

EXECUTIVE ASSESSMENT: MEASURING HIGH LEVEL PERFORMANCE

Article dated September 2009

Author(s): Angus McDonald

Introduction

There are a number of tests of high level reasoning ability. Among the most popular are the Graduate & Managerial Assessment (GMA) tests published by ASE. These are aimed at academically successful job candidates and are designed to provide an additional measure of high-level ability and suitability for senior management roles. In particular the GMA Abstract Reasoning Test has provided a good predictor of strategic thinking, and the ability to operate at different levels of complexity, for over 20 years.

The new Profiling for Success (PFS) Abstract Reasoning Test, which is based on the same principles as the GMA(A), offers an equivalent measure of thinking skill, but is shorter, available at different levels of difficulty, and can be closely tailored to the needs of the user. This briefing paper explores the overlap between the PFS(A) and the GMA(A), and makes the point that the new test is an ideal alternative to the GMA.

Objective

The objective was to research the relationship between the two tests in order to determine whether the PFS(A) and GMA(A) measure the same constructs, and can effectively be considered as alternate forms of the 'same' test. In particular whether they are comparable in terms of overall difficulty, and so are able to place test takers in their corresponding order of ability.

Approach

Both tests provide measures of flexibility of thinking and the capacity to perceive new patterns, devise new methods and operate at different levels of analysis. The GMA(A) is available in two forms, of equal difficulty; and is targeted at the top 10-15% of the population. The PFS(A) can be constructed flexibly from an item bank, and also has two fixed forms – this time aimed at different levels of ability: Level-4 has the same target as GMA(A), but Level-3 covers the top 40% of the population.

To compare the tests, all three were administered during single sessions at two schools. Order effects were controlled by dividing each group into two, with the order of presentation of each test counterbalanced.

Sample

Data was gathered during the first quarter of 2007 from two groups of Year-12 students: a boys-only comprehensive and a girls-only independent school. There were 78 male participants with a mean age of 16.7 years (SD=0.7), and 48 female participants with a mean age of 16.4 years (SD=0.5), yielding 116 usable sets of data.

There was a selective component to the sample, in particular from the girls-only school, however it still provides a better cross-section of the 'real' ability range in the population than using graduates, or managers themselves.

Results

Various detailed analyses were performed, including a comparison of difficulty levels for the PFS(A) tests and the GMA(A), the latter having two scoring protocols: harsh and lenient.

The overall facility levels* for the PFS tests were 0.65 (Level-3) and 0.54 (Level-4) respectively. The corresponding figures for the GMA were 0.63 (lenient scoring) and 0.31 (harsh scoring).

This implies that the GMA(A) is more difficult than the PFS(A) Level-3, but not quite as difficult as the PFS(A) Level-4, when using lenient scoring. The harsh scoring protocol suggests that the GMA(A) is

much harder than the PfS(A) Levels 3 & 4.

However given the comparability of the GMA(A) and PfS(A) when using lenient scoring it is clear that the change in 'difficulty' is an artefact of the lenient/harsh scoring, rather than anything to do with the test itself. So in short, the tests are of comparable difficulty.

Additional statistics reveal that they are also found to be of comparable difficulty by the same test takers. The results below shows pairs of correlations (the figures in brackets being corrected for the reliability of the two tests):

PfS 4 / PfS 3 = 0.73 (0.80)

GMA lenient / PfS 3 = 0.71 (0.78)

GMA lenient / PfS 4 = 0.64 (0.71)

GMA harsh / PfS 3 = 0.68 (0.86)

GMA harsh / PfS 4 = 0.69 (0.79)

GMA harsh / GMA lenient = 0.80 (0.89)

All the corrected figures are in excess of 0.70, a threshold that is usually recognised as the point at which the tests can be considered as alternate forms of each other.

Conclusions

The PfS(A) tests provide an up-to-date assessment of the same constructs as the GMA(A). They offer a more time efficient measure of high level performance, taking 12 minutes to complete, rather than 30 minutes, and provide a powerful alternative to the GMA(A) test.

The fact that the PfS(A) uses a bank of pre-calibrated items also offers organisations the flexibility to tailor tests to their needs, and to have increased confidence in high-stakes assessment situations.