Raven, J. (2005, October 1). Raven's Progressive Matrices: The "Flynn Effect" Continues. *WebPsychEmpiricist*.
Retrieved October 1, 2005, from http://wpe.info/papers_table.html

WPE WebPsychEmpiricist

Raven's Progressive Matrices: The "Flynn Effect" Continues

10/1/05

John Raven
Edinburgh, Scotland

Abstract

(i) Data collected with the Coloured Progressive Matrices test in Poland and Korea strongly suggest that the "Flynn Effect" (viz. the successive increase in scores with each new birth cohort) is continuing. (ii) So far as can be judged from the available data, the increase in scores affects people of all levels of ability (and is not confined to the "less able"), but severe methodological problems face those who wish to compare the relative size of the gains made by people of different levels of ability. (iii) This across-the-board effect confirms that the increase is not due to the remediation of obvious environmental defects. (iv) Further cross-cultural data added to the data pool in recent years confirm that most of the hypotheses commonly put forward to explain the increase are invalid. (v) Increases in life expectancy over the same period of time as data for the Progressive Matrices exist invalidate most of the inferences (e.g. that "the tests cannot save themselves") which Flynn initially drew from his data. (vi) These same data also call into question the common inference that the single-dimensional properties of the Raven Progressive Matrices (demonstrated using Item Response Theory) suggest that a single psychological process (e.g. "speed of neural transmission") underlies the variance in scores.

Correspondence concerning this article should be sent to: John Raven, 30 Great King St., Edinburgh EH# 6QH Scotland, ++44 (0) 131 556 2912, jraven@ednet.co.uk, or WPEList@yahoogroups.com

Raven's Progressive Matrices: The "Flynn Effect" Continues
10/1/05

In the late 1970s a number of authors (e.g. Bouvier, 1969, Thorndike, 1975, 1977, Garfinkel, & Thorndike, 1976, Schaie, 1983, and Raven, 1981) showed that scores on tests of eductive (or meaning-making) ability (often misnamed "fluid intelligence"), but not reproductive ability (often misnamed "crystallized intelligence"), had been increasing dramatically over time. In the mid 1980s, Flynn (1984a&b, 1987) supported these observations by bringing together huge quantities of data, mostly for the Raven Progressive Matrices, for 18-year-olds. The bulk of these data had been collected from military conscripts by the military services of many countries and stretched back over several decades. The increase thereafter became known as the "Flynn Effect". Prompted by the publication of Flynn's data, the author arranged for a standardization of the Standard Progressive Matrices (SPM) among British adults (see Raven, J., Raven, J. C., & Court, J. H., 2000/04 and Raven, J. 2000a for data first published in the 1993 edition of Raven, Raven, & Court). The SPM had last been standardized in the UK by Foulds and J. C. Raven in the early 1940s (see Raven, 2000b for a description of the methodology used in both studies). Comparison of the two data sets revealed a dramatic increase over time. An important incidental result of the study was, however, to strongly confirm the accuracy of Flynn's speculation that data from cross-sectional studies (i.e. studies in which a single sample of a cross section of the adult population had been tested at a particular point in time, as in both the studies already mentioned) which had previously been interpreted as showing that eductive ability declines with age might, in reality, mainly be due to the previously unsuspected inter-generational increase.

A digression: The Raven Progressive Matrices tests

For the benefit of those familiar with the Raven Progressive Matrices tests, it may be helpful here to say something about the tests themselves. They consist of a series of wallpaper-like patterns demonstrating serial change (of increasing complexity) in two dimensions. Each pattern or design has one bit missing. Those taking the test are asked to find the missing piece from a number of options. The tests are widely regarded as among the best indices of eductive ability (viz. the ability to find meaning within complexity), with eductive ability itself often being regarded as one of the two main components of Spearman's ***g***. An illustrative item together with a brief review of the theoretical basis and validity of the test and a summary of variation in scores with time and culture will be found in Raven (2000b). More detail will be found in Raven, Raven and Court (1998, revised and updated 2003, and 2000, updated 2004).

Possible explanations of the intergenerational increase … and their validity

Thorndike proffered a number of possible explanations of the increase - such as increasing access to television, changes in educational practice, smaller family sizes, and changes in disciplinary practices. However, as the various studies brought together by Neisser (1998) show, none of these hypotheses hold up. This conclusion has been strongly reinforced in our own data (see Raven, 2000b) which show remarkable consistency in the norms obtained in most cultures with a tradition of literacy at any point in time but dramatic increases in the scores obtained in all cultures over time. A selection of these data have been combined with the results of more recent studies in Table 1 below. This similarity in the norms obtained in different cultures is all the more striking in view of the fact that there are major differences between them in such things as age of entry to, and duration of, education, the nature of the education provided, calligraphy, the child-rearing practices advocated and adopted, family sizes, access to television and cultural values. The only plausible "explanation" of the increase that remains after most of the others have been discredited (and promoted, in particular, by Greenfield in Neisser's 1998 book) - i.e. increasing familiarity with diagrams and "puzzles" - is undermined by the fact that the increase has occurred on all measures of eductive ability whether these are verbally based or not (Thorndike, 1977; Schaie, 1994; Bouvier, 1969; Flynn, 2000). Indeed Flynn (2000) has shown that the rate of increase on the WISC sub-tests is directly proportional to those sub-tests' loadings on eductive ability.

Although there is not space here to present the data, it is worth mentioning that other analyses of these cross-cultural data have also strongly confirmed that the concept of eductive ability is, in scientific terms, every bit as secure as "hardness", "life expectancy", and "height" (see Raven, Prieler, & Benesch, 2005). This is demonstrated by the fact that the Item-Response-Theory (IRT) -based Item Characteristic Curves (ICCs) for most items of the RPM increase incrementally and in step with those for other items of similar difficulty, with the whole forming a coherent sequence. The closest analogy is with a measure of high-jumping ability where the frequency with which people clear the bar set at any one height relates directly to the frequency with which they can clear it at other heights. The display of the tests' ICCs is remarkably stable over time and culture. A hint of the strength of the claim may be given by saying that the correlations between the IRT-based indices of item difficulty established in different social and cultural groups typically lie between 0.96 and 1.00. Similarly, the predictive validity of the test to external criteria is not only similar across groups but also continuous between them, i.e. the regression lines of the RPM against other measures of performance are continuous across the interface between groups having different mean scores (Raven, 2000a; Raven, Raven, & Court, 1998/2003, 2000/2004; Schmidt & Hunter, 1998; Gottfredson, 1997).

Returning now to our discussion of possible explanations of the observed increases in scores over time, it is important to note that most of the hypotheses that have been advanced have relied on single-factor explanations (such as "improved diet" or "changes in educational practice") mediated by single underlying processes - such as "speed of neural transmission".
Quite apart from the data presented in Neisser's book and in the author's publications, the chances of any such explanation, or Flynn's more recent basket ball analogy (see Dickens & Flynn, 2001; Flynn & Dickens, 2001), being correct are seriously challenged by two parallel findings: First, there has been a constant, year by year, increase in life expectancy over the same period as Raven Progressive Matrices scores exist (in the UK, life expectancy has virtually doubled over the last century). And height has been progressively increasing in a similar way. (For both see Bodzsar & Susanne,1998 - especially the chapter by Stefancic & Tomazo-Ravnik.)

Although increases in height and life expectancy have proved as inexplicable as the RPM increase, there are important insights to be gained from reflecting further on these data. First, despite the fact that life-expectancy is, like eductive ability measured by the RPM, assessed using a "Rasch" scale, this in no way suggests that a single underlying variable (analogous to "speed of neural transmission") is responsible for the variance. Nor is there any reason to believe that the factors that are responsible for the increase over time are the same as those which produce the within-birth-cohort variance. These data can also be used to illustrate the illogicality of many of the arguments Flynn initially advanced to discredit the concept of "intelligence" itself and current tests as meaningful measures of it in particular ("the tests cannot save themselves"). He argued, for example, that back-projection of the curves he plotted for the increase in RPM scores over the last century to the time of the Greeks led to the conclusion that the Greeks must have been incredibly stupid. Since that could not be true he concluded that modern tests of "intelligence" must be meaningless. Yet no one would infer from the fact that back-projection of the life expectancy graphs would suggest that the Greeks' life expectancy must have been very short indeed and that the measures must therefore be meaningless. Nor would they infer that our measures of life expectancy are meaningless because there are differences between men and women, socio-economic groups, and ethnic groups. Yet Flynn used all of these things to cast doubt on the meaningfulness of measures of eductive ability – although is main point was that the environment had been having a previously unsuspected influence.

It is also important to note that the increases in height and life expectancy over time have not reduced the variance within each birth cohort. Tall people have got still taller. The same is also true of the variance in scores on measures of eductive ability. The observed increases are *not* due to remediation of environmentally-induced birth defects. It has long been obvious that the intergenerational increases in Standard Progressive Matrices scores has resulted in a serious ceiling effect among more able individuals born more recently. (Currently, the SPM fails to discriminate above the 75^{th} percentile among older adolescents and young adults.) Unfortunately, this fact has often been overlooked by those who have asserted that the main increase in scores has been among the less able and inferred that it mainly arises from remediation of obvious environmental deficiencies. In reality, the fact that the true SPM scores of the more able have continued to increase in line with their previous increases with birth cohort is easily documented by looking at the increases in the *Advanced* Progressive Matrices scores of this group (see Raven, 2000b).

Assessing differential gain: A methodological digression

Despite the fact that we have shown that the scores of people at all levels of ability have increased over time, it is nevertheless still often asserted that the less able have gained "more" than those of the more able. We have already illustrated some of the problems which make it difficult to substantiate such hypotheses, but there are actually much more basic problems to be overcome. Although discussion of these is something of a digression here, they are nevertheless quite striking and may be of considerable interest to readers of this article because they invalidate virtually all conclusions drawn from studies in which attempts have been made to generate and interpret change scores calculated by, for example, subtracting pre-treatment from post-treatment scores on a particular test in an attempt to measure individual differences in such things as "responsiveness to instruction (learning potential)" or "reactivity to a (particular) drug or treatment".

The problems are best illustrated using some Figures initially prepared by Prieler and discussed in more detail in Raven et al. (2000/04), Prieler and Raven (10/20/02), and Prieler and Raven (2002).

Figure 1 illustrates the problem of determining gain among high ability respondents in a way which would make it possible to meaningfully compare that gain with the gain of those with average or low scores. Figure 2 provides a parallel illustration for people of low and average ability.

If use is made of a test which has the Test Characteristic Curve (TCC) shown on the left in Figure 1 (Test 1), the mean score of the high ability group increases from A at time 1 (i.e. in year one or before training or the administration of a drug) to B at time 2 (perhaps a year later or after training). This is a relatively small increase. But if we use the more difficult test (Test 2), whose TCC is shown on the right, the same increase in the latent ability of the high ability group shows up as a huge increase in raw score, moving from X to Y.

But if one looks at Figure 2 one sees exactly the opposite effect at the other end of the scale. The apparent increase in score from Time 1 to Time 2 is huge on Test 1 and trivial on Test 2.

Putting the two cases together, it is obvious that, if a researcher employs Test 1 to try to assess differential change over time or the impact of an intervention, the relative gains of the low ability group are huge while those of the high ability group are trivial. On the other hand, if the researcher employs Test 2, exactly the opposite findings emerge.

The general, and vitally important, conclusion illustrated using these examples is that the apparent magnitude of any real increase in ability over time or as a result of some intervention (such as a developmental experience, a stress-inducing event, or administration of a drug), depends on (a) the difficulty level of the test used, (b) the shape of its Test Characteristic Curve, and (c) the sector of the curve on which the change occurs.

This makes it virtually impossible, without employing techniques like those developed by Fischer and Prieler (and described in an Appendix to Raven et al., 2000/04), to (1) make any meaningful statement about the *relative* magnitude of gains or losses of high, medium, and low ability groups, (2) generate meaningful individual change scores (i.e. to measure individual differences in learning potential or responsiveness to stress), or (3) to decide whether a change in an individual's score represents an improvement or deterioration worthy of note.

The intergenerational increase continues

Reverting once more to our discussion of the changes over time, we must now note that, perhaps because the increases were so unexpected and contradicted faith in the invariance and genetic determination of "intelligence", the belief that the observed and documented increase in eductive ability had run its course has been widely espoused.

Unfortunately, the figures in Table 2, which presents normative data for the *Coloured* Progressive Matrices, do not exactly lead one to have great faith in this. Although the data come from a number of different countries, they nevertheless convey an unmistakable impression that the increase in scores across birth cohorts documented in so many robust earlier studies is continuing. The Table shows that the 50th percentile for six year olds has increased from 17 in the UK and many other countries around 1982, to 19 in Poland in 1991, and to 24 in Korea and Poland in 2001-03. Note that examination of only the higher percentiles might well mislead researchers since the CPM has only 36 items and thus runs into a ceiling effect among these age groups. (Incidentally, one of the reasons why my colleagues and I hardly ever present the results of studies with the RPM in terms of means and standard deviations [and still less frequently in terms of deviation IQs], more “sophisticated” though they may seem, is that they tend to preclude detailed examination of data.)

Summary

The main evidence supporting the claim that scores on tests of eductive ability have been increasing over time was first summarized. New data demonstrating similarity in the norms obtained in a variety of cultures with a tradition of literacy at any point in time were then added. These disconfirm most of the hypotheses commonly advanced to explain the increase over time. It was then shown that there are severe methodological problems (which are not widely appreciated) confronting those who wish to compare the relative size of the gain among the more and less able. Many of these have been overlooked by those who have claimed that the gains have been much greater among the less able and concluded that the gains must therefore mainly been due to remediation of environmental defects. The data relating to the increase in eductive ability scores were then juxtaposed with data on life expectancy. It was suggested that, by analogy, these increases in life expectancy over much the same period of time invalidated most of the arguments Flynn originally put forward in support of his conclusion that “the tests cannot save themselves”. These same data also called into question the common inference that the Item-Response-Theory-based properties of Raven’s Progressive Matrices tests suggest that a single psychological process (such as “speed of neural transmission”) underlies the observed variance in eductive ability (which some psychologists see as lying at the heart of **g**, and even “intelligence” itself). In this way, these data call the quest for an explanation of the “Flynn Effect” based on effects on an assumed single underlying psychological process into serious question. Finally, a comparison between earlier data and data recently collected with the Coloured Progressive Matrices in Poland and Korea strongly suggest that eductive ability scores are continuing to increase with each new birth cohort.

References

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.

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.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. Psychological Review, 108, 346-369.

Flynn, J. R. (1984a). IQ gains and the Binet decrements. Journal of Educational Measurement, 21, 283-290.

Flynn, J. R. (1984b). 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, and K. Webb (Eds.), The nature of intelligence (Novartis Foundation Symposium 233) pp. 202-227. Chichester, England: Wiley.

Flynn, J. R, & Dickens, W. T. (2001). Great leap forward: a new theory of intelligence. New Scientist, 21 April, 44-47.

Foulds, G. A., & Raven, J. C. (1948). Normal changes in the mental abilities of adults as age advances. Journal of Mental Science, XCIV(394), January, 133-142.

Garfinkel, R., & Thorndike, R. L. (1976). Binet item difficulty: Then and now. Child Development, 47, 959-965.

Gottfredson, L. S. (Ed.). (1997). Intelligence and social policy. Intelligence (Whole Special Issue), 24, 1-320.

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), The rising curve. Washington, DC: American Psychological Association.

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.

Neisser, U. (Ed.). (1998). The rising curve. Washington, DC: American Psychological Association.

Prieler, J. A., & Raven, J. (10/20/02) The measurement of change in groups and individuals, with particular reference to the value of gain scores: A new IRT-based methodology for the assessment of treatment effects and utilizing gain scores. WebPsychEmpiricist http://www.wpe.info/papers_table.html http://home.earthlink.net/~rkmc/vault/priravf/prirav.pdf

Prieler, J. A., & Raven, J. (2002). The measurement of change in groups and individuals with particular reference to the value of gain scores: A new IRT-based methodology for the assessment of treatment effects and utilizing gain scores. Psiholoska obzorja/ Horizons of Psychology, 11(3), 119-150.

Raven, J. (1981). Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.1: The 1979 British standardization 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., Prieler, J., & Benesch. M. (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. http://wpe.info/papers_table.html

Raven, J., Raven, J. C., & Court, J. H. (1998, revised and 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, revised and updated 2004). Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices. San Antonio, TX: Harcourt Assessment.

Schaie, K. W. (Ed.). (1983). Longitudinal studies of adult psychological development. New York: Guilford Press.

Schaie, K. W. (1994). The course of adult intellectual development. American Psychologist, 49(4), 304-313.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124(2), 262-274.

Stefancic, M., & Tomazo-Ravnik, T. (1998). Fifty-two years of secular trend in Ljubljana school children. In E. B. Bodzsar and C. Susanne (Eds.), Secular Growth changes in Europe. Budapest: Eotvos University Press.

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.

Figure 1
Illustration of changes in raw scores on "easy" and "difficult" IRT-Based tests of cognitive ability for identical changes in latent ability
High ability group only

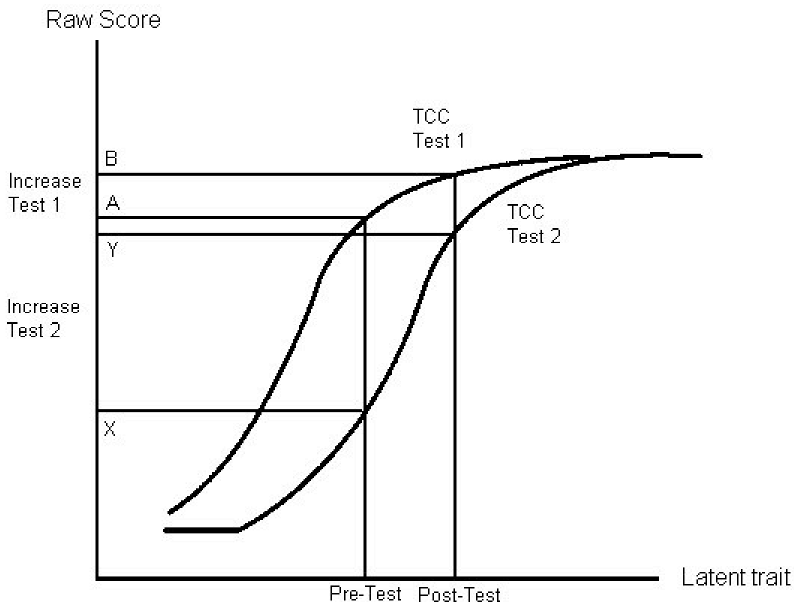

Figure 2
Illustration of changes in raw scores on "easy" and "difficult" IRT-Based tests of cognitive ability for identical changes in latent ability
Low ability group only

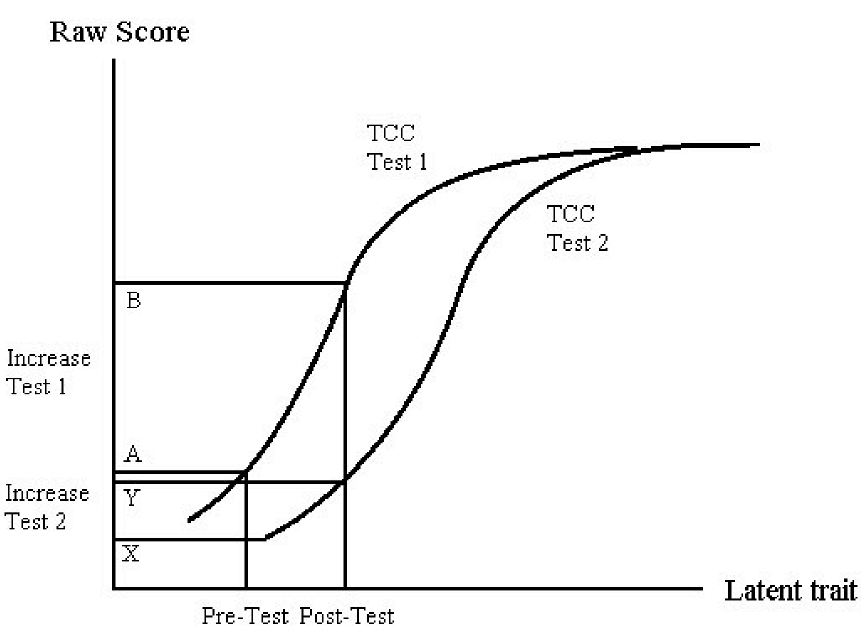

Table1
Some indications of cross-cultural stability
Standard Progressive Matrices
Selection of cross-cultural and birth cohort norms
95th, 50th, and 5th percentiles only
Most European and similar norms omitted

Age in Years (Months)

	8.5	8.5	9	9.5	9.5	10	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10
	8(3)	8(0)	8(9)	9(3)	9(0)	9(9)	10(3)	10(3)	10(3)	10(3)	10(5)	10(3)	10(0)	
	To	to	to	to	to		to	to	to	to	to	to	to	
	8(8)	8(11)	9(2)	9(8)	9(11)	10(2)	10(8)	10(8)	10(8)	10(8)	10(10)	10(8)	10(11)	
Perc	UK	**KW**	UK	UK	**KW**	UK	UK38	UK	TW	PRC	PL	US	**KW**	P&M
95	42	**40**	44	46	**43**	48	48	49	49	50	47	46	**45**	46
50	**31**	**20**	**33**	**36**	**27**	**38**	33	**39**	**38**	**35**	**35**	**34**	**32**	**28**
5	13	**10**	14	14	**11**	17	13	22	19	13	12	15	**12**	11

	11	11.5	11.5	11.5	11	12	12.5	12.5	12.5	12	13	13.5	13.5	14
	10(9)	11(3)	11(3)	11(0)		11(9)	12(3)	12(0)	12(3)		12(9)	13(3)	13(0)	13(9)
	To	to	to	to		to	to	to	to		to	to	to	to
	11(2)	11(8)	11(8)	11(11)		12(2)	12(8)	12(11)	12(8)		13(2)	13(8)	13(11)	14(2)
Perc	UK	UK	QA	**KW**	P&M	UK	UK	**KW**	US	P&M	UK	UK	**KW**	UK
95	50	51	46	**48**	49	52	53	**50**	50	52	54	54	**52**	55
50	**40**	**41**	**38**	**37**	**33**	**41**	**42**	**40**	**39**	**39**	**43**	**44**	**42**	45
5	24	25	18	**16**	12	26	27	**19**	21	14	28	29	**23**	30

	14.5	14.5	14	14	14	15	15.5	15.5	15.5	15.5	20	20	25	20
	14(3)	14(0)	13(9)	13(0)	13(0)	14(9)	15(3)	15(5)	15(0)	15(3)	18	18	20	18
	To	to	to	to	to	to	to	to	to	to	to	to	to	to
	14(8)	14(11)	14(2)	14(11)	14(11)	15(2)	15(8)	15(10)	15(11)	15(8)	22	22	29	22
Perc	UK	**KW**	UK38	AR64	AR00	UK	UK	PL	**KW**	US	UK42	UK92	B	TN
95	56	**53**	53	49	56	57	57	55	**54**	56	55	59	58	56
50	**46**	**45**	44	39	48	**47**	**47**	**47**	**46**	**46**	44	**54**	**49**	**47**
5	33	**26**	23	22	36	33	33	26	**29**	29	23	39	34	19

Notes.
AR (Argentina). The data were supplied by Lilia Rossi Case, Rosa Neer, and Susana Lopetegui. The 1964 data were collected by Direccion de Psicologia - Buenos Aires from 880 children studying in La Plata – Buenos Aires. The year 2000 data were collected by Lilia Rossi Case and her colleagues. The sample consisted of 1,740 young people who were studying, or had finished, high school or secondary level equally distributed between males and females, plus students at public and private schools of La Plata – Buenos Aires, selected according to geographical and socio economic criteria. Full details of the study can be found in Cayssails (2001).

B (Belgium). Data collected between 1984 and 1990 by J. J. Deltour by asking students taking a course in psychometrics each to test 10 adults with equal numbers from each of four educational levels (i.e. not in such a way as to match the total population proportions from each level). The sample was neither stratified for age nor socio-economic status. See, Deltour (1993).
P&M (Pune and Mumbai [Bombay], India). A carefully drawn sample of 5,161 Mumbai (Bombay) and 5,127 Pune young people were tested under the supervision of Professor C. G. Deshpande, by selected personnel from the Department of Applied Psychology, University of Mumbai and the Jnana Prabodhiai Institute of Psychology. The 78 schools involved included Government, Government Aided, and Private Schools teaching in Marathi, English, Hindi, and Gujarathi in the correct proportions. Full details are published by Manasayan (Delhi) as a Supplement to the Indian edition of the SPM Manual.
PL (Poland). Data from the 1989 Polish standardisation. See Jaworowska & Szustrowa (1991).
PRC (People's Republic of China). Data from a 1986 study of 5,108 respondents drawn from all main cities of China. Testing organised by Professor Hou Can Zhang of Beijing Normal University. See Raven, Raven, & Court (2000, updated 2004 [first published 1986]).
QA (Qatar). Data collected by Alanood Mubarak Ahmad AL Thani, Umm Alqura University, Saudi Arabia as part of a Masters degree programme. A random sample of 1,135 children drawn from 7 boys' and 7 girls' public elementary schools in Doha City was tested.
TN (Tunisia). Data collection organised by Riadh Ben Rejeb between 2000 and 2002 following a sampling design intended to yield 5 men and 5 women in each 5-yearly age group between 15 and 60 in each of 6 geographic areas of the country, but which, in fact, yielded a total sample of 509
TW (Taiwan). Data collection from 2506 young people organised by Emily Miao. See Miao (1993).
UK (United Kingdom of Great Britain and Northern Ireland). Main $8^{1}/_{2}$ -15 year olds' data obtained from a nationally representative sample of UK schoolchildren, excluding those attending special schools, tested in 1979 (See Raven, J., 1981). 20 year olds' data derived from the 1992 standardisation of the SPM and APM in Dumfries, Scotland (See Raven, J., Raven, J. C., & Court, J. H., 2000). 1938 and 1942 data put together by J. C. Raven and collaborators following procedures described in Raven, J. (2000).
US (United States of America). National norms compiled by weighting and combining a series of norms for School Districts having known demographic compositions and, as far as possible, derived from representative samples of those districts. See Raven, J. (2000a).

Filename as independent table: CPM50AT.doc

Table 2
Some indications of continuing increases over time
Coloured Progressive Matrices
2004 Data for Poland in the context of 2001 Korean norms and earlier UK and Polish norms
90th, 50th, and 10th percentiles only

Age:		4		
Perc	UK82	PL91	K01	PL04
90		19	21	22
50		**14**	**15**	**18**
10		9	10	12

Age:		5				6		
Perc	UK82	PL91	K01	PL04	UK82	PL91	K01	PL04
90	20	23	26	26	22	27	30	28
50	**15**	**15**	**19**	**22**	**17**	**19**	**24**	**25**
10	10	11	12	17	12	13	15	20

Age:		7				8		
Perc	UK82	PL91	K01	PL04	UK82	PL91	K01	PL04
90	27	29	32	31	31	31	34	33
50	**19**	**20**	**28**	**27**	**23**	**23**	**30**	**29**
10	14	14	20	22	16	14	24	23

Age:		9				10		
Perc	UK82	P91	K01	PL04	UK82	PL91	K01	PL04
90	33	32	35	34	34		35	35
50	**27**	**26**	**32**	**30**	**31**		**33**	**31**
10	18	15	27	25	22		29	25

Notes.
K01 (Korea). Collection of the data in 2001 was organized by Prof. Ho-chan Lim of the College of Rehabilitation of the Korean Nazarene University. All children were attending nursery homes (n=238), kindergartens (n=291), or elementary schools (n=746). Testing was individual or in small groups. Further information (in Korean) is included in the Korean edition of the RPM Manual.
PL04 (Poland). Collection of the 2003-04 Polish data was organized by Irena Grochowska. In 2003, all children aged 3 to 10 living in Zalesie Gorne, a small village near Warsaw were tested using the Board Form of the test. In 2004 the Board and Book Forms of the test were administered to every second child attending 3 schools in Warsaw and 3 in surrounding districts. Because there was little difference between the Board and Book Form norms, these have been merged in the above table. Total, n = 756.
PL91 (Poland). 1991 data come from Jaworowska, A., & Szustrowa, T. (2003), Test Matryc Ravena w wersji Koloroweij. Warszawa: Pracownia Testow Psychologicznych. Total, n = 588.
UK82 (United Kingdom of Great Britain and Northern Ireland). Data comes from the 1982 standardization of the CPM in Dumfries (see Raven, Raven, & Court, 1999). Total n = 598.
n.b: The CPM has only 36 items. As a result there is a ceiling effect in the older age groups.