and methods

Participants

A sample made up by 988 subjects of both sexes, aged 13-18, was selected: The subjects attended different secondary educational institutions and covered the whole spectrum of students in the different secondary schools of the city of La Plata. Tables 1 and 2 show the distribution of the sample –proportional to the universe considered– according to the educational mode.

Table 1: distribution of subjects in the sample according to educational mode general scale

Educational mode	Subjects
Technical Schools	152
Secondary Schools	285
University Schools	235
Private Schools	266

Table 2: distribution of subjects in the sample according to educational mode advanced scale.

Educational mode	Subjects
Technical Schools	82
Secondary Schools	286
University Schools	190

Private Schools	166

Instruments

The General Scale, version 1993, of the Raven's Progressive Matrices Test was administered to the whole sample, whereas the Advanced Scale, Series I and II, was administered to 724 subjects of the previous sample.

Procedures

The scales selected were administered collectively (Group-class). For the first one, no time limit was given so that this would allow us to asses intellectual capacity without the intervention of speed in the task. After a 15-day interval the same subjects were given the Advanced Scale with both its series. Series I was administered with a time limit of 10 minutes and immediately followed the administration of Series II, with a time limit of 40 minutes, according to the procedures provided for by the author, in order to evaluate intellectual efficiency. The difference in the number of subjects included in the administration of both scales was due to students being absent on the day of the second administration.

From the results obtained, percentile norms were constructed corresponding to the group examined, for each of the Scales applied. Direct scores of scales were compared to those of 1964 and to that constructed for the sample examined.

Descriptors were estimated for all the groups examined and according to educational mode, in order to establish possible differences among them.

Statistical tests were also applied in order to estimate significant differences ($p < 0.05$ and $p < 0.01$) among:

1. Direct scores of the total samples of 1964 and 1997/8 scales.
2. Mean scores according to age groups (13-16 and 17-18 years of age).
3. Mean scores according to educational mode (General and Advanced Scale, Series II)

A correlation was ultimately made among the direct scores of the General Scale and the Advanced Scale, Series II: $r = 0.68$

Results

Percentile norms for the student population en La Plata, between 13-18 years of age. Percentile norms of the General and Advance Scales, Series II are shown on Tables 3 and 4.

Table 3: percentile norms for the general scale

PERCENTILE	13-14 YEARS	15-16 YEARS	17-18 YEARS
99	59	59	58
95	56	56	57
90	55	55	56
75	52	52	53
50	48	49	50
25	44	44	46
10	39	40	43
5	36	36	40

Table 4: percentile norms for the advanced scale series II

PERCENTILE	13-14 YEARS	15-16 YEARS	17-18 YEARS
99	29	30	32
95	26	28	30
90	25	26	29
75	22	24	25
50	18	19	21
25	14	15	17
10	10	10	13
5	8	8	9

Comparison of direct scores of subjects examined in the 1995-1998 study with those published for the city of La Plata in 1964

The results obtained evidence an increase in the direct scores necessary to reach each Percentile in the 1997-1998 study if compared with those of 1964. Such increase is an indicator of the average increase observed in the different age groups (Table 5).

Table 5: Increase in direct scores

Percentile	Point increase
P 5	14
P10	12
P25	10
P50	9
P75	8

P90	6
P95	6

The greatest difference is evidently seen in the direct scores of the lowest Percentiles, though the increase is manifest along the whole percentile scale.

Comparison of mean scores according to age groups (13-16 and 17-18 years of age)

The analysis of mean score differences in samples was carried out with a significance level of $p < 0.05$. The general average for the age group 13–16 is significantly lower than that for 17–18 years of age in both scales. These results confirm what Raven says about a poorer performance for the younger ones. According to him, probably "the capacity to perform comparisons and reason using analogy be, at such age, an intellectual development too recently acquired to systematically apply it".

The lowest the general average, the groups have shown to have a more heterogeneous performance.

Series I turns out to be very easy. It seems it is used merely as training for solving Series II, that is why results, from now on, shall only be presented for Series II.

Table 6: results obtained in each scale according to age group

Age	General scale		Advanced series I		Advanced series II	
	Mean	Bias	Mean	Bias	Mean	Bias
13 – 16 years	47.76	6.19	10.05	1.66	18.60	6.30
17 – 18 years	49.5	5.44	10.31	1.61	20.78	6.20

Comparison of mean scores in subjects participating in different educational modes, according to age group

For the 13-16-year-old group, the mean scores of schools pertaining to the University is significantly higher than the average of the total sample; whereas in the 17-18- year-old group, the most important difference is observed with the private schools.

The most homogeneous groups turn out to be those of the University in both age groups. The most heterogeneous ones are those present at the Secondary and Technical Schools (Table 7).

Table 7: results in the general scale for each educational mode

	Age

Educational mode	13-16 years		17-18 years	
	Mean	Bias	Mean	Bias
Technical Schools	46.54	7.06	47.91	5.23
Secondary Schools	46.25	6.31	49.65	5.96
University Schools	49.85	5.02	49.15	4.80
Private Schools	47.68	6.53	51.16	4.99

For the Advanced Scale, Series II, schools pertaining to the University show the highest average in the scores awarded and the Secondary and Technical schools, the lowest (Table 8).

Table 8: results obtained in the series ii for each educational mode

Educational mode	**Age**			
	13-16 years		17-18 years	
	Mean	Bias	Mean	Bias
Technical Schools	17.26	4.80	19.50	6.21
Secondary Schools	16.25	5.98	20.26	6.10
University Schools	21.52	5.92	22.58	5.14
Private Schools	18.5	5.39	21.07	6.22

Discussion and conclusions

From the results obtained in the administration of the General Scale and the comparison with those obtained in the typification carried out in 1964 in subjects with similar characteristics, a statistically significant increase is observed in the mean scores for all age groups under consideration, especially the scores required to get 5, 10, 25 and 50 percentiles.

As to the comparison of the measures showing the main trend according to educational mode, the sample corresponding to the *Bachillerato de Bellas Artes* seems to get the highest average, if compared with the results corresponding to subjects coming from the other educational modes. This would lead to state that the characteristics of the test seem to be related with the development of particular skills.

As for the results of the General and Advanced Scales, the average for the age group 13-16 turns out to be significantly lower than that of the age group 17-18 in both Scales.

One may wonder about the theoretical framework for such findings. First, it is necessary to highlight that tests undergo periodical revision, for they become obsolete (materials, norms and procedures for administration) and it is of upmost importance to improve them in order to overcome the limitations informed about its use within a long period of

application. This results in new versions with modifications. Several investigations have demonstrated that if the same subject is administered a previous and a current version of a test, he gets lower scores in the second one; that is, score criteria become progressively more demanding.

The observation made on the increase of intellectual performance leads us to those stated by Flynn on the constant increase of absolute values in every single important test, for all ages, and to the questions on the possibility of an increase in intelligence 0etween generations. The answer provided by the author accounting for the phenomenon described say there is no increase in intelligence, for intelligence tests do not examine general intelligence but rather some of its aspects, as in the case of Raven's test –related to the solution of abstract problems. Other possible alternatives are related to the increase in the schooling rate in populations corresponding to different generations. However, in our own results, we are dealing with subjects with the same schooling level. At the same time, a consideration is made that education or the school organisation are part of the cultural life and that this life is in constant change. This assumes a knowledge of the world of different characteristics: with demands, duties, as well as longings of various types. Progress may be said to be translated in new ways of thinking that come associated to a greater stimulation. Here we can highlight the visual stimulation now at work in our present day society. Sartori states video is transforming *homo sapiens*, product of a written culture, into a *homo videns*, for whom image has dethroned the word "everything ends up being visualised".

As regards the increase of mean scores, it is worth noting that in the particular case of Raven's Test, the material of the General Scale has undergone no change at all, except for the order of presentation. Subjects whose adolescence is spent in different periods, when faced with the difficulties the test presents have more resources to provide correct answers because of the differential qualities of their life experience.

If we take into account that each generation is increasingly surrounded with images, all the richer as we progress, which suppose not only watching them but analysing and interpreting them, we can very well ask ourselves the following: such visual analysis and interpretation corresponds to a particular form of intelligence effectively developing when in contact with complex visual media? That being the case, such particular form of intelligence can be postulated as the responsible one for the increase in scores? Or can it be a change in the type of logical reasoning used that would imply a passage from the norms of thought involving deductive concatenation to logical rules that we could call "binary", formed in the contact with the multimedia and the cyberspace and which are

expressed in dychotomic propositions –such as yes-no, it is-it is not, it corresponds-it does not correspond, it is adequate-inadequate?

When results obtained by the students from different schools were compared, differences favouring the group from the *Bachillerato en Bellas Artes* seem to advocate for an answer in the affirmative for the first questions expressed above.

This work does not offer conclusions in that respect, but rather questions aiming at opening up a space for reflection and future study.

The authors' intention is to continue updating the scales of Raven's for both scales, widening the spectrum of ages examined, and to use as contrast figures –of present and future results– the performance of the subjects in verbal tests in order to establish if the same phenomenon can be detected, in search for far reaching explanations to this problem.

Notes

1. *Editorial Paidós* published in its evaluation folders, norms obtained from 100 students attending the National School "Carlos Pellegrini" belonging to the University of Buenos Aires and from 200 students of Economic Sciences in the Federal District and Mar del Plata.

References

1. Anastasi, A. (1980). T*ests Psicológicos*. Madrid: Aguilar.

2. Casullo, M.M. (1996). *Evaluación Psicológica y Psicodiagnóstico.* Buenos Aires. Facultad de Psicología UBA. Catálogos Editora.

3. Crombach, L. (1998). *Fundamentos de los Tests Psicológicos*. Madrid: Biblioteca Nueva.

4. Daniel, W. (1995). *Estadística Aplicada a las Ciencias Sociales y a la Educación*. Colombia: Mc Graw Hill.

5. Flynn, J. (1984*). The mean IQ of Americans: massive gains* 1932-1978. *Psychological Bulletin.*

6. Flynn, J. (1987). *Massive IQ gains in 14 nations: what IQ tests really measure. Psychological Bulletin.*

7. Flynn, J., Rossi Casé, L. (2001). *Standardization of Raven´s in Argentina 1964 -2000: IQ gains huge; gender differences nil.* Rn preparación.

8. Fogliatto, H. (1991). *Cuestionario de Intereses Profesionales C.I.P*. Buenos Aires: Guadalupe.

9. Guilford, J. P. & Merrifield, P. R (1960). *The structure-of-intellect model:its uses and implications. Report from the Psychological Laboratory, 24.* USC.

10. Ministerio de Cultura y Educación de la Nación. Secretaría de Programación y Evaluación Educativa. Subsecretaría de Evaluación de la Calidad Educativa. Dirección Nacional de Evaluación. Argentina (1997). *Primer operativo de evaluación de Finalización del Nivel Secundario.*

11. Neisser, U. (1998). Sommes-nous plus intelligents que nos grands-parents? *La Recherche, Nº 309, Mayo 1998, pp. 46 – 52.*
12. Padua, J. (1993). Técnicas de Investigación Aplicada a las Ciencias Sociales. México: F.C.E.
13. Raven, J.; Raven, J.C. & Court, J. H. (1993). *Test de Matrices Progresivas. Manual*. Buenos Aires: Paidós.
14. Raven, J.C. (1975) *Test de Matrices Progresivas. Escala General. Manual*. Buenos Aires: Paidós.
15. Sartori, G. (1998). *Homo videns. La sociedad teledirigida*. Buenos Aires: Taurus.
16. Spearman, CH. (1955). *Las habilidades del hombre*. Buenos Aires: Paidós.

Received: July 2001
Revision received: October 2001
Accepted: December 2001