

Spearman's Raven Legacy

John Raven*

SCOTLAND

Spearman's discovery of *g* is well known. Less well known is that Spearman saw *g* itself as being made up of two psychologically very different abilities which nevertheless worked closely together. These he termed *eductive* and *reproductive* abilities. The first term comes from the Latin root *educere* and thus refers to "the ability to draw meaning out of confusion". The second refers to the ability to recall acquired information. Research over the intervening century has strongly confirmed Spearman's observations (Carroll 1993, Horn, 1994) if not the fluid-crystallized terminology introduced by Cattell and Horn. As Horn (1994) now also recognises, one is *not* a crystallised form of the other. The abilities are distinct at birth, have different genetic origins, are influenced by different aspects of the environment, and predict different things in life. But Spearman also observed that *g* itself had emerged from analyses of the correlations between the tests employed in a so-called "educational" system which does not merit the name because it does not "draw out" the diverse talents of those involved. He also noted that indexing those talents meant developing an alternative psychometric model. He made the point as follows: "Every normal man, woman and child is a genius at something ... the problem is to identify at what ... this must be a most difficult task because it occurs in only a minute proportion of circumstances ... this cannot be done with any of the procedures in current use ...".

* John Raven can be contacted at:
30 Great King Street,
Edinburgh EH3 6QH, Scotland
Phone: ++ 44 131 556 2912
Fax: ++ 44 131 556 3754
e-mail: jraven@ednet.co.uk

The purpose of this article is to summarise some of the research that my father and I have conducted in both the narrower and wider of these areas in such a way that readers whose appetite is whetted by what is said can find their way to relevant recent publications.

Ageing and Cognitive Ability

My father was one of a number of psychologists who thought that they had, through cross-sectional studies, conclusively demonstrated that eductive, but not reproductive, ability declined steadily from about 22 years of age onward.

The interpretation of these well established data was dramatically called into question by Flynn (1984, 1987). We have added three things to Flynn's conclusion. Whereas Flynn brought together *Raven Progressive Matrices (RPM)* data from 18 year olds in several countries, we¹ compiled cross-age, cross-sectional, data from samples of the British population tested around 1942 and 1992. What these data show is that the eductive ability of the 1922 birth cohort – who were 20 when tested in 1942 and 70 when tested in 1992 – has *not* declined as they aged. Instead, what one sees is a continuous, and dramatic², increase in the scores of people of all levels of ability with date of birth. The data that most psychologists had previously interpreted as showing that reasoning ability declines with age thus reveal something much more interesting: The ability to make meaning out of confusion has been improving dramatically over the last century and that increase has been environmentally induced.

Other data we have assembled show that the features of the environment that are responsible are

not those on which psychologists and sociologists have focussed most attention. Data collected by researchers working in many different countries support the conclusion that the norms are, at any point in time, similar across most cultures with a tradition of literacy. Since these cultures vary enormously in their state of economic development, their family sizes, their access to television, education, and computers and so on, this cross-cultural stability invalidates virtually all the explanations put forward by psychologists to explain the increase over time³.

But, in fact, recently published data from another area clearly show that much of the controversy Flynn provoked should never have occurred. Over the century for which normative data for the RPM exist, life expectancy has increased equally steadily and dramatically. For example, the life expectancy at birth of British males has increased from 45 to 75 years.

It is salutary to ponder the implications of these data. Flynn suggests that backward projection of the RPM data to the time of the Greeks indicate that the Greeks must have been incredibly stupid. He then uses this argument, along with others, to arrive at the conclusion that data on the differences between cultural groups have little meaning, asserting that “the tests cannot save themselves”. But does the life expectancy data not equally suggest that the Greeks must have had remarkably short lives? And does it follow that the differences between the life expectancy of blacks, whites, men, and women are meaningless? And do the data suggest that one should look for a single factor explanation of the increase – equivalent to changes in education, family size, and the availability of social puzzles? Above all, do the data suggest that whatever environmental variables are responsible for the increase over time are also responsible for the intra-generational and inter-group variance?

Implications

But, I hear you asking, what has this to do with developments in testing? Well, these studies, and the work summarised by Carroll (1993), *might* have had enormous implications for the conceptual framework that lies behind measures of cognitive ability. In practice, they do not. But the changes over time have created a serious ceiling effect on the *Standard Progressive Matrices (SPM)* among adolescents and young adults ... although better discrimination among the less able. It has therefore been necessary to develop a new test – the *Standard Progressive Matrices Plus* (Raven, J. C., Styles, I., & Raven, M. A., 1998) – to restore the

discriminative power at the upper end of ability that the SPM had when it was first developed.

Talent Diversity

We may turn now to our work in the other areas noted by Spearman that were mentioned earlier. As Spearman hinted, the main barrier to moving forward inheres in the measurement model espoused by psychologists (which was actually wished upon them by so-called educators who did not merit the name⁴). The difficulty posed by the model can perhaps be made clearest by asking “Where would biologists have got to if they had tried to describe all the variance between animals or plants in terms of 1, 2, or even 16 ‘variables’, then tried to describe the variance between the environments inhabited by those organisms in terms of nine variables, and then sought to assess the effects of the environments on the animals by intercorrelating the ‘animal’ and ‘environmental’ variables?”

The effects of psychologists’ failure to develop an appropriate measurement paradigm are serious and may be illustrated from a brief example. Psychologists are often asked to evaluate and compare the effectiveness of educational programmes. In particular, they have been asked to evaluate “progressive education”⁵. Many of those who have advocated or provided “progressive” education have not done so on the grounds that it improved reading, writing, and arithmetical skills but rather on the grounds that it created opportunities for pupils to develop characteristics like self-confidence, creativity, and initiative. More importantly, it presented an opportunity to nurture a diversity of talents and thus create cultures of enterprise or initiative, that is to say, to nurture emergent group properties.

Since there are no good measures of these qualities – and especially of the diversity of talents nurtured – and because the Joint Committee on Standards for the Evaluation of Educational Programmes and Policies (1981) insists that evaluators should only use tests of proven reliability and validity – evaluators have typically gone about their business using only tests purporting⁶ to measure reading, writing, and counting. It then emerges that “progressive education” does not enhance such scores. (Stallings and Kaskowitz [1974] even found that it *depressed* them whilst at the same time *increased* Progressive Matrices scores.) These findings have then been used to justify closure of such programmes.

Note that this has occurred despite two vitally important observations that can be made about some of the programmes. First, since they are the *only* educational programmes which draw out the talents of the pupils, they are the only ones that merit description as “education”. Second, these are the *only* educational programmes which set out to nurture the qualities that are required to transform our society in such a way that our species, and the planet as we know it, will have any chance of survival.

Implications

Because of these limitations, most of the evaluations can only be described as both scientifically incompetent and unethical⁷. Interestingly, however, further reflection reveals that the quality of an assessment, whether of an individual or of an educational or social process, is to be judged from its *comprehensiveness*, not from the accuracy of the assessment of one or two of its components.

The nature of the paradigm shift required to develop measures of the qualities mentioned can be illustrated using another example. Spearman drew attention to the specificity problem. Because they are difficult and demanding activities, people will only display the initiative, creativity, and self-confidence of which they are capable while carrying out activities that they are strongly motivated to carry out. The activities they are strongly motivated to carry out are specific and legion, varying from a predisposition to put drinks at ease, to raiding banks, to creating political turbulence. So, if one wants to make any meaningful assessment of someone’s ability to be creative, take initiative, or display a high degree of internal locus of control, one must first discover what it is that they are strongly motivated to do. Then one must set about identifying the components of competence they bring to bear to carry out such activities effectively. That is to say, as Spearman noted, the question is: “In the course of undertaking what kind of activity do they display their creativity, initiative, self-confidence, etc.?” not “*How* creative is this person?”

The process is analogous to a chemical analysis or the classification of a biological specimen and its effective implementation therefore depends on the development of a framework akin to Dalton’s

atomic theory or Linnaeus’s classification of animals and plants.

External Determinants of Behaviour

There is not space here to outline where we have got to in our quest to develop such a framework and set of analytic (or “assessment”) procedures. Interested readers will have to turn to our *Competence in the Learning Society* (Raven & Stephenson, 2001). Nor, indeed, is there space to even hint at the work we have done on the nature and assessment of developmental environments as they express themselves in homes, schools, and workplaces. But I cannot close without mentioning that, in the end, it has been this study of the ecological context of behaviour that has produced by far the most radical change in our thinking. This is because it has emerged that human behaviour is not mainly determined by personal talents, abilities, attitudes, and values but by an invisible network of external social forces which constrain the individual.

The point may again be made via an analogy. Prior to Newton, if things moved or changed direction it was because they were possessed of “animal spirits”. After Newton, it was because they were acted upon by a network of often invisible forces, documentation of which enabled one not only to predict the existence of previously unknown planets but also to harness those invisible forces to do useful work for mankind – such as by adding keels to sailing boats to harness the invisible counter-reaction of the sea to the wind to enable the sailing boats to head into the wind.

Implications

If we are to understand and predict behaviour – and thus gain control over our destinies – it is essential to map and harness an analogous network of social forces.

Our task as psychometricians is, therefore, to, on the one hand, build on the tools we have developed to implement a *descriptive* framework for thinking about the variance in human abilities and environments (and their interactions) and, on the other, to develop the tools required to map the network of social forces which overwhelmingly determine behaviour⁸.

Notes

- 1 See the 1998 edition of the General section of the Manual for Raven's Progressive Matrices and Vocabulary Scales (Raven, Raven, & Court, 1998).
- 2 Fifty percent of our grandparents would be assigned to Special Education classes in the U.S. if their scores were compared with today's norms.
- 3 See Neisser (1998) for a review of the status of the explanations put forward.
- 4 Actually, the roots of this problem are very deep indeed. To expose them, it is necessary to engage in the kind of socio-cybernetic analysis advocated later in this article. Interested readers should turn to Raven (2002).
- 5 See Raven (1994) for a review of these studies.
- 6 In Raven (1991) and elsewhere, I show that these tests actually lack construct validity.
- 7 See Raven (1997, 2000) for fuller discussion.
- 8 Whereas much of our work in connection with the former has been summarised in Raven & Stephenson (2001), our work on socio-cybernetics currently is only available in an unpublished paper by Raven & Navrotsky (2000).

References

- Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. New York: Cambridge University Press.
- 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.
- 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.
- Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects and materials. New York: McGraw Hill.
- Neisser, U. (Ed.) (1998). The rising curve. Washington, DC: American Psychological Association.
- Raven, J. (1991). The tragic illusion: Educational testing. New York: Trillium Press.
- Raven, J. (1994). Managing education for effective schooling: The most important problem is to come to terms with values. Unionville, NY: Trillium Press.
- Raven, J. (1997). Educational research, ethics and the BPS. Starter paper and peer reviews by P. Mortimore, J. Demetre, R. Stainthope, Y. Reynolds, & G. Lindsay. Education Section Review, *21*, 3–26.
- Raven, J. (2000). Ethical dilemmas. The Psychologist, *13*, 404–406.
- Raven, J. (7/7/2002). Intelligence, engineered invisibility, and the destruction of life on Earth. WebPsychEmpiricist: http://wpe.info/papers_table.html
- Raven, J., & Navrotsky, V. (2000). The development and use of maps of socio-cybernetic systems to improve educational and social policy, with particular reference to sustainability. Paper presented to a meeting of Research Committee 51, International Sociological Association, Panticosa, Spain.
- Raven, J., Raven, J. C., & Court, J. H. (1998-2001). Manual for Raven's Progressive Matrices and Vocabulary Scales. Sections 1-7 and Research Supplements 1 and 3. Oxford, England: Oxford Psychologists Press. Some of this material may be more easily retrieved from the website www.jcravenltd.com.
- Raven, J., & Stephenson, J. (Eds.). (2001). Competence in the learning society. New York: Peter Lang.
- Raven, J. C., Styles, I., & Raven, M. A. (1998). Raven's Progressive Matrices: SPM Plus Test Booklet. Oxford, England: Oxford Psychologists Press.
- Spearman, C. (1926). Some issues in the theory of g (Including the law of diminishing returns). Address to the British Association Section J Psychology, Southampton, England, 1925. Reprint of original bound in: London: Psychological Laboratory, University College: Collected Papers.
- Stallings, J., & Kaskowitz, D. (1974). Follow through classroom observation evaluation 1972 –1973. Menlo Park, CA: Stanford Research Institute (Report URU-7370).