The Outstanding Properties of the *Standard Progressive Matrices* ***Plus*** test

John Raven
30 Great King St., Edinburgh EH3 6QH
http://www.eyeonsociety.co.uk
jraven@ednet.co.uk

Version Date: 12 April 2000

The *Standard Progressive Matrices* ***Plus*** was introduced just over 20 years ago[i] but does not seem to have been adopted as widely as might have been expected.

This may be partly because the evidence of its outstanding psychometric properties has been buried in the chapters of a book that has not been widely purchased or cited[ii].

The purpose of this note is to bring this material together.

The test was developed to restore the ability which the original Standard Progressive Matrices had to discriminate between people of above average ability. This had been eroded by the inter-generational increase in scores widely known as the "Flynn effect".

In developing the new test particular care was taken to retain the unique features of the *Progressive Matrices* items[iii].

Nevertheless, it was a great surprise to find that the psychometric properties of the test which emerged were quite so outstanding.

The test has:

1. A set of items which almost exactly conform to the criteria for an ideal test meeting the specifications of *Item Response Theory*. Graphical representations of the results (see below) demonstrate that the abilities required to solve the problems form a continuous series, building on, and extending, those required to solve the easier items. No new abilities or metamorphoses in thinking are required. Thus the test is measuring "the same thing" at all levels.
2. The increase in difficulty from one item to another is almost equal throughout the scale. Combined with (1) this indicates that the scale has the properties of a foot-rule or meter stick. That is, it, perhaps uniquely in psychology, offers an *interval* scale. This is of fundamental importance if one wishes to, for example, assess the relative impact of some educational innovation on "more" vs "less" able people[iv].
3. The test works, and works in the same way, for people from different social and ethnic backgrounds[v].

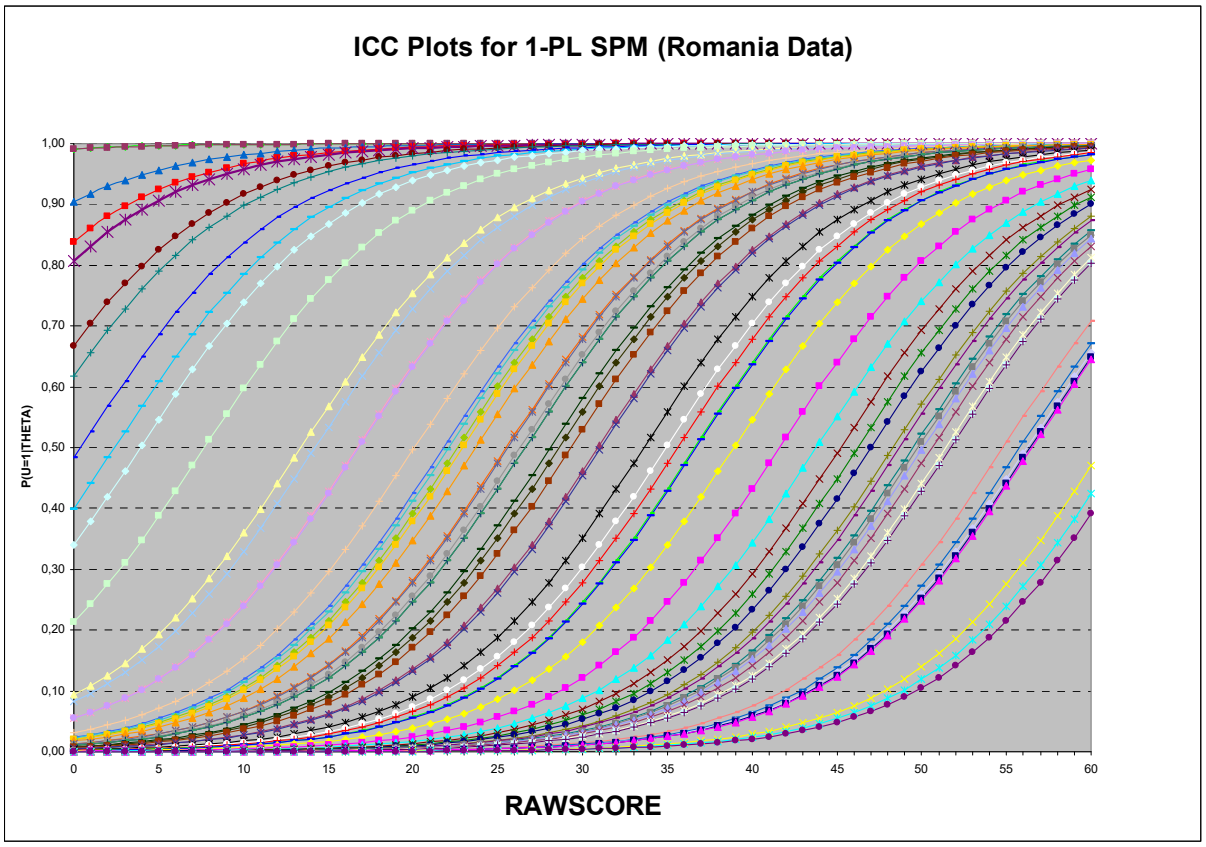
The evidence supporting these statements is contained in the Figures below[vi].

The graphs in the first two Figures show Item Characteristic Curves for all the items in the test.

The graph for each item plots the proportion of respondents with each total score who get the item right. Thus we see that while many low ability people fail to get the easier items right, 100% of more able people do so. On the other hand, while most low ability people fail to get the most difficult items right (those that do so do so as a result of randomly selecting the correct answer from one of the options available) more of the more able do so.

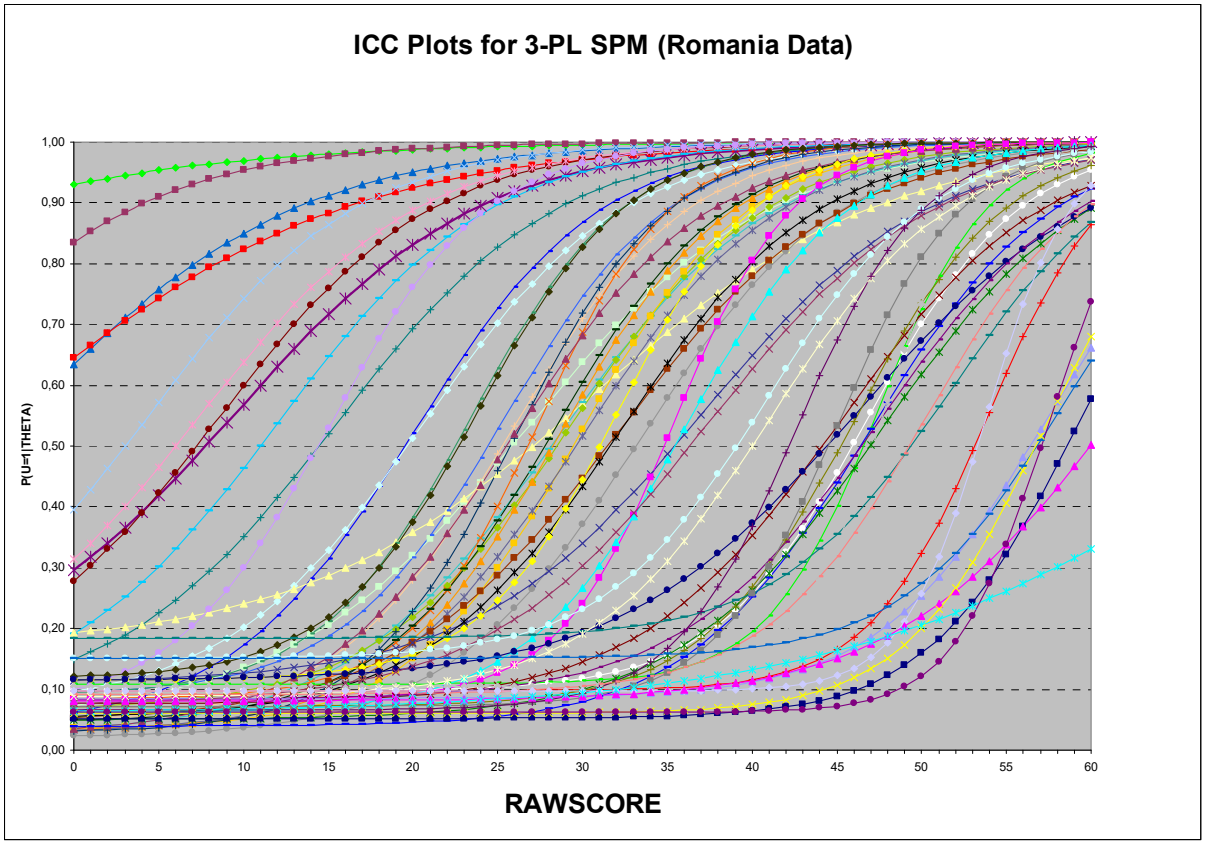
The program used to generate the graphs plotted in the first figure – known as a 1 parameter model – has smoothed the raw data rather heavily.

The 3-parameter plot shown in the next Figure creates a more realistic impression.



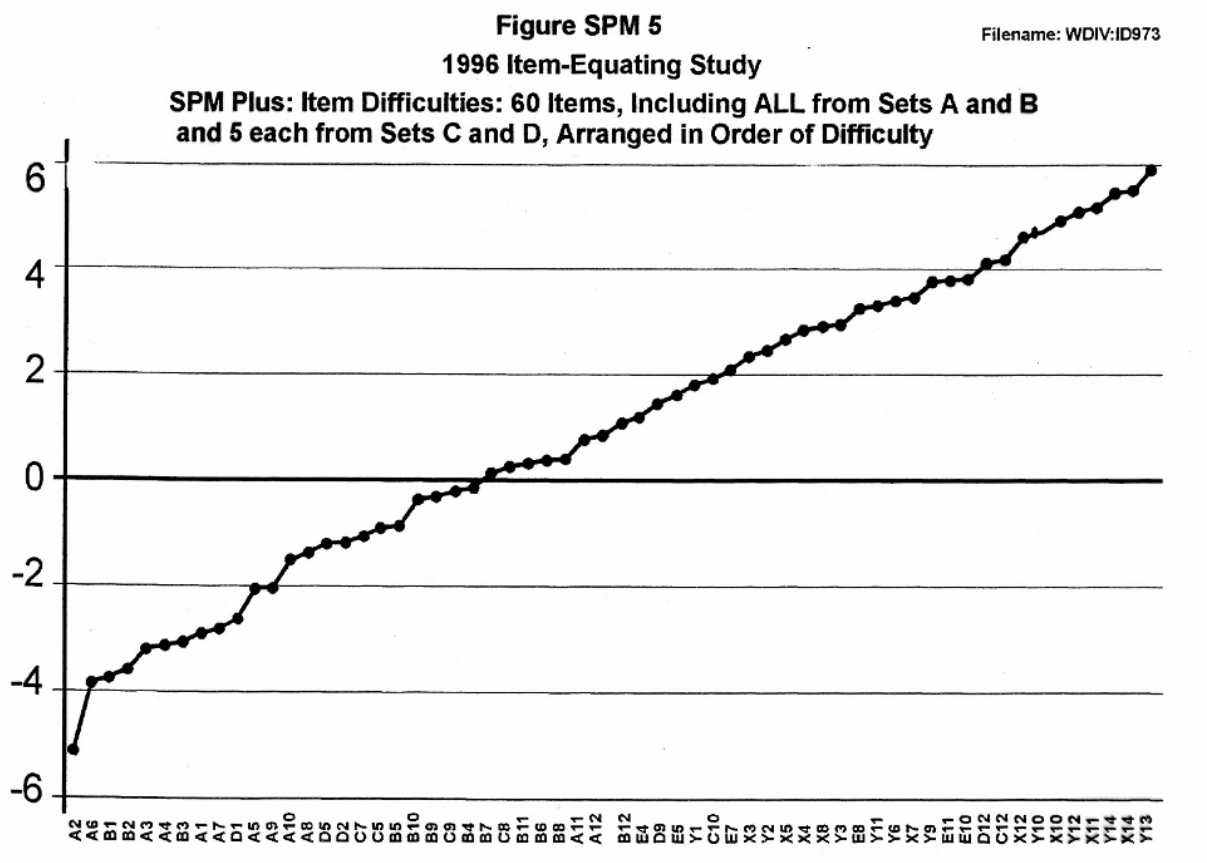
Although few of those who use the off-the-shelf statistical packages appear to understand it, the mathematical indices generated by these packages indicate how closely the set of items in a test conform to an ideal 1-parameter IRT model.

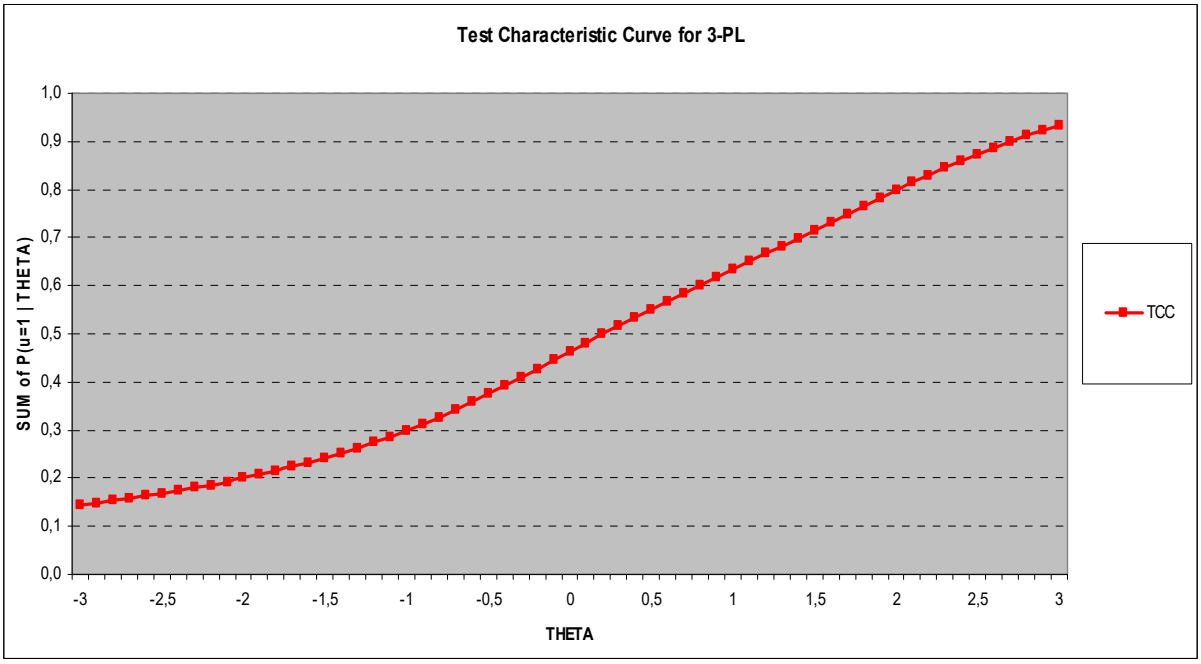
The graphs derived from a 3 parameter plot show more deviance from this ideal.



As mentioned above, what these Figures show is that, at least to a considerable extent, the abilities required to solve each more difficult item build on, and extend, those required to solve the easier items. No “new” abilities are required and there are no transformations or “metamorphoses” in the abilities required to solve the more difficult items.

What is more, as the next Figures show, the increases in item difficulty from one item to another is approximately equal at all points in the scale.



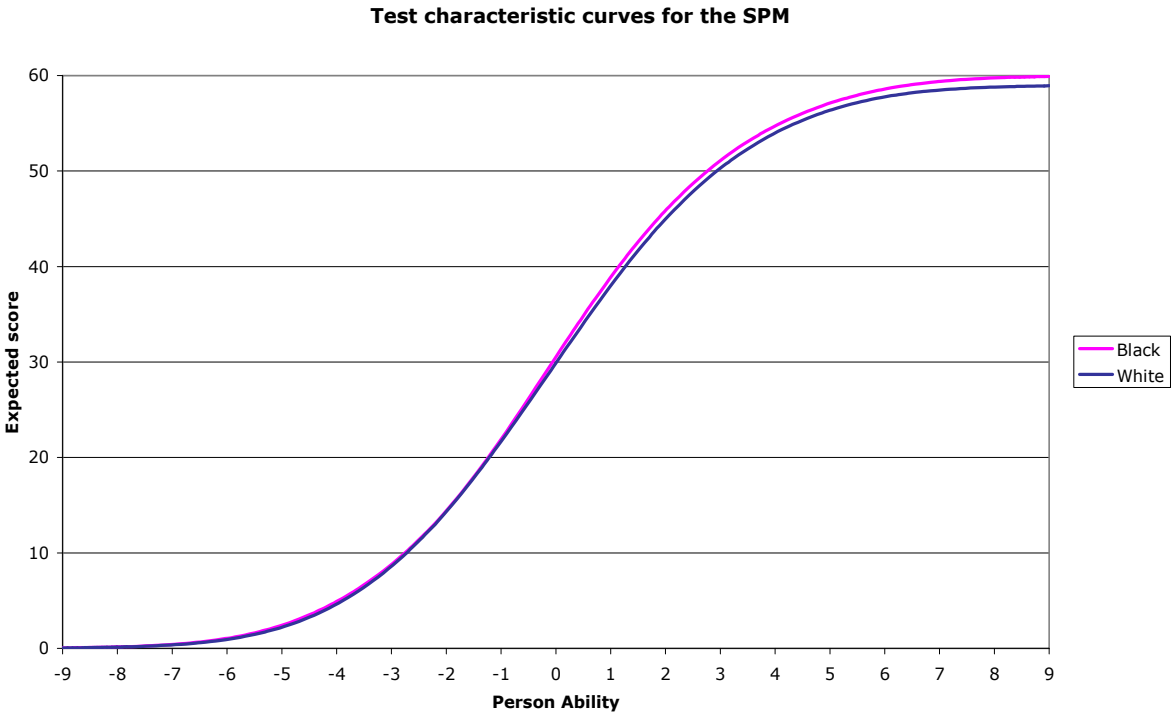


Taken together, what these plots show is that the test is not merely measuring "the same thing" at different points in the scale *but the **differences** between any two scores at any point in the scale* are to all intents and purposes ***equal.***

In technical terms, what we have here is an *interval* scale analogous to a foot-rule or meter stick.

This is of vital importance when, for example, trying to compare the differential effect of some (e,g, educational) intervention on those with high and low scores. As Prieler and Raven (2008) show, the use of tests lacking this property has resulted in endless misleading conclusions as, for example, in most of the studies concluding that low scorers benefit more than high scorers from educational intervention programmes.

Finally, as is illustrated in the graph below[vii] and demonstrated in numerous studies summarised in the test *Manuals*[viii], there is no Differential Item Functioning between e.g. black and white respondents. The same score means *the same thing* regardless of the group the respondent belongs to.



References

Prieler, J. & Raven, J. (2008). Problems in the measurement of change (with particular reference to individual change [gain] scores) and their potential solution using IRT. Chapter 7 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. Also available at: http://eyeonsociety.co.uk/resources/UAIChapter7.pdf

Raven, J. (2008). The need for, and development of, the SPM *Plus*. 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 3, pp. 103-112). Also available at http://eyeonsociety.co.uk/resources/UAIChapter3.pdf

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: 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 5, pp. 127-159). Also available at: http://www.wpe.info/papers_table.html and http://eyeonsociety.co.uk/resources/UAIChapter5.pdf

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, Including the Parallel and Plus Versions*. San Antonio, TX: Harcourt Assessment.

Raven, J., Rust, J., & Squire, A. (2008). *Manual: Standard Progressive Matrices - Plus Version - and Mill Hill Vocabulary Scale.* London, England: NCS Pearson, Inc.

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.

Taylor, N. (2008). Raven's *Standard and Advanced Progressive Matrices* among adults in South Africa. In J. Raven & J. Raven (Eds.) *opus cit,* Chapter 15. Also available at: http://eyeonsociety.co.uk/resources/UAIChapter15.pdf

---

[i] Raven, J.C. et al (1998), Raven, J. *et al* (2000, updated 2004), Raven, J. (2008)

[ii] Raven, J. (2008), Raven, Prieler, & Benesch (2008), Taylor, N. (2008)

[iii] Raven, J. et al (2000, updated 2004), Raven, J. (2008)

[iv] Prieler, J. and Raven, J. (2008)

[v] Raven, J. et al (2000, updated 2004), Taylor, (2008).

[vi] The first four diagrams come from Raven, J., Prieler, J. & Benesch, M. (2008), the fifth from Taylor, N. (2008).

[vii] Reproduced from Taylor, N. (2008).

[viii] Raven, Raven, and Court (2000, updated 2004).