

Comparison of the PfS and GMA Abstract Reasoning Tests

Introduction

This paper provides a comparison of the Profiling for Success (PfS) Abstract Reasoning tests and Graduate and Managerial Assessment (GMA) Abstract Reasoning test. It describes a study in which test takers completed PfS and GMA, with the purpose of demonstrating that the PfS tests assess the same constructs as GMA at a similar level of ability.

The PfS and GMA Abstract Reasoning tests are both based on Bongard problems. Originally developed as a method of evaluating computer vision and problem-solving programs, Bongard problems present test takers with two sets of shapes. Each set of shapes in a set, typically labelled 'Set A' and 'Set B', have one or more features in common, though the two sets are not related to each other. Test takers have to identify the common features for each set and then determine whether further 'test shapes' share the common features of Set A, Set B or do not share the features of either set.

Abstract reasoning tests based on Bongard problems have a number of distinct features from other abstract or non-verbal reasoning tests. First, their origins as challenging problems for computer vision programs mean that their format differs from many abstract tests (e.g. series completion tests and matrices) which emphasise convergent thinking that can be more readily replicated by computers. Instead, Bongard problems require both divergent and convergent thinking, so assessing a style of thinking that more closely mimics human problem-solving. Second, not all of the information in each of the Bongard problems is relevant to its solution. As with 'real life' problems and challenges, this means that part of the task is to identify what information is part of the solution and what is 'noise' that can be ignored. Finally, the nature of test items developed from Bongard problems provides an efficient format for ability testing, as respondents can be asked to classify a number of items for each set of shapes. Each item is treated as a separate test question and scored accordingly, though GMA also has a 'harsh' scoring method that awards one mark only if all five test shapes in a group have been answered correctly.

This paper presents a comparison of the PfS Abstract Reasoning tests Levels 3 and 4 with the GMA Abstract Reasoning Test Form A. Each of these is a high-level assessment of abstract reasoning ability. PfS Level 3 is aimed at approximately to top 40% of the population in ability terms, whilst both PfS Level 4 and GMA cover the top 10 to 15% of ability.

Study design

Data was collected during the first quarter of 2007 from two groups of Year 12 students, one at a boys-only comprehensive school and another at a girls-only independent school. There were 78 participants from the boys school with a mean age of 16.7 years (SD=0.7) and 48 from the girls school with a mean age of 16.4 (SD=0.5).

One test session was conducted at each school, during which the respondents completed three tests: PfS Abstract Levels 3 and 4, and GMA Form A. To allow for order effects, each group was divided into two and the order in which the tests were administered was counterbalanced.

Comparison of the PfS and GMA Abstract Reasoning Tests

Sample characteristics

The mean raw scores and standard deviations (in brackets) for the PfS and GMA Abstract tests is shown below. The pooled data shows the scores from all candidates with the following four rows showing the scores from each of the four groups that were tested.

	PfS Abstract 3	PfS Abstract 4	GMA Abstract – lenient scoring	GMA Abstract – harsh scoring
Pooled data (n=116)	38.9 (10.2)	32.1 (11.2)	72.1 (15.6)	7.1 (4.0)
Group 1 – PfS 4, PfS 3, GMA (n=31)	32.8 (9.4)	24.9 (9.7)	62.4 (11.5)	4.1 (3.2)
Group 2 – GMA, PfS 4, PfS 3 (n=37)	38.5 (10.7)	29.4 (10.7)	72.0 (15.7)	6.6 (3.7)
Group 3 – PfS 3, PfS 4, GMA (n=23)	40.1 (8.6)	39.5 (8.4)	77.0 (18.6)	10.0 (2.9)
Group 4 – GMA, PfS 3, PfS 4 (n=25)	45.8 (7.0)	40.4 (6.5)	82.4 (7.9)	9.7 (2.4)

The administration of the three tests was counterbalanced as far as possible to control for practice or fatigue effects. Differences between the four groups in terms of mean raw test scores were explored using post-hoc tests, and showed the following:

- Group 1 had significantly lower scores than Groups 3 and 4 on all tests, and were also significantly lower than Group 2 on GMA lenient and harsh scores,
- Group 2 had significantly lower scores than Groups 3 and 4 on PfS 4 and GMA harsh scores, and were significantly lower than Group 4 on PfS 3 and GMA lenient scores,
- Group 3 was not significantly different from Group 4 on any of the test scores.

Though it is difficult to conclude conclusively from the design what effect on performance test order had, comparison of the four groups indicates that the observed differences can be at least partially accounted for by variations between the schools and the groups themselves. For example, Group 1 consistently performed lower than Group 2 though they were from the same school, as did Group 3 when compared to Group 4. Further comparisons of the scores from individual tests also suggest against a strong order effect. For example PfS 3 scores are higher in Group 1 than 2 and in Group 3 than 4 suggesting better scores when the test is taken later (practice effect), but the GMA scores from these groups follow the opposite pattern. Given the similarity of the PfS and GMA tests such opposite effects seem unlikely, further supporting the view that the observed differences are due to the composition of the groups.

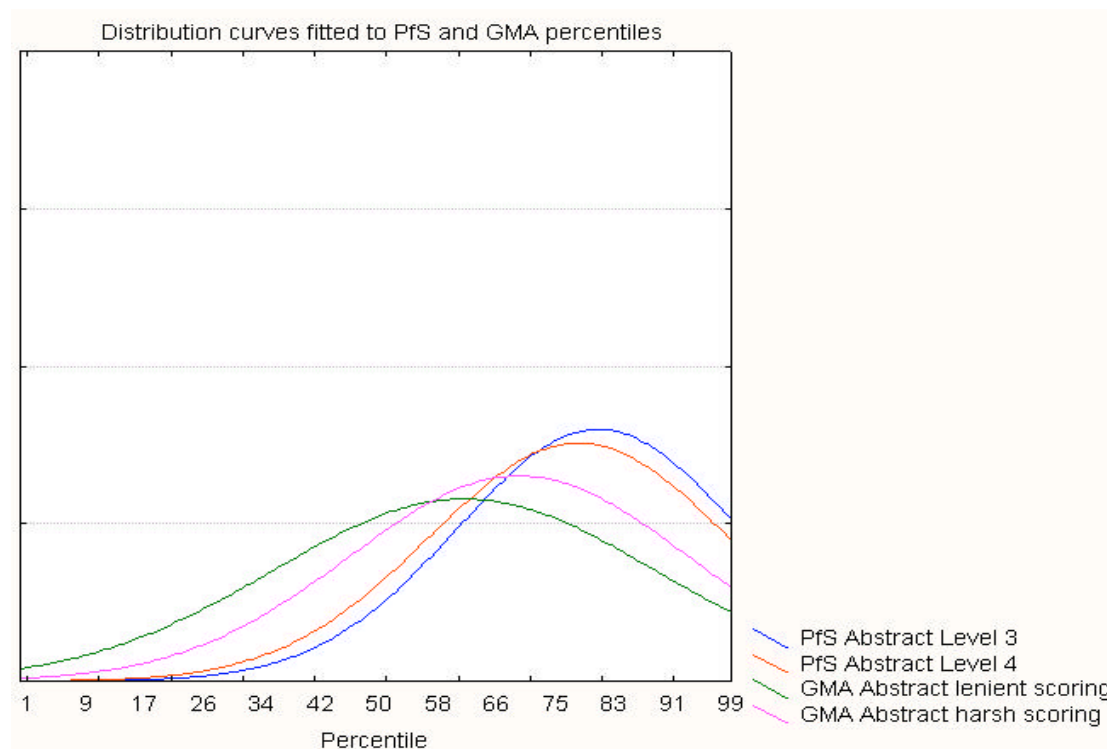
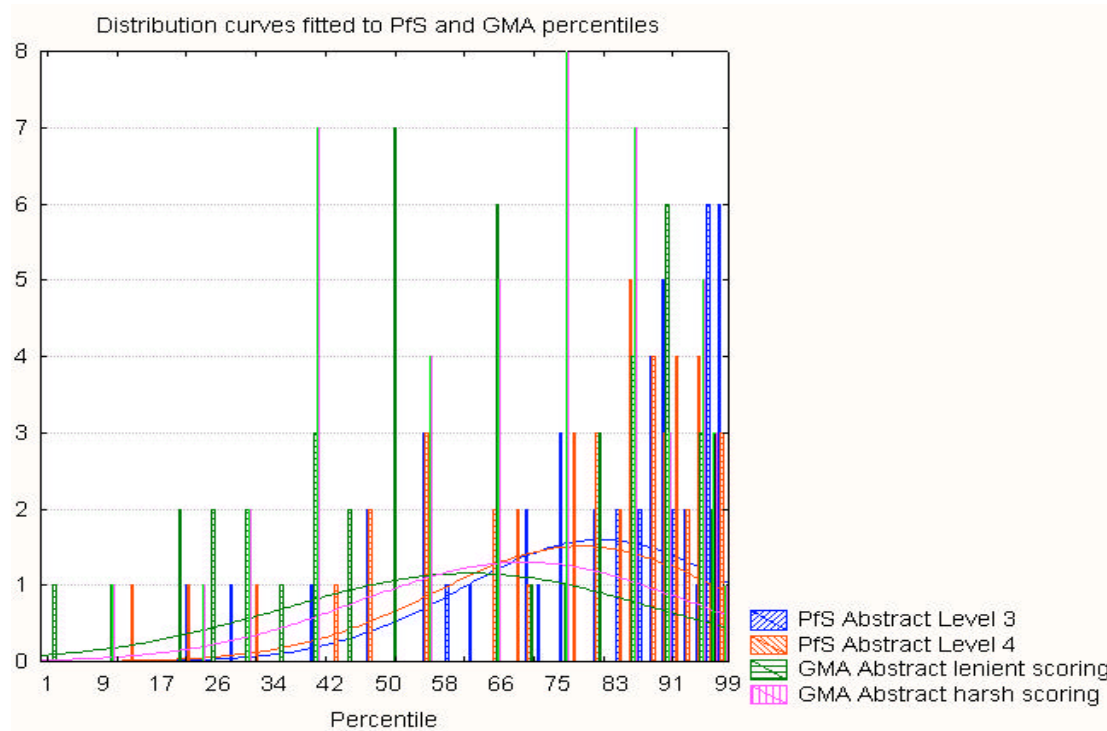
On the basis of this data, the four groups were combined and treated as a single sample for the following analyses.

It is also notable from that mean scores on PfS Abstract Level 3 are higher than those on Level 4. Given that Level 3 is targeted at undergraduates and Level 4 at postgraduates and experienced professionals these differences support the more difficult nature of the Level 4 Abstract test, though this difference is marginal in the case of Group 3.

Comparison of the PfS and GMA Abstract Reasoning Tests

Comparison of difficulty levels

A comparison of the difficulty levels of the PfS and GMA tests can be made by looking at percentile equivalents and facility levels. The percentile frequency distributions and curves fitted to these distributions are shown below, with the second chart showing the fitted distribution curves in detail without the frequency bars for clarity.



Comparison of the PfS and GMA Abstract Reasoning Tests

An examination of the fitted distribution curves gives the clearest comparison of the PfS and GMA percentiles, as it includes a 'smoothing' of the peaks that occur due to the modest sample size. It is notable that, despite the GMA being widely considered as being a very challenging test, the majority of respondents fell above the mid-point of the 50th percentile. Correspondingly, the distributions show evidence of a ceiling effect, though given that all the tests have further 'stretch' at the top end (e.g. the GMA norm group gives a percentile of 99 for scores of 100 or more out of 115) this issue could be rectified by the development of a higher-performing norm group.

The four percentile distributions are broadly concurrent, but they do show that the GMA with lenient scoring produces a broader distribution of scores with a lower mean percentile than the PfS Level 3 or 4 Abstract Reasoning tests. It needs to be recognised however, that this comparison of percentiles is not comparing 'like-with-like' as the percentiles for the three tests were each derived from a separate norm group separated in time by almost 20 years.

A second way of comparing the PfS and GMA tests is by deriving an overall facility estimate for each test. Facility levels can be estimated by dividing the mean raw score by the number of test items, to give a value from 0 to 1 for each test. Values closer to 1 indicate an easier test and those closer to 0 a harder test. The values for PfS 3 and 4 were 0.65 and 0.54 respectively, as would be expected from the mean scores of these tests noted above. The mean GMA raw scores for lenient and harsh scores were 0.63 and 0.31 respectively.

These figures indicate that the difficulty level of GMA is comparable to that of the PfS tests; GMA being more difficult than PfS Level 3 but not quite as difficult as PfS Level 4 when using lenient scoring. Using harsh scoring has a noticeable effect on these comparisons, resulting in a value that shows it to be much harder than the PfS Level 3 or 4 tests. Given the comparability of GMA to PfS when using lenient scoring, however, this change in difficulty is clearly due to the nature of the scoring rather than the test *per se*.

Association between PfS and GMA Abstract tests

The difficulty levels discussed above show that the PfS and GMA tests are of comparable difficulty, but alone do not indicate whether they are found to be of comparable difficulty by the same test takers. The correlational analyses here indicate the extent to which respondents performed comparably on each test. The correlations between the scores from the three tests are shown below (scatter plots can also be seen in the appendix), with the first figure showing the raw correlation and the second in brackets the correlation corrected for the reliability of the two tests in question.

	PfS 3	PfS 4	GMA lenient
PfS 4	0.73 (0.80)		
GMA lenient	0.71 (0.78)	0.64 (0.71)	
GMA harsh	0.68 (0.86)	0.69 (0.79)	0.80 (0.89)

All uncorrected correlations between PfS tests and GMA are 0.64 or greater, and when correct for reliability are 0.71 or greater. These figures indicate a good degree of association between the assessments and all exceed the 0.70 threshold, typically recognised the point at which tests can be considered to be alternative forms of each other.

Comparison of the PfS and GMA Abstract Reasoning Tests

Conclusions

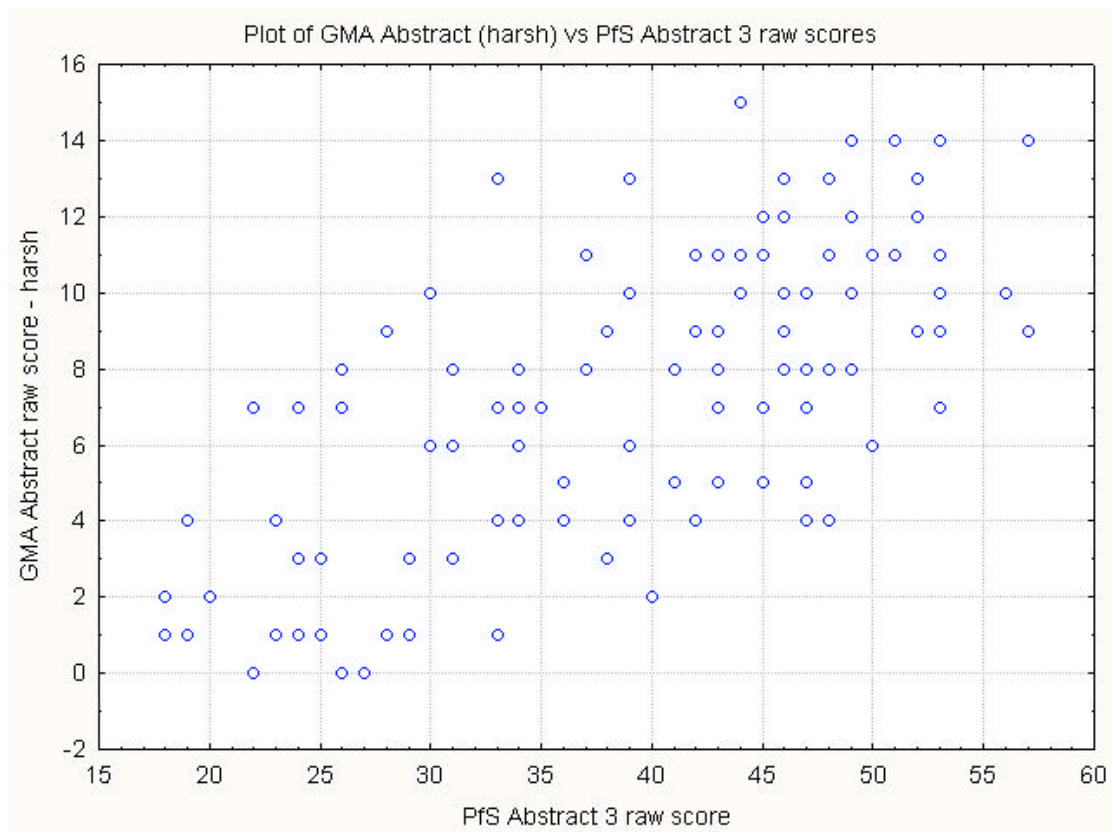
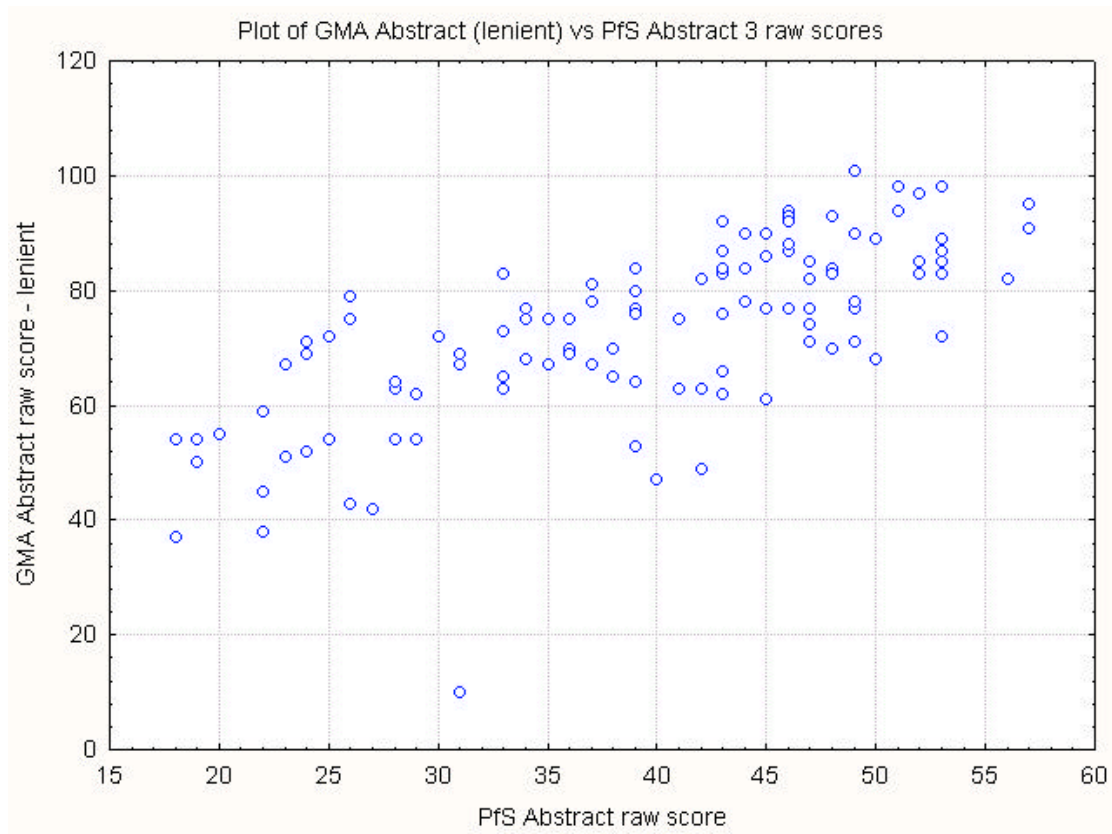
The current study has demonstrated the comparability of the PfS and GMA Abstract Reasoning tests. Comparability has been demonstrated both in terms of the overall difficulty level of the tests and their ability to place respondents in corresponding order of ability. Indeed, the close association between PfS Levels 3 and 4 and GMA mean that they can be treated as alternate versions of the same test.

The percentile distributions produced from the tests were broadly similar, though these did show the PfS tests to be slightly easier than GMA. This is most likely to have arisen, however, from the differences in the norm groups that the percentiles were derived from. It is likely that PfS and GMA could be further aligned by re-norming.

To conclude, the PfS Abstract Reasoning tests provide an up-to-date assessment of the same construct as the GMA Abstract Reasoning test. The PfS tests are efficient in terms of time (12 minutes compared to 30) and can be delivered as paper-based assessments or online. The tests provide an effective and fair way of identifying a type of ability known to be a significant predictor of learning potential and job success and, by utilising the bank of pre-calibrated items on which the tests used here are based, offer organisations the flexibility of tailoring PfS Abstract Reasoning tests to meet their specific needs.

Comparison of the PfS and GMA Abstract Reasoning Tests

Appendix: Scatter plays of the PfS Level 3 and 4 against GMA



Comparison of the PfS and GMA Abstract Reasoning Tests

