Chapter 13

Kuwaiti norms for the Classic SPM in an International Context*

Ahmed Abdel-Khalek** and John Raven

Abstract

A probability sample (*n*=8,410) of Kuwaiti school students aged 8-15 responded to Raven's *Standard Progressive Matrices*. The test was administered, untimed, in group sessions. In this paper, the smoothed summary age norms for Kuwait (which will themselves be of interest to many psychologists and others working in Kuwait and neighbouring countries) are first compared with what have become the standard international reference data for such work, namely the 1979 British norms. Thereafter, the Kuwait norms are compared with those reported for several other countries and set in the context of accumulating data on changes over time. The results show that, while some, as yet unidentified, features of the environment do have a dramatic effect on scores, aspects of the environment that many people would have expected to have a significant effect (such as differences in calligraphy) are much less important than might have been thought.

Background

Raven's Standard Progressive Matrices (SPM) test was constructed to measure the eductive component of ***g*** as defined in Spearman's theory of cognitive ability (Raven, Raven, & Court, 1998, updated 2003, p. G1).

* An earlier version of this chapter was published in *Social Behaviour and Personality: An International Journal* (2006), *4,* 169-179.

** The research reported in this chapter was supported by Kuwait University under Grant No. OP01/02. The authors gratefully acknowledge the able assistance of Research Administration at that university.

Kaplan and Saccuzzo (1997) stated that "research supports the Raven Progressive Matrices (RPM) as a measure of general intelligence, or Spearman's ***g*** factor… In fact, the Raven may be the best single measure of ***g*** available" (p. 359).

In the same vein, Jensen (1998) maintained that "in numerous factor analyses, the Raven tests, when compared with many others, have the highest ***g*** loading and the lowest loadings on any of the group factors. The total variance of Raven scores in fact comprises virtually nothing besides ***g*** and random measurement error" (p. 541). He added that Raven's Progressive Matrices is often used as a "marker test of Spearman's ***g***. That is, if it is entered into a factor analysis with other tests of unknown factor composition, and if the Matrices has a high loading on the general factor of the matrix of unknown tests, its ***g*** loading serves as a standard by which the ***g*** loadings of the other tests in the battery can be evaluated" (p. 38).

By the same token, Lynn, Allik, Pullman, and Laidra (2004) stated that "the Progressive Matrices is widely regarded as the best test of abstract or non-verbal reasoning ability, and this is itself widely regarded as the essence of "fluid intelligence" and of Spearman's ***g***" (p. 1250). Mackintosh (1996, p. 564) has described it as "the paradigm test of non-verbal, abstract reasoning ability".

This view is not, of course, universally accepted. Indeed, Raven, Raven and Court (1998, 2000) refer to several studies which suggest a loading on spatial ability, and a review of the extensive literature dealing with this topic from the point of view of researchers keen to distinguish "Working Memory" from ***g*** has been provided by Ackerman, Beier, and Boyle (2002).

The Standard Progressive Matrices test enjoys good psychometric characteristics (see: Court & Raven, 1995; Kline, 2000; Murphy & Davidshoffer, 1998). A huge body of published research bears on the validity of this test (Gregory, 1992). Therefore, it has gained widespread acceptance and use in many countries all over the five continents (Irvine & Berry, 1988). No other test has been so extensively used in cross-cultural studies of intelligence. Lynn and Vanhanen (2002) summarized a plentiful number of studies based on normative data for the test has been collected in 61 countries. For all these reasons, Kaplan and Saccuzzo (1997) concluded that "with its new worldwide norms and updated test manual, the Raven holds promise as one of the major player in the testing field in the 21st century" (p. 361).

The Arab countries are in a great need of standardized intelligence tests with local norms. Indeed, the three series of the Progressive Matrices test, i.e., the Standard, Coloured, and Advanced, are already available in the majority of the Arab countries. The Standard Progressive Matrices test has been administered to different samples in most Arab countries. However, the vast majority of these studies remain either unpublished or published in Arabic.

In 1988, Abdel-Khalek found that the test-retest reliability reached .82 among Egyptian college students. A clear general factor with high loadings was extracted from the five Sets of the test. A factor analysis of the total score on the test and four subscales of Thursone's Primary Mental Abilities yielded a general and high loaded factor, on which the Matrices loading was .77, denoting high concurrent validity.

More recently, Abdel-Khalek and Lynn (2006) examined the sex differences on the test and found a small sex difference of .08 *sd* (1.2 IQ points) favouring girls.

Using a Kuwaiti sample of school children (n=968), Abdel-Khalek (2005) found the test-retest reliability ranged from .69 to .85, while internal consistency assessed by the alpha coefficient ranged between .88 to .93 denoting good temporal stability and internal consistency. The loadings of the five sets on the only salient factor ranged from .72 to .89 indicating the good factorial validity of the scale.

The objective of the current investigation was primarily to create Kuwaiti norms for the *Standard Progressive Matrices* (SPM) test, but these norms are presented here in an international context since the comparative data that have emerged are of considerable importance to cognitive psychology.

Method

Participants

A sample of 8,410 8-15 year olds was recruited. All of them were Kuwaiti citizens and students in the governmental schools in the six districts in Kuwait. In each district one elementary, intermediate and secondary school for both boys and girls were randomly chosen. The selection of school districts used a stratified random sampling procedure. The test was administered to at least 60 students in each age group of boys and the same for girls in each of the six districts of Kuwait.

The Test

The original, 1958, version of the SPM (Raven, J. C., 1958) was employed but adapted in the sense that, in the Arabic test booklets, the main matrix and the six or eight alternatives were transposed to read from right to left following the custom of Arabic writing.

Procedure

The SPM was administered to students by a group of competent and trained testers. The testers in the boys' schools were male, and female in girls' schools. In every class, testing was carried out by a tester and an assistant. Testing was carried out in whole classes of 25-30 students. Verbal instructions were given to the students on how to do the test. The test was given without time limits. The testing was carried out in the year 2002. The raw data of the completed answer sheets were scored by computer.

Results and Discussion

Table 13.1 presents the Kuwaiti norms in the context of the 1979 norms for Great Britain. It will be seen that, although the scores obtained in Kuwait are, in general, considerably lower than the much earlier UK norms, especially among the younger and less able pupils, the general impression is one of rather surprising similarity. There is not space here to speculate on the reasons for the decline in the discrepancy among the older age groups, but it would be inappropriate not to draw attention to the ceiling effect which now restricts the variation in scores among the more able from age 12 onwards (and which exacerbates the non-Gaussian within-age distributions which, among other things, make it inappropriate to process the data in terms of means and Standard Deviations, let alone IQ scores).

In the earlier version of this paper which was acknowledged on the title page of this chapter, data from an earlier, smaller, sample of Kuwaiti pupils were included in a table comparing a wide range of international norms. Since a modified version of this Table now appears in the *General Introduction* to this book that, rather long, table has been omitted here.

Nevertheless, it would be inappropriate to pass on without comment - especially as the preparation of that table was prompted by the arrival

Table 13.1. *Standard Progressive Matrices*
Smoothed 2006 Norms for Kuwait in the Context of 1979 British Standardisation

	Age in Years (Months)										
	7 6(9) to 7(2)		7½ 7(3) to 7(8)		8 7(9) to 8(2)		8½ 8(3) to 8(8)		9 8(9) to 9(2)		9½ 9(3) to 9(8)
Percentile	UK	**KU**	UK	**KU**	UK	**KU**	UK	**KU**	UK	**KU**	UK
95	34	**29**	37	**33**	40	**38**	42	**40**	44	**42**	46
90	32	**25**	35	**29**	38	**34**	40	**37**	42	**39**	44
75	26	**19**	30	**22**	33	**26**	36	**31**	38	**33**	41
50	**19**	**14**	**22**	**16**	**25**	**18**	**31**	**21**	**33**	**24**	**36**
25	14	**12**	15	**12**	17	**13**	22	**14**	25	**15**	28
10	12	**10**	12	**10**	14	**11**	17	**11**	17	**12**	19
5	10	**9**	11	**9**	12	**10**	13	**10**	14	**10**	14
n	138	**348**	148	**495**	174	**273**	153	**413**	166	**387**	198

(continued)

	Age in Years (Months)										
	10½ 10(3) to 10(8)		11 10(9) to 11(2)		11½ 11(3) to 11(8)		12 11(9) to 12(2)		12½ 12(3) to 12(8)		13 12(9) to 13(2)
Percentile	UK	**KU**	UK	**KU**	UK	**KU**	UK	**KU**	UK	**KU**	UK
95	49	**46**	50	**47**	51	**48**	52	**48**	53	**50**	54
90	47	**43**	48	**44**	49	**46**	50	**46**	51	**48**	52
75	43	**39**	44	**41**	45	**43**	46	**44**	47	**44**	49
50	**39**	**32**	**40**	**35**	**41**	**37**	**41**	**38**	**42**	**39**	**43**
25	33	**23**	34	**27**	36	**30**	37	**32**	38	**33**	39
10	27	**15**	29	**18**	31	**20**	31	**22**	32	**23**	33
5	22	**13**	24	**14**	25	**15**	26	**16**	27	**17**	28
n	194	**405**	187	**436**	164	**418**	164	**401**	174	**395**	185

(continued)

	Age in Years (Months)										
	14 13(9) to 14(2)		14½ 14(3) to 14(8)		15 14(9) to 15(2)		15½ 15(3) to 15(8)		16 15(9) to 16(2)	16½ 16(3) to 16(8)	17 16(9) to 17(2)
Percentile	UK	**KU**	UK	**KU**	UK	**KU**	UK	**KU**	**KU**	**KU**	**KU**
95	55	**52**	56	**54**	57	**54**	57	**54**	**54**	**54**	**54**
90	54	**50**	54	**51**	55	**52**	55	**52**	**52**	**52**	**52**
75	50	**47**	50	**49**	51	**49**	51	**49**	**49**	**49**	**49**
50	**45**	**43**	**46**	**44**	**47**	**45**	**47**	**45**	**45**	**45**	**45**
25	42	**37**	42	**39**	42	**40**	42	**40**	**40**	**41**	**41**
10	36	**31**	36	**32**	36	**33**	36	**33**	**33**	**34**	**34**
5	30	**26**	33	**29**	33	**30**	33	**31**	**31**	**31**	**31**
n	196	**404**	189	**445**	191	**418**	171	**391**	**466**	**425**	**234**

of the Kuwait data. Two things strike one immediately. The first is the similarity between the normative data collected by different people using different sampling procedures in these different countries. The second is the wide within-age variance in the norms from within each country.

The similarity in the norms across countries having such different calligraphies, such different reading and writing systems, such different values, such different educational systems, such different child rearing practices, such different family sizes, such differential access to television, and at such different stages in economic development strongly suggests that cultural variation in these socio-demographic characteristics has much less impact that is commonly assumed.

Furthermore, the variance *within* countries reconfirms this observation. If these cultural variables *did* have the impact on scores that is often asserted they would surely influence the within-culture variance. Everyone in each of these cultures is exposed to much the same cultural environment, yet it seems that it neither restricts nor enhances the within-cultural variance.

From the data for 11 year olds, it would seem that the norms for the 50th and lower percentiles in India, Kuwait, and Qatar lag increasingly behind.

Missing from the table are some data that many people find embarrassing and which lack political correctness. The data in question have to do with Blacks in the USA and South Africa, many Native American groups (with the exception of the Eskimos), and other groups lacking a tradition of literacy.

It would in fact have been misleading to have included these data in the Table because most of the samples leave much to be desired. Nevertheless such data as exist (see Raven, 2000, Court & Raven, 1995, for a summary) taken together with the data for South Africa and Indian tribal areas included in this volume reveal huge differences between these groups and the Kuwait data reported in Table 13.1 as well as that summarised in Table 1.2 of the General Introduction to, and Overview of, this book.

Changes Over Time

It is important to contextualise these observations by again emphasising that data presented in other chapters in this volume reveal dramatic changes in scores over time. This means that some aspects of the environment *do* have an enormous (and previously unsuspected) effect

on scores - but again without significantly reducing the within-age variance. On the other hand, the cross cultural data reported in this book - including that presented in this Chapter - clearly show that the features of the environment that most people first think of as explanations for the change over time have much less effect than was previously believed.

In an effort again to avoid creating a misleading impression it is important to draw attention to the fact that similar increases in scores over time have been documented on a wide range of *verbal* measures of *eductive* (i.e. "meaning making" or "reasoning") ability (see, e.g. Bouvier, 1969; Schaie & Willis, 1986; Flynn, 2000). In other words, they have not been limited to pictorial or diagrammatic tests of the kind we have been concerned with here. This undermines yet another of the "explanations" of the change over time that most commonly spring to mind.

It follows from these observations that the increase is *not* due to such things as schools attempting to enhance levels of "creativity" by encouraging children to tackle non-verbal puzzles or handle computer games or any of the explanations most widely favoured by psychologists and listed by Thorndike (1975, 1977) as possible explanations of the increase he had documented in the norms for the Stanford-Binet test.

From the point of view of seeking an explanation of the increase over time it is, however, perhaps still more important to note that an increase of exactly the same magnitude has occurred in height and life expectancy. It is worth dwelling on some of the implications of this.

First, no one would conclude from the fact that life expectancy is measured by a Rasch Scale analogous to that used to measure eductive ability via the RPM that the variance must be determined by some single underlying ability analogous to speed of neural processing. Nor would they seek a single factor explanation of the increase over time. Nor would they conclude from the fact that backward projection of the increase to the time of the Greeks that the Greeks must have had impossibly short life expectancies and therefore that the measures must be devoid of meaning. Nor would they conclude from the fact that there are ethnic and socio-economic differences in life expectancy that both the measures themselves and the differences between groups have no meaning. And nor would they expect that the same factors as are responsible for the within cohort variance are the same as those responsible for the increase over time. Yet all of these claims have been made by Flynn or others in connection with the increase in Raven Progressive Matrices (RPM) scores.

Finally, it is worth nothing that the causes of the increase in height and life expectancy and the ethnic and cultural differences associated with them have proved just as difficult to pin down as those on the RPM.

Despite these caveats, it is essential, when seeking to interpret the similarities and differences between cultures at any point in time that were summarised in earlier chapters of this volume, to bear in mind that the cross-birth-cohort data show that changes in the environment not only *can*, but *have*, had effects which completely swamp the differences between cultural groups.

Conclusion

Normative data for the Standard Progressive Matrices derived from testing a large representative sample of young people in Kuwait are expected to be of considerable interest to psychologists, teachers, and others working in Kuwait and neighbouring countries. However, when viewed in the context of parallel data from several other countries and cultures, the data acquire a much wider significance in that they reveal remarkable similarity in the norms across cultures at any point in time accompanied by dramatic change over time.

The data clearly show that variation in features of the environment that many people would have expected to markedly influence scores – such as variance in calligraphy, educational systems, and cultural norms – have much less effect than many people would have expected whilst as yet unidentified features of the environment have a much greater effect than many people would have suspected.

In this a context, such cross-cultural differences as remain appear to merit less attention than might otherwise have seemed to have been the case.

And, when the cross-birth-cohort data are themselves compared with similar data relating to life expectancy, the logic of many arguments put forward by psychologists would seem to be, at best, highly questionable.

References

Abdel-Khalek, A. M. (1988). Egyptian results on the Standard Progressive Matrices. *Personality and Individual Differences, 9,* 193-195.

Abdel-Khalek, A. M. (2005). Reliability and factorial validity of the Standard Progressive Matrices among Kuwaiti children ages 8 to 15 years. *Perceptual and Motor Skills, 101,* 409-412.

Abdel-Khalek, A. M., & Lynn, R. (2006). Sex differences on the Standard Progressive Matrices and in educational attainment in Kuwait. *Personality and Individual Differences, 40*, 175-182.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131,* 567-589.

Bouvier, U. (1969). *Evolution des Cotes a Quelques Test.* Belgium: Centre de Recherches, Forces Armees Belges.

Cayssails, A. (2001). *Carpeta de Evaluacion Escale General*. Buenos Aires, Argentina: Paidos.

Court, J. H., & Raven, J. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 7: Research and References: Summaries of Normative, Reliability, and Validity Studies and References to All Sections*. San Antonio, TX: Harcourt Assessment

Deltour, J. J. (1993). *Echelle de Vocabulaire Mill Hill de J. C. Raven: Adaptation Francaise et normes comparees du Mill Hill et du Standard Progressive Matrices (PM38). Manuel et Annexes*. Braine le Chateau, Belgium: Editions L'Application des Techniques Modernes SPRL.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29-51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

Flynn, J. R. (2000). IQ gains, WISC subtests and fluid ***g***: ***g*** theory and the relevance of Spearman's hypothesis to race. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The Nature of Intelligence (Novartis Foundation Symposium 233)* pp. 202-227. Chichester, England: Wiley.

Gregory, R. J. (1992). *Psychological testing: History, principles, and applications*. Boston: Ally & Bacon.

Irvine, S. H., & Berry, J. W. (Eds.) (1988) *Human abilities in cultural context*. Cambridge: Cambridge University Press.

Jaworowska, A., & Szustrowa, T. (1991). *Podrêcznik do Testu Matryc Ravena: Wersja dla Zaawansowanych (1992). Polskie standaryzacje: Uczniowie 12;6-19;5 i studenci*. Warsaw: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

Jensen, A. R. (1998). *The **g** factor: The science of mental ability*. Westport: Praeger.

Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological testing: Principles, applications, and issues* (4th ed.). Pacific Grove: Brooks/Cole.

Kline, P. (2000) *Handbook of psychological testing* (2nd ed.). London: Routledge.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.

Lynn, R., Allik, J., Pullman, H., & Laidra, K. (2004). Sex differences on the progressive matrices among adolescents: Some data from Estonia. *Personality and Individual Differences, 36*, 1249-1255.

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science, 28*, 559-572.

Miao, E. S. Y. (1993). *Translation of J. Raven, J. C. Raven, & J. H. Court, Manual for Raven's Progressive Matrices Tests*. Taiwanese Edition. Taiwan: Chinese Behavioural Science Corporation.

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing*: *principles and applications* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Raven, J. (1981). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.1*: *The 1979 British Standardisation of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, Together With Comparative Data From Earlier Studies in the UK, US, Canada, Germany and Ireland*. San Antonio, TX: Harcourt Assessment.

Raven, J. (2000a). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.3 (Second Edition): A Compendium of International and North American Normative and Validity Studies Together with a Review of the Use of the RPM in Neuropsychological Assessment by Court, Drebing, & Hughes*. San Antonio, TX: Harcourt Assessment.

Raven, J. (2000b). The Raven's Progressive Matrices: Change and stability over culture and time*. Cognitive Psychology, 41*, 1-48.

Raven, J., Raven, J. C., & Court, J. H. (1998, updated 2003) *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1*: *General Overview*. San Antonio, TX: Harcourt Assessment.

Raven, J., Raven, J. C., & Court, J. H. (2000, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3*: *The Standard Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Raven, J. C. (1958). *The Standard Progressive Matrices.* London: H. K. Lewis. An earlier version of this test was known as *Progressive Matrices (1938)* and was also published by H. K. Lewis. The test was subsequently published by OPP Ltd. (Oxford) and now by Harcourt Assessment, San Antonio, TX.

Schaie, K. W., & Willis, S. L. (1986). *Adult Development and Ageing* (2nd edition). Boston: Little Brown.

Thorndike, R. L. (1975). *Mr. Binet's Test 70 Years Later*. Presidential Address to the American Educational Research Association.

Thorndike, R. L. (1977). Causation of Binet IQ decrements. *Journal of Educational Measurement, 14*, 197-202.