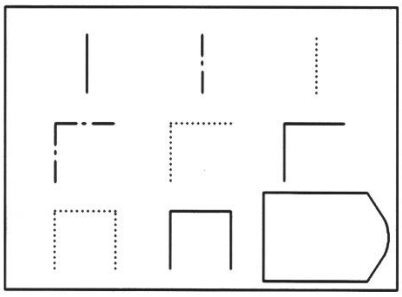
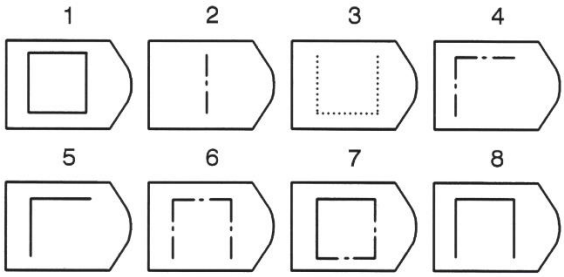
Main article 3,707 words
Including Notes and References 4,943
Version date: 7 September 2025

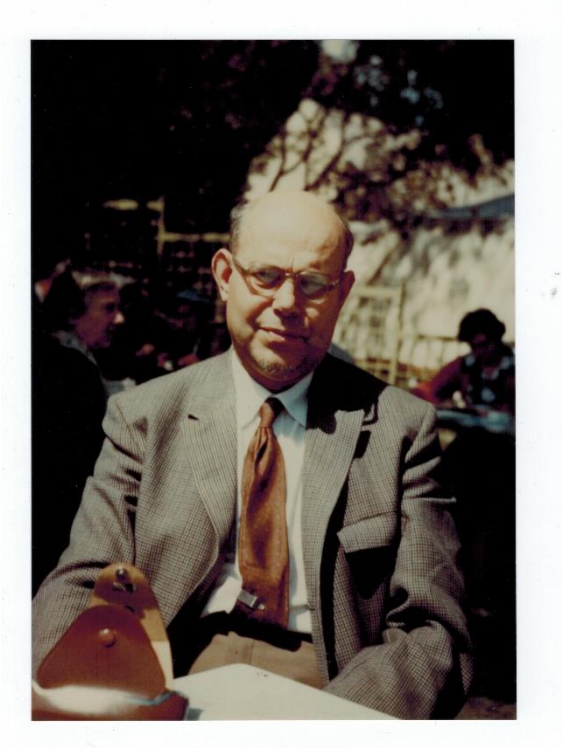
Highlights from 90 Years With the *Raven Progressive Matrices*[*]

John and Jean Raven





[*]We need to do more than acknowledge the help of someone who prefers to remain anonymous in preparing this paper.



Raven's *Progressive Matrices* first saw the light of day in 1936[1] and thus turns 90 in 2026.

So we thought it would be good to have some kind of retrospective review.

The 90-year development of the Raven Progressive Matrices.

The publication of the original test was quickly followed by the *Coloured Progressive Matrices* and, much later, what became known as the *Advanced* Progressive Matrices. And, later still, the *Parallel Versions.* Much later, as work with the test demonstrated an intergenerational increase in scores[2] (the "Flynn effect"), *The Standard Progressive Matrices* ***Plus***[3] was developed[4] to restore the discriminative power that the original test had when it was developed. Importantly this last test consisted of parallel versions of the original items or others constructed following the same principles, thus making it possible to continue the kind of research that had contributed to the establishment of the "Flynn effect"[5].

In the interim, *Raven's Progressive Matrices* have (largely through mandatory testing programmes by military and educational systems) directly influenced the lives and livelihoods of many million people worldwide.

What do the *Raven Progressive Matrices* tests measure?

It is well known that Spearman[6] was among the first to draw attention to the fact that most human abilities are positively correlated, with some being more highly correlated than others. This cluster of abilities he named ***g,*** vigorously declining to name the factor "general intelligence". Less well known is the fact that he saw ***g*** as being made up of two distinct abilities – not factors – which worked closely together. These he termed *eductive* (from "educe", i.e. 'draw out', meaning) and *reproductive* abilities. He wrote "To understand the respective natures of eduction and reproduction – in their trenchant contrast, in their ubiquitous cooperation and in their genetic interlinkage – to do this would appear to be for

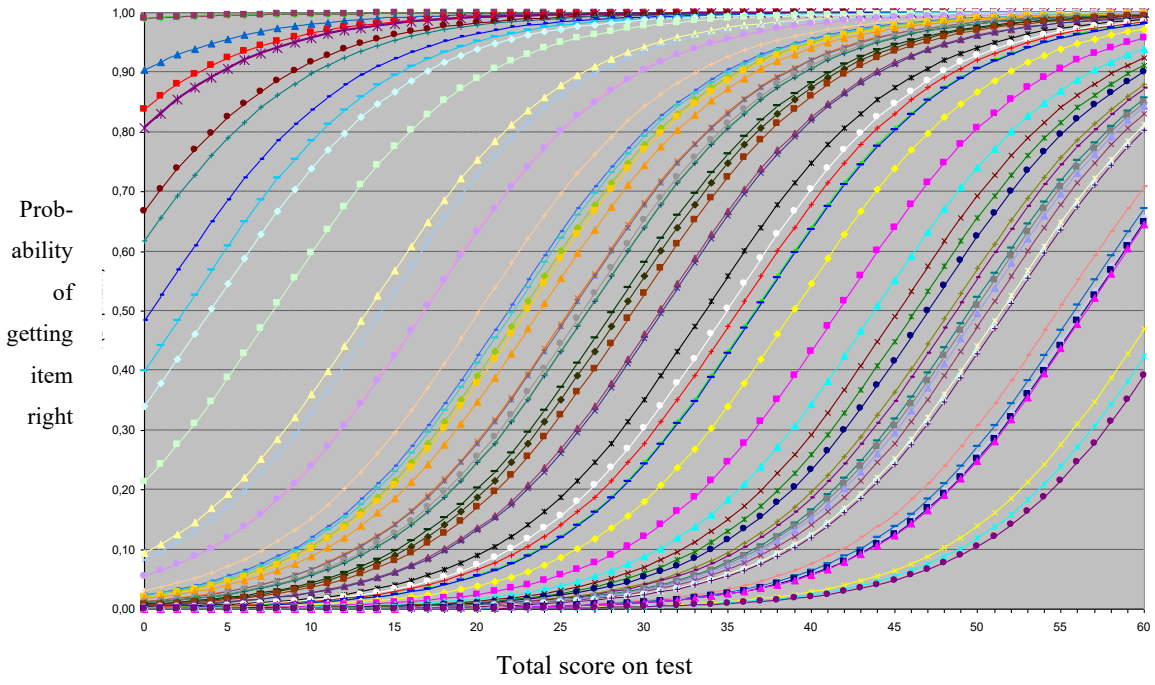
the psychology of individual abilities and even for cognition in general, the very beginning of wisdom". Cattell[7] made somewhat similar observation but did consider them factors and named them "fluid" and "crystallised" intelligence. However, as his colleague John Horn[8] later pointed out, reproductive ability is not a crystallised form of eductive ability. The abilities differ at birth, respond to different environmental stimuli, and predict different things in life.

Raven created measures of these two constructs via his *Progressive Matrices* and *Vocabulary* tests[9].

In designing these tests Raven adopted a measurement model that was not widely understood at the time and, indeed, contradicted the measurement model adopted by most psychometricians (most of whom were factor analysts, who generate scores on factors by first identifying the items having high loadings on a factor and then calculating how many of those items a particular individual endorses). In contrast, Raven sought to develop a series of increasingly difficult items and demonstrate that these items did indeed form a scale somewhat analogous to a foot-rule or meter stick.

Later versions of this model have become known as *Item Response Theory.* The results of applying this in the course of standardising the *Standard Progressive Matrices* ***Plus*** test (of which more below) in Romania are shown in the Figure below[10].

Figure 1
Standard Progressive Matrices *Plus*
Romanian Data
1-Parameter Model Item Characteristic Curves for All 60 Items



In effect[11], what each of the graphs in the Figure shows is the number of people with each total score that got the item right. Thus, if one looks at the graphs for the items about one-

quarter of the way along, few people with scores less than 12 get them right but the proportion doing so increases until almost everyone with a score over 35 gets them right.

What this shows is that the abilities required to solve the more difficult items build on and extend those required to solve the easier items; they form a continuous series. There are no leaps, plateaux, or "metamorphoses"[12]. More specifically, the earlier "perceptual" items form part of exactly the same series as the more "analytic" items.

In effect, what these analyses, taken together with those summarised below, show is that eductive and reproductive abilities are every bit as "real" and measurable as "energy".

Stability of test properties across socio-economic and cultural groups.

It has often been said that the "abstract" nature of the items disadvantages certain groups. Vodegel Matzen[13] set out to test this directly by making the items more realistic, i.e. substituting such things as hats, bananas, and faces for the "abstract" components of the items whilst retaining their logic. Somewhat surprisingly, the effect was to make all the items easier for everyone but did not substantially change their order of difficulty or differentially benefit "disadvantaged" groups.

This robustness has been repeatedly demonstrated by correlating the item difficulties established separately in a wide range of socio-economic and cultural groups (see, for example, tables below[14 15]).

Table 1

Correlations Between SPM Item Difficulties as Determined Separately for Eight Socio-Economic Status Groups.
(Decimal Point Omitted)

1								
2	99							
3	99	99						
4	98	99	99					
5	97	98	99	99				
6	98	99	99	99	99			
7	95	96	98	98	99	98		
8	95	96	98	98	99	99	99	
	1	2	3	4	5	6	7	8

Table 2

Standard Progressive Matrices

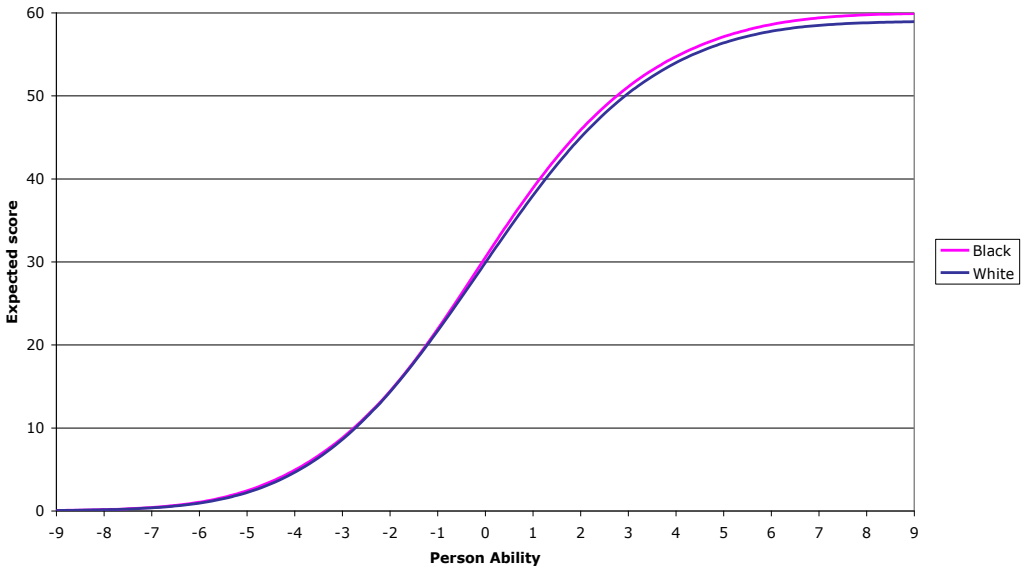
Correlations Between Item Difficulties Calculated Separately Within Specified Groups

		Westown Black	Westown White	Westown Hispanic	Westown Asian	Westown All	Des Moines	China
Westown	Black							
	White	98						
	Hispanic	100	98					
	Asian	98	99	98				
	All	99	99	100	99			
Des Moines		99	99	99	97	99		
China		95	94	94	96	95	96	
UK 1979		99	97	99	98	99	99	97

Unlike the Westown sample, the Des Moines sample covers all ages not just adolescents.
(Decimal point omitted and rounded to two decimal places).

This has been further striking illustration of this can be seen in the following Figure, generated by Nicola Taylor[16], which shows the Test Characteristic Curves for Black and White applicants for jobs in the mines of South Africa[17] (before looking at the figure, note that there are huge differences in the mean scores of miners from different backgrounds).

Figure 2
Test Characteristic Curves for the SPM



The general conclusion from such work is that the same score means the same thing in all (or most) cultural groups. Cultural differences are real and cannot be attributed to such things as "test bias"[18]… but this does not mean that they cannot be "explained".

For example, the norms for Indian tribal areas[19] are well below those for urban populations[20]. This is not surprising given that those concerned often live in mud huts that are regularly washed away, have limited means of earning a living, very little formal education, and low levels of literacy. What is surprising is that the test scales in the usual way and has similar variance of scores within age groups.

Similarities and differences across socio-economic, cultural, and ethnic groups.

Documenting differences in cognitive abilities across socio-economic, cultural, and ethnic groups is extremely problematic because of the difficulty of generating representative samples.

Numerous studies based on unrepresentative "convenience" "samples" have been published, thereby generating endless confusion.

For the 1979 British standardisation of the *Standard Progressive Matrices*[21], the Centre for Environmental Studies first conducted a cluster analysis of types of social-economic areas within the UK. This resulted in the identification of 7 clusters. The standardisation was conducted on representative samples from within each of these. It emerged that there were major differences between the scores attained in these areas. The results from the different areas were combined and weighted to give overall norms for the UK. These provided reference data against which to view the results obtained in other countries and regions such as Ireland, various cities in the USA[22], China, Poland, Romania, Slovenia, Kuwait, and Pakistan[23]. The general conclusion seems to be that the more one approaches representative samples, the more similar become the norms for societies with a tradition of literacy *at any point in time.*

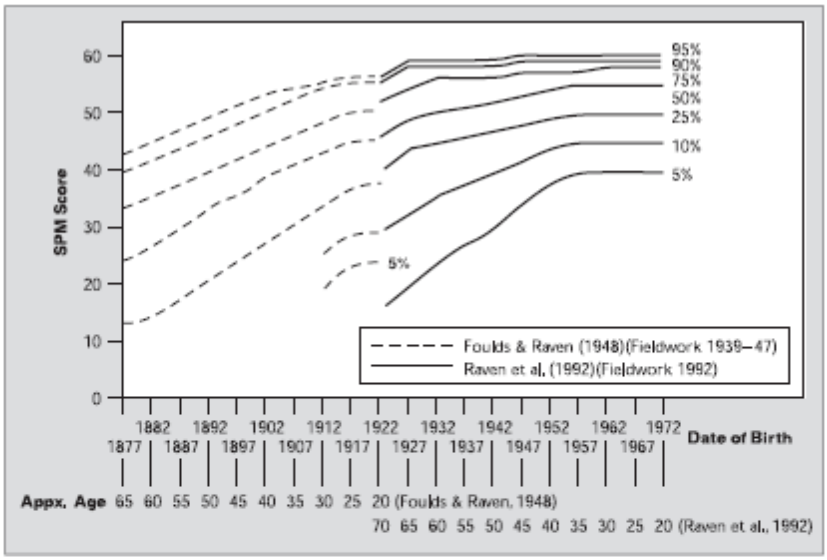
Changes over time (i.e. the "Flynn effect").

Most test publishers provide data purporting to show how scores change with age. These are, however, mostly derived from a series of cross-sectional rather than longitudinal studies.

Flynn[24] shocked the world by suggesting that these apparent changes with age were largely attributable to the date on which the studies had been conducted rather than age.

The results of re-plotting the British 1940s and 1979 data by date of birth are shown in Figure 3.

Figure 3
100 Years of Eductive Ability, by Date of Birth



The figure graphs the percentile norms obtained by adults of different ages (and thus dates of birth) on the *Standard Progressive Matrices* when a sample was tested circa 1942 and in 1992. It will be seen that those born in 1922 and tested circa 1942 (and thus approximately 20 years of age) obtained similar scores to those born in 1922 and tested in 1992 when they were 70 years of age.

It is clear that there has been a continuous increase in scores over the years and that the *Standard Progressive Matrices* (SPM) is no longer discriminating between the scores of those in the top 25th percentile. This has become known as the "Flynn effect". What is more, the cross-cultural data available for the SPM shows that the increase has occurred in many cultures with very different educational and social systems and that, therefore, the most common explanations (education, access to media, changes in family sizes, etc) do not hold up. More likely it is due to the same things that have led to a worldwide intergenerational increase in height[25]. In passing it may be noted that eductive ability scores have increased more than reproductive ability despite the latter seemingly being more dependent on education and media.

Some recent studies suggest a possible decline in scores which parallels the simultaneous worldwide decline in height which most likely stems from similar complex causes - although psychologists have tended to point toward digital dependency associated with a need for reduced cognitive effort in the AI era. In fact, the picture is extremely complicated, with some studies reporting not merely no decline but a continuing increase[26].

Heritability.

Over the years, there had been considerable dispute over estimates of the heritability of "intelligence" arising from small studies and the nature of the test used. This was largely brought to rest with the publication of the results of what became known as the Minnesota Twin Study, which relied in part on the RPM[27]. Significantly, the actual statistical estimate (later rounded to 70%) for the heritability of IQ was identical with the much-disputed estimate of Cyril Burt.

"Intelligence", social mobility, and social structure.

Also born in 1936 was the second tranche[28] of another hugely important study in which the author found himself to be a participant.

For several decades, the entire population of 11 year olds in Scotland took an intelligence test with a view to their allocation to secondary schools.

The Scottish Council for Research in Education arranged to collect more comprehensive data on, and thereafter to re-test and re-interview, samples of those born in 1921 and 1936

Two studies were involved in the latter. In the first of these, a team at SCRE arranged for the sample participants to be re-visited and re-tested on a regular basis until they were 27 years old. Much later, SCRE contacted a sample of those living in the Lothian region of Scotland and followed them up until today[29].

Among other things, this second study revealed that the correlation between the test scores at age 11 and age 77 was 0.64[30]. Thus IQ seems to be more stable than might be thought.

Hope[31] gained access to the records of the first of these data pools and showed, among other things, that some two thirds of social mobility, *both upward and downward,* could be accounted for by 11 year old's test scores relative to those expected of their background. Much the same was true in the US, although it took the US to age 40 to achieve the degree of

sorting by intelligence that Scotland achieved at age 11. In other words, whatever about its ability to predict a wide variety of things[32], "IQ" predicts *level* of job attained and retained[33]. High SES parents are much less able to keep their less able children "up" than is commonly thought.

Problems associated with the measurement of change.

Numerous researchers have set out to assess the differential effect of some intervention on *Progressive Matrices* scores, often at different levels of ability.

This turns out to be extremely problematic.

Here's the problem.

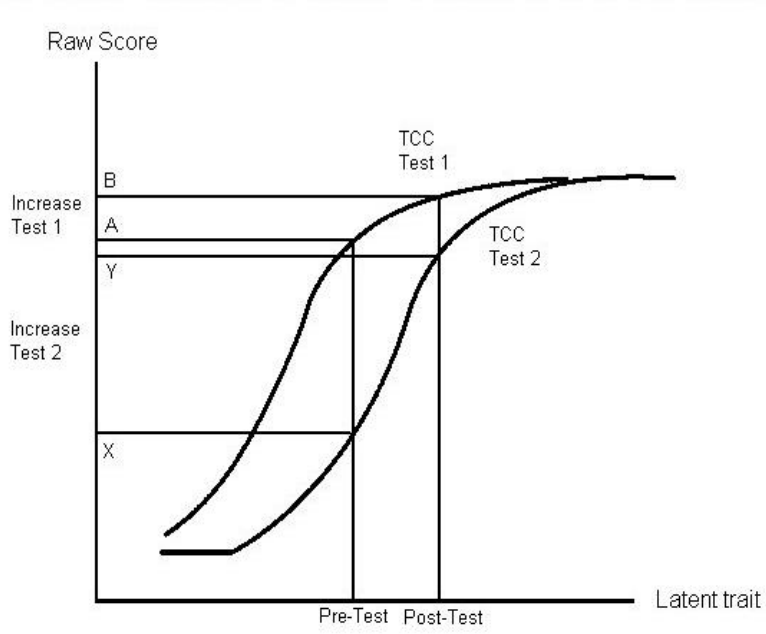
On most tests, a difference between, say, a score of say 7 and 11 does not mean the same thing as a difference between 13 and 17. Furthermore, any given score can be achieved in a number of different ways. In short, the metrics are *arbitrary*[34]. This makes it very difficult to assess change – e.g. to compare the gains made by more and less able groups or individuals in response to an intervention – in a meaningful way[35].

But the problems do not end there.

Even with tests which both conform to Item Response Theory and have items which increase regularly in difficulty, there are still problems arising from non-linear test characteristic curves.

Figure 4 illustrates the problem.

Figure 4
Illustration of Changes in Raw Scores on "Easy" and "Difficult" Measures of Ability for Identical Changes in Latent Ability



High ability personnel only.

From Prieler and Raven (2008)

If we seek to compare the relative gains by a high and low ability group in response to some intervention and we employ a test having the Test Characteristic Curve shown on the left in Figure 4, the mean score of the high ability group increases from A at the pretest (i.e. before the intervention) to B at posttest (i.e. after the intervention). This appears to be a relatively small increase. But if we use the more difficult test shown on the right, the same increase in score on the latent trait shows up as a *huge* increase in raw score, moving from X to Y.

Clearly, the same effect occurs at the other end of the scale.

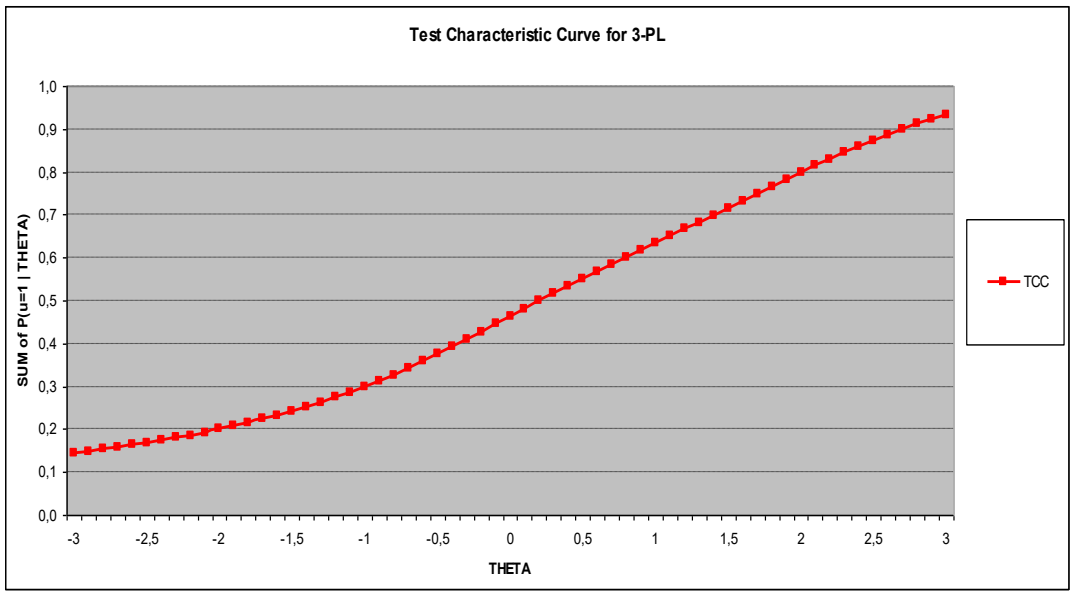
The point is clear: The relative gains of low and high ability groups will appear to be very different depending on the test used.

More generally, the apparent magnitude of any real increase in latent ability depends (a) the general difficulty level of the test relative to the ability tested (b) the shape of the test characteristic curve and (c) the sector of the curve at which the change occurs.

But can the problem be overcome by using a test with a *linear* Test Characteristic Curve?

Maybe the SPM+ yields such a thing.

Here is it's Test Characteristic Curve[36]:



Physiological bases.

Numerous researchers, going right back to Spearman's very early work, have attempted to link various measures of reaction time to "intelligence". It was therefore something of a surprise to learn that Deary[37], had established correlations of around -0.40 between the RPM and average choice reaction time. (Note, however, that the latter is not a timed test; rather it measures how long someone needs to accurately discriminate between the length of two lines.).

The differential assessment of eductive and reproductive abilities to illuminate other issues.

Raven intended the differential assessment of eductive and reproductive abilities to be useful in individual assessment and, for all we know, this may have been the case in clinical, counselling, and school psychology. However, there have been few reports of this happening in the research literature. One exception may be the researches of MacKay and Dawson in relation to autism. MacKay[38], with great difficulty, first established the meaningfulness of "autism". He then noticed that it seemed to be identifiable from a major discrepancy between eductive and reproductive abilities. Not having access to Raven's *Vocabulary Tests,* Dawson[39] had treated the whole Wechsler test as a measure of reproductive ability and explored the implications of discrepancies between the Weschler and RPM scores. MacKay[40] reports that her findings become still clearer if one substitutes one of Raven's vocabulary scales for the Wechsler.

AI and the Matrices.

In an extraordinary set of developments, which include the utilisation and development of numerous 'RPM look-alikes"[41], the RPM have been used, on the one hand, to understand machine learning and, on the other hand, AI has been used to understand the processes which lie behind the solution of RPM problems[42]. At the time of writing, it seems that it has been both more difficult than expected for AI to solve some of the classic items and to generate more difficult items, but, whatever about that, the field seems to be exploding.

*****

Some Limitations of the Framework.

Authors of papers submitted to journals are routinely asked to declare the limitations of their work. But never have we seen any as devastating as Spearman's own critique of his own work. Thus he wrote: "Every normal man, woman, and child is … a genius at something … It remains to discover at what … This must be a most difficult matter, owing to the very fact that it occurs in only a minute proportion of all possible abilities. It certainly cannot be detected by any of the testing procedures at present in current usage."[43].

In point of fact, "measuring", or at least recording, these diverse abilities depends on the adoption of a very different "psychometric" model - which J.C.Raven sought to develop throughout his life and which we have sought to elaborate[44].

Be that as it may, the hegemony of what might be termed a single-factor concept of "ability" has had a devastating effect on the lives of millions of individuals, workplaces, and society[45].

The tests are widely misused, with the use of cut-off scores in, for example, entry to special educational programmes[46], while pervasive, being perhaps the least important. Wild generalisation about the meaning and implications of the scores is perhaps more common.

The assessment framework, with its excessive reliance on the notion of "ability", is recursively linked to the legitimisation and production of a dysfunctional hierarchical society.

There is not space to elaborate on this here, but readers might like to turn to *Intelligence, engineered invisibility, and the destruction of life on earth,* chapter 19 in *Uses and Abuses of Intelligence*[47].

References

American Psychologist. (2006). Special Issue on arbitrary metrics, especially articles by Kazdin and Embretson. *American Psychologist 61(1).*

Bouchard, T.J., Lykken, D.T., McGue, M., Segal, N.L. & Tellegen, A. (1990*).* Sources of human psychological differences: The Minnesota study of twins reared apart*. Science. 250, (4978): 223–8.*

Bouchard, T.J. & McGue, M. (2003). *Journal of Neurobiology, 54 (1)*, 4–45. *doi:10.1002/neu.10160. PMID 12486697*

Cattell, R.B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193.

Dawson, M., Soulières, I., Gernsbacher, M.A. & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, August, 18(8), 657–662. doi:10.1111/j.1467-9280.2007.01954.x.

Deary, I.J., Whalley, L.J., Crawford, J.R., Starr, J.M. (Undated). *The Stability of Individual Differences in Mental Ability From Childhood to Old Age*. University of Edinburgh: Department of Psychology.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H.C. (2004). The Impact of Childhood Intelligence on Later Life: Following Up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86*(1), 130–147. https://doi.org/10.1037/0022-3514.86.1.130

Deary, I.J., Michelle, T., Taylor, D., Hart, C.L., & Wilson, V., Smith, G.D., Blane, D., & Starr, J.M. (2005). Intergenerational social mobility and mid-life status attainment: Influences of childhood intelligence, childhood social factors, and education. *Intelligence*, *33(5)*, 455-472.

Deary, I.J., Gow, A.J., Taylor, M.D., Corley, J, Brett, C., Wilson, V., Campbell, H., Whalley, L.J., Visscher, P.M., Porteous, D.J. & Starr, J.M. (2007). *The Lothian Birth Cohort 1936: A study to examine influences on cognitive ageing from age 11 to age 70 and beyond.* BMC Geriatr. 2007 Dec 5;7:28. doi: 10.1186/1471-2318-7-28. PMID: 18053258; PMCID: PMC2222601.

Deary, I.J., Pattie, A. & Starr, J.M. (2013), The stability of intelligence from age 11 to age 90 years: the Lothian birth cohort of 1921. *Psychological Science*, 24(12), 2361-2368.

Deshpande, C.G., & Patwardhan, V. (2008). *Standard Progressive Matrices* Norms for Indian Tribal Areas. In J. Raven & J. Raven (Eds.) *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics*. Unionville, New York: Royal Fireworks Press; Edinburgh, Scotland: Competency Motivation Project; Budapest, Hungary: EDGE 2000; Cluj Napoca, Romania: Romanian Psychological Testing Services SRL. (Chapter 16, pp. 392-401).

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

Gottfredson, L.S. (1997). Why ***g*** matters: The complexity of everyday life. *Intelligence, 24*, 79-132.

Hope, K. (1984). *As Others See Us: Schooling and Social Mobility in Scotland and the United States*. New York: Cambridge University Press.

Horn, J.L. (1994). Theory of fluid and crystallized intelligence. In R.J. Sternberg (Ed.), *Encyclopedia of Human Intelligence*, pp. 443-451. New York: Macmillan.

MacKay, T., Knapp, M., Boyle, J.M., Iemmi, V., Connolly, M. & Rehill, A. (2018). *The microsegmentation of the autism spectrum: Economic and research implications for Scotland.* Edinburgh: The Scottish Government. (pages 63-75). https://www.gov.scot/publications/microsegmentation-autism-spectrum/

MacPherson, J.S. (1958). *Eleven Year Olds Grow Up*. London: University of London Press.
Małkiński, M. & Mańdziuk, J. (2025). Deep Learning Methods for Abstract Visual Reasoning: A Survey on Raven's Progressive Matrices. *ACM Computer Surveys,* 57(7), Article 166 (February). Also available at https://www.researchgate.net/publication/358260660_Deep_Learning_Methods_for_Abstract_Visual_Reasoning_A_Survey_on_Raven%27s_Progressive_Matrices.
Maxwell, J.N. (1961). *The Level and Trend of National Intelligence: The Contribution of the Scottish Mental Surveys*. London: University of London Press.
Maxwell, J.N. (1969). *Sixteen Years On*. Edinburgh: Scottish Council for Research in Education.
Prieler, J.A., & Raven, J. (2008). Problems in the measurement of change (with particular reference to individual change [gain] scores) and their potential solution using IRT. In J. Raven & J. Raven (Eds.) *Uses and Abuses of Intelligence* (cp. cit.) chapter 7, pp. 173-210. Also available at http://eyeonsociety.co.uk/resources/UAIChapter7.pdf and https://www.researchgate.net/publication/340900090_Problems_in_the_Measurement_of_Change_with_Particular_Reference_to_Individual_Change_Gain_Scores_and_their_Potential_Solution_Using_IRT
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*. H.K.Lewis/Oxford Psychologists Press/San Antonio, TX: Harcourt Assessment. (This Supplement also contains the results of the Irish standardisation.)
Raven, J. (1984/1997). *Competence in Modern Society: Its Identification, Development and Release*. Unionville, New York: Royal Fireworks Press. www.rfwp.com (First published in 1984 in London, England, by H. K. Lewis.) Also available at: https://www.researchgate.net/publication/337925795_Competence_in_Modern_Society and http://eyeonsociety.co.uk/resources/Competence-in-Modern-Society-John-Raven.pdf
Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1-48.
Raven, J. (2000). *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. (2008). Intelligence, engineered invisibility, and the destruction of life on earth. In J. Raven & J. Raven (Eds.) *Uses and Abuses of Intelligence* (op cit), chapter 19, pp. 431-471. Also available at http://www.eyeonsociety.co.uk/resources/UAIChapter19.pdf and https://www.researchgate.net/publication/255580747_Intelligence_Engineered_Invisibility_and_the_Destruction_of_Life_on_Earth_The_nature_development_and_identification_of_talents_going_well_beyond_g_the_network_of_social_forces_rendering_them_invisible
Raven, J. (2014). *Our incompetent society (with a discussion of some of the competencies needed to transform it).* http://eyeonsociety.co.uk/resources/Incompetent-society-v3.pdf Also available at https://www.researchgate.net/publication/337928723_Our_Incompetent_Society
Raven, J. (2021) Commentary on 'The *Raven's 2 Progressive Matrices* Tests and Manual, (Pearson, 2018)' *Testing International* Volume 46, pages 14 – 25. Aso available at http://eyeonsociety.co.uk/resources/TI-December-2021-edited.pdf and https://www.researchgate.net/publication/354190448_Commentary_on_The_Raven's_2_Progressive_Matrices_Tests_and_Manual

Raven, J., Prieler, J. & Benesch, M. (2008). Using the Romanian data to replicate the IRT-based Item Analysis of the SPM+: Striking achievements, pitfalls, and lessons. In J. Raven & J. Raven (Eds.) *Uses and Abuses of Intelligence* (op. cit.) chapter 5, pp. 127-159. Also available at: http://eyeonsociety.co.uk/resources/UAIChapter5.pdf and https://www.researchgate.net/publication/340899446_Using_the_Romanian_Data_to_Replicate_the_IRT-_Based_Item_Analysis_of_the_SPM_Striking_Achievements_Pitfalls_and_Lessons_Chapter_5_in_Raven_J_Raven_J_ed_Uses_and_abuses_of_intelligence_published_by_Roy

Raven, J., Raven, J. C., & Court, J. H. (2000, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Sections 1 to 7 with three Research Appendices. San Antonio, TX: Harcourt Assessment. The *Manual* is not available online, but a summary of the contents of the *Sections* will be found at. https://www.researchgate.net/publication/368919615_Manual_for_Raven's_Progressive_Matrices_and_Vocabulary_Scales_Summary_of_Contents_of_All_Sections

Raven, J., & Raven, J. (Eds.). (2008). *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics.* Unionville, New York: Royal Fireworks Press; Edinburgh, Scotland: Competency Motivation Project; Budapest, Hungary: EDGE 2000; Cluj Napoca, Romania: Romanian Psychological Testing Services SRL. Also available at http://eyeonsociety.co.uk/resources/fulllist.html#uses_and_abuses and https://www.researchgate.net/publication/340665403_Uses_and_Abuses_of_Intelligence_Studies_Advancing_Spearman_and_Raven's_Quest_for_Non-Arbitrary_Metrics

Raven, J, & Raven, J. (2018). *John Carlyle Raven and his legacy. http://eyeonsociety.co.uk/resources/J-C-Raven-Biography.pdf.;*and *https://www.researchgate.net/publication/341150479_JOHN_CARLYLE_RAVEN_1902_-_1970_and_his_legacy*

Raven, J.C. (1936). *Mental Tests Used in Genetic Studies: The Performances of Related Individuals in Tests Mainly Eductive and Mainly Reproductive*. M.Sc. Thesis, University of London.

Raven, J.C., Styles, I., and Raven, M. (1998). *Raven's Progressive Matrices: SPM **Plus** SETS A-E.* Oxford, England: Oxford Psychologists Press/The Psychological Corporation/Pearson test publishers.

Scottish Council for Research in Education (1933). *The Intelligence of Scottish Children*. London: University of London Press.

Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. London, England: MacMillan.

Taylor, N. (2008). Raven's *Standard and Advanced Progressive Matrices* among adults in South Africa. In J. Raven & J. Raven (Eds.) *Uses and Abuses of Intelligence.* (Op. Cit.), chapter 15, pp. 371-391.

Vodegel-Matzen, L.B.L. (1994). Performance on Raven's Progressive Matrices. Ph.D. Thesis, University of Amsterdam

Zhuo, T. & Kankanhalli, M. (2020). Solving Raven's Progressive Matrices with Neural Networks. *ArXiv:2002.01646v2 [cs.CV]* 6 Feb 2020. Also in Doctemp.

[1] Raven, J.C. (1936)
[2] Raven, J. (2000)
[3] Raven, J.C., Styles, I., and Raven, M. (1998)
[4] The enormous amount of international work required to do this is summarised in Raven, J., Raven, J.C. & Court, J.H. (2000, updated 2004). Particular mention should, however, be made of the work of Irene Styles in

using Item Response Theory to help select the final items and the work of Styles and M.A.Raven in generating the new items.

[5] Numerous "look-alike" versions also appeared, especially on the internet. Pearson Test Publishers, who had taken over the publication of the RPM materials around 2004, also published the *Raven's Progressive Matrices 2* test. However, since many of the items are not presented in matrix format we do not need to consider it further here (but see Raven, 2021).

[6] E.g. Spearman (1927)

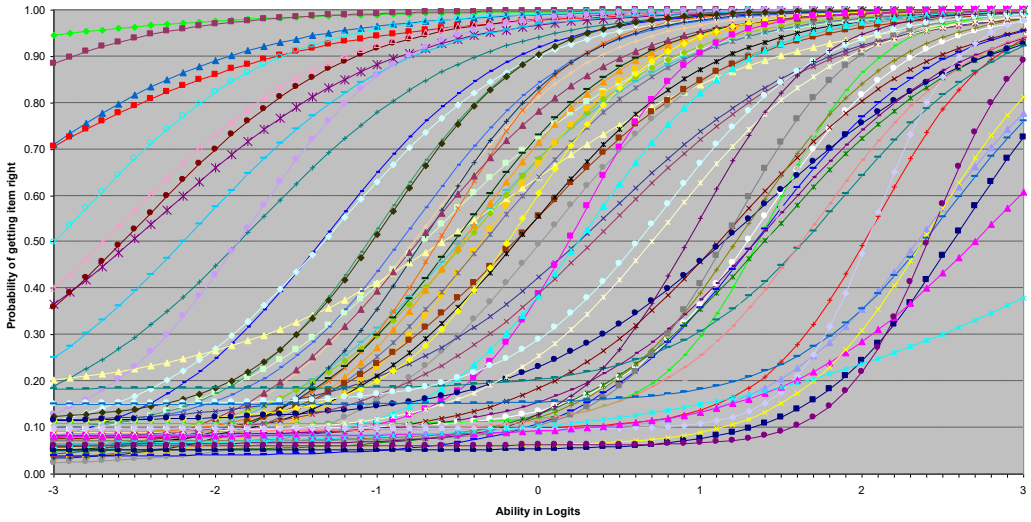
[7] Cattell (1943)

[8] Horn (1994)

[9] There were/are several versions of each of these.

[10] From Raven, Prieler and Benesch (2008).

[11] The picture is, in reality, somewhat more complicated in that the Figure is in fact based on derivative indices rather than raw scores. Furthermore the Figure comes from fitting a 1-parameter model, which heavily smooths the graphs. As can be seen from the Figure below, fitting a 3-parameter model yields a less satisfying picture.



[12] This does not mean that these things do not occur in individual development.

[13] Vodegel Matzen (1994)

[14] From *Research Supplements 1* and 3 in the *Manual for Raven's Progressive Matrices* (Raven, J.C. et al 2000, updated 2004)

[15] "Westown" is a US city that did wish to be identified.

[16] Taylor (2008)

[17] Prospective miners still come from very different tribes who speak different languages and often do not understand each other, never mind English.

[18] This became particularly important when it emerged that there were significant differences between the scores obtained by pupils studying in school districts catering for different socio-economic groups within the USA. These led to numerous lawsuits accusing those concerned of "test bias" which were ameliorated by our results.

[19] India is a vast country, having a population of 1,600 million (compared with the approximately 40 million of most European countries). Tribal areas are to be found throughout from the Indian ocean to the mountains of the Himalaya.

[20] Deshpande & Patwardhan (2008)

[21] Raven (1981)

[22] Raven (2000)

[23] The results were reported in various editions of the *Manual* (Raven, Raven & Court [2000 updated 2004]) and *Uses and Abuses of Intelligence* (Raven & Raven, 2008).

[24] Flynn, (1984)

[25] For a full discussion see Raven, J. (2000).

[26] https://en.wikipedia.org/wiki/Flynn_effect

[27] Bouchard, et al. (1990, 2003)

[28] See Scottish Council for Research in Education (1933, MacPherson, J.S. (1958) and Maxwell (1961 &1969) for tranche 1. In passing, note that the Maxwell books contain the first clear evidence of what became known as "The Flynn Effect".
[29] Deary et al. (2004, 2007)
[30] Deary, et al. (undated, 2013)
[31] Hope (1984)
[32] Gottfredson (1997)
[33] A parallel study of with the later sample of the same birth cohort by Deary et al (2005) came to similar conclusions.
[34] A special issue of *The American Psychologist* 2006, vol 61, was devoted to this topic.
[35] A more detailed discussion will be found in Prieler & Raven (2008).
[36] This may solve the problem for this test which is known to yield similar results for different ability groups, but unfortunately, there can be no guarantee that, just because the overall distribution is linear the distributions for different ability groups will be similar.
[37] I have not been able to trace the original reference for this, but Deary (personal communication) reports similar results from the Lothian Birth Cohort study.
[38] MacKay, et al. (2018)
[39] Dawson, et al. (2007)
[40] Personal communication.
[41] Małkiński, M. and Mańdziuk, J. (2025)
[42] Zhuo and Kankanhalli, (2020)
[43] Spearman (1926)
[44] See Raven and Raven (2018). More specifically, recognising and recording multiple talents – or "competencies" – requires a "descriptive" framework analogous to that used in chemistry or biology. (see Raven (1984/1997)
[45] Raven (2014)
[46] It is not widely appreciated how unreliable are the basic norms for tests. To illustrate: Suppose one has a random sample (obtaining which is itself a problem) of 1000 people aged 5 to 15, that means one has 50 people in each 6-monthly age group. That means the 5th percentile for any age group is actually the score of the 2.5th person in that age group in the sample. Such figures are obviously extremely unreliable. So the published figures have been smoothed to take account of those for adjacent age groups and higher percentiles. Figures for percentiles below 5 (or above 95) are then extrapolations based on assumptions about the shapes of the within-age distributions – which are rarely Gaussian. Beyond that, young children are developing rapidly at the ages at which allocation to special education programmes are typically made, so choice of the next age category (the scores for which may be considerably higher) may be more appropriate. Considerable discretion, rather than strict adherence to guidelines, is required (even if the procedure is justified at all).
[47] Raven, J. (2008)