

# Reasoning Tests

# User's Guide

abstract

numerical

verbal

# Profiling for Success: Reasoning Tests

## User's Guide v1.4

**Roy Childs, John Gosling**  
**(and Angus McDonald who produced the original User Guide on which this is based)**

### Contents

Introduction	7
Overview of the PfS Reasoning Tests	7
Section One: Using Reasoning Tests for selection and development	11
Why use Reasoning Tests?	11
Section Two: Selecting and administering the Reasoning Tests	21
Introduction	21
Selecting appropriate tests	21
Using the PfS Reasoning Tests as locator tests	24
Administering paper-based tests	26
Overview of administration	26
Planning the test session	27
Materials	28
The test session	29
Administering computer-based tests	31
Supervised assessment	31
Unsupervised assessment	32
Technical requirements for computer-based tests	35
Section Three: Scoring and review of test results	37
Overview of scoring and test scores	37
Qualitative analysis of results	39
Scoring paper-based tests	40
Scoring computer-based tests	41
Using the online report generator with paper-based tests	42
Review of test results	42
Communicating test results	42
Conducting a review session	44
Section Four: Development of the Reasoning Tests	47
Test formats	47
Verbal reasoning format	48
Numerical reasoning format	48
Abstract reasoning format	49
Item writing	50

Pre-trialling item reviews	50
Trialling	51
Item analysis	51
<b>Section Five: Technical information</b>	<b>55</b>
Introduction	55
Reliability	55
The concept of reliability	55
Reliability statistics	56
Standard error of difference	61
Bias	64
A commentary on interpreting bias data	73
Validity	74
Face validity	75
Content validity	75
Construct validity	76
Criterion validity	81
<b>Section Six: Making comparisons – using norms and the IRT scale</b>	<b>85</b>
Introduction	85
Developing a common scale – the IRT scale	86
How to use the IRT scales and the norms	89
Using the norms and IRT Conversion Table – an illustration	89
<b>References</b>	<b>91</b>
<b>Appendix One: Explanations of practice questions</b>	<b>93</b>
<b>Appendix Two: Sample test reports</b>	<b>99</b>
<b>Appendix Three: Comparison tables – IRT Scores to Raw Scores</b>	<b>111</b>
<b>Appendix Four: Core Norm tables</b>	<b>125</b>
Introduction to the norm tables	125
Core norms for closed tests	127
<b>Appendix Five: Norm table descriptions for Open tests</b>	<b>141</b>
<b>Appendix Six: Norm table supplements for the Closed tests</b>	<b>149</b>
<b>Appendix Seven: Details of the IRT process for creating a common scale across all tests</b>	<b>175</b>

## List of Tables

<b>Table 1:</b> Correspondence between the PfS Reasoning Tests and level of ability as indicated by the level of educational attainment .....	23
<b>Table 2:</b> Summary descriptions for combinations of speed of working and accuracy .....	39
<b>Table 3:</b> Score bands for communicating results and their relationship to T-scores and percentiles.....	43
<b>Table 4:</b> Timings and number of items in each of the PfS Reasoning Tests .....	53
<b>Table 5:</b> Mean, SD, sample size, number of items, internal consistency and SEM for the PfS Reasoning Tests .....	59
<b>Table 6:</b> Mean and SD for first time and retest candidates, and test-re-test reliabilities for bespoke versions of the PfS Reasoning Tests .....	59
<b>Table 7:</b> Best estimates (underestimates) for the Alternate Form (AF) reliabilities with means SD's and sample sizes .....	61
<b>Table 8:</b> SED for the PfS Reasoning Tests at 68%, 80% and 95% confidence levels .....	63
<b>Table 9:</b> Mean raw scores and standard deviations for males and females on the PfS Reasoning Tests.....	67
<b>Table 10:</b> Mean raw scores and standard deviations for 'whites' and 'non-whites' on the PfS Reasoning Tests .....	68
<b>Table 11:</b> Mean test scores and effect sizes for different ethnic groups based on the open Level 2 PfS Reasoning Tests .....	69
<b>Table 12:</b> Cross tabulation between appointment decision and aggregated ethnic group categories .....	70
<b>Table 13:</b> Raw score means on the three reasoning tests for each ethnic group .....	71
<b>Table 14:</b> Correlations between PfS Reasoning Tests (raw scores) and respondents' age .....	72
<b>Table 15:</b> Intercorrelations of the PfS Reasoning Tests .....	77
<b>Table 16:</b> Associations between PfS Reasoning Tests and the GMAT .....	78
<b>Table 17:</b> Associations between PfS Abstract Tests and GMA Abstract form A .....	79
<b>Table 18:</b> Intercorrelations between the Verbal, Numerical and Abstract Reasoning Tests and existing reasoning tests .....	80
<b>Table 19:</b> Correlations between Numerical, Verbal and Abstract Reasoning Tests and Memory and Attention Test.....	81
<b>Table 20:</b> Associations between GCSE English, maths and science grades and PfS Reasoning Tests.....	82
<b>Table 21:</b> Association between UCAS points, degree class and PfS Reasoning Tests .....	82
<b>Table 22:</b> Summary details of the PfS linking samples .....	85

**List of Figures**

Figure 1:The predictive validity and popularity of different assessment methods. 14  
Figure 2:The normal distribution curve, Z-score, T-score and percentile scales ..... 38  
Figure 3:Verbal - IRT scaled score to published percentile..... 87  
Figure 4:Numerical - IRT scaled score to published percentile..... 87  
Figure 5: Abstract - IRT scaled score to published percentile ..... 88

## **Acknowledgements**

The authors would like to thank the following:

Mark Parkinson for his intelligent contributions to and critiques of this User Guide

Barry Sexton for his invaluable statistical advice and analysis

Linda Paxton, Elvira Botta and Janet Newman for editing, proof -reading and commenting



## Introduction

Profiling for Success (PFS) is the name for a comprehensive range of psychometric tests and questionnaires. These cover important areas in which there are significant individual differences and these are grouped into areas/themes as follows:

1. Personality and Style
2. Motivation and Energy
3. Relationships and Connection
4. Capability and Thinking

This User Guide covers the PFS Reasoning Tests which are from the Capability and Thinking area. These tests measure Verbal (V) and Numerical (N) reasoning – and a form of Abstract (A) reasoning (N.B. Abstract reasoning tests cover a variety of different skills – to be discussed later). As such it covers tests that have been well established as measures of cognitive ability – and which have been shown to predict success in a wide range of situations and jobs. However, in choosing and using the PFS Reasoning tests the reader is encouraged to consider the fuller range of PFS Capability and Thinking tests. This is because the traditional V, N and A tests cannot adequately capture the complexity of human cognitive functioning. Of course they simulate an important ability (i.e. to think and process information quickly and accurately) but they do this in a way that does not capture the way in which people think and make decisions in the real world. Rarely do people have all the information they require to make a decision in one place at one time. Rarely is that information clear and unambiguous – modern life involves different skills which are increasingly important in the real world of work. For example, the ability to deal with multiple tasks and to handle varying levels of complexity and ambiguity. These skills are covered by different tests from PFS called the Memory and Attention Test (MAT) and the Decision Analysis Test (DAT), which are covered in different User Guides.

## Overview of the PFS Reasoning Tests

The PFS Reasoning Tests were developed to provide a reliable but easy to administer battery of cognitive skills. They offer a flexible approach to the assessment of reasoning abilities for selection and development purposes. The tests cover three areas of reasoning abilities:

- Verbal** – The ability to understand written information and determine what follows logically from the information.
- Numerical** – The ability to use numerical information to solve problems.



**Abstract** – The ability to identify patterns in abstract shapes and generate and test hypotheses.

As the benefits of psychometric assessments are increasingly recognised and test usage grows, new ways of assessing abilities are needed. The PfS Reasoning Tests meet these needs by offering both paper- and computer-based assessments that can be used with a wide range of ability groups.

The key features and benefits of the PfS Reasoning Tests are:

- **Flexible delivery options** – the paper- and computer-based (online) tests allow for traditional individual or group administration, or remote assessment. Through remote assessment it is possible to include test data earlier on in the assessment process. For example, including test information in the first sift alongside application forms or CVs gives more information on which to base decisions, so potentially enhancing the accuracy of decisions and increasing the efficiency of the selection process.
- **'Open' and 'closed' test versions** – closed versions of each of the tests are available for use under supervised conditions where the identity of test takers can be closely monitored. Open access versions are also available for use in situations where remote, unsupervised administration is appropriate. These different versions have been developed to meet the increasing need to test candidates remotely, largely as a result of the growth in internet assessment and the demand for the use of tests for guidance and other development purposes, as well as the more established approach of supervised assessment.
- **A choice of tests for different levels of ability** – the PfS range covers a very wide range of ability from the General Population to those working at high conceptual levels in academia and research. This range has been established using Item Response Theory (IRT) methodology so that, with each choice of test level, there is a clear step level in difficulty. This encourages the use of an appropriate level of difficulty for different purposes, and also provides a way of equating test scores across different levels of test.
- **Common formats across a wide ability range** – whilst the Verbal, Numerical and Abstract Reasoning Tests span a wide range of ability levels, from school leavers to experienced managers, they all use common test formats. If necessary, the appropriate test level can be identified by administering a test as a 'locator' among a group of current employees. This process is readily achieved through the use of online tests and guidance is given on how to do this in the User's Guide.

- **Detailed reports and analysis** – separate computer-generated reports are available for test users and test takers. For test takers these reports give broad band results so that there is less likelihood of small differences in score being misinterpreted. These reports also contain an analysis of speed and accuracy which is linked to a narrative on test-taking strategy and suggesting areas for consideration and development. Test user/administrator's reports present full test data (raw scores and standardised scores) as well as the analysis of speed and accuracy linked to interview prompts. Summary versions of reports for test takers and test administrators are also available.

This User's Guide provides test users/administrators with the information they need to understand, use and interpret the Verbal, Numerical and Abstract Reasoning Tests which make up the PfS Reasoning Tests. **Section One** summarises research on the importance of reasoning abilities for successful job performance and training, and describes the rationale behind the PfS Reasoning Tests. Administration of paper- and computer-based versions of the tests is covered in **Section Two**. **Section Three** deals with scoring and feedback. The development of the PfS Reasoning Tests is described in **Section Four**, and **Section Five** provides technical information on the tests and their functioning. It is recommended that all users should read at least **Sections Two** and **Three** before using any of the tests.

In addition to the information contained in this User's Guide, the test publishers offer consultancy, training and general support in using and interpreting the results from these Reasoning Tests and other assessments. In the interests of promoting best practice you are welcome to contact Team Focus Ltd on + 44 (0)1628 637338, e-mail [teamfocus@teamfocus.co.uk](mailto:teamfocus@teamfocus.co.uk) for further enquiries and support.



## Section One: Using Reasoning Tests for selection and development

### Why use Reasoning Tests?

The use of reasoning tests for selection and development is well-established in many organisations. Surveys show that usage continues to increase (e.g. CIPD, 2006; Jenkins, 2001), with new organisations discovering the benefits that properly applied psychometrics can bring and established users expanding their use of psychometrics. The use of online tests as part of the selection process has also grown rapidly in recent years, with figures showing a rise from 6% in 2002 to 25% in 2006 (CIPD, 2004; 2006). The figure was more or less the same in 2013, however if employers are asked about all selection focused ability testing, the percentage rises to 45% (CIPD, 2013).

When used sensitively, with due regard for both their strengths and limitations, there are many good reasons for using psychometric tests. The most compelling reason for using well constructed psychometric reasoning is that they provide information on a person's potential or development needs – information that is hard to obtain by other methods. All benefits of psychometric assessments ultimately come from the fact that they get at something that is hard to observe and which adds to the accuracy of prediction, thus helping the decision-making or development process. Well-informed decisions, in turn, help organisations to grow and develop. It is now well-established that tests of general mental ability, of which reasoning is a core feature, should normally feature as a significant factor in the decision-making process. Some of the evidence suggests that they can be the best single predictor of job performance and success on work-related training courses (Schmidt and Hunter, 1998).

To contribute to the decision-making process, psychometric tests have to discriminate among the people who take them. Here, discrimination is the ability to identify real differences between test takers' potential, not the pejorative sense of discrimination where one group is favoured over another for reasons unrelated to true potential.

Changes in the education system, particularly the increasing number of students in further and higher education, have made psychometric tests valuable decision-making tools for employers for three reasons:

- The growth in the number of courses and qualifications makes it difficult to evaluate applicants with very different qualifications.
- The increasing number of students obtaining top grades means that academic qualifications have lost much of their ability to discriminate between people.

- Standards of education vary considerably between institutions and courses. Psychometric tests overcome these variations by providing a 'level playing-field' for people to demonstrate their current ability and potential.

This last point touches on the increasingly important issue of fairness in selection. A very significant reason for using psychometrics is that they can, when used well, make the assessment of applicants fairer. Of course, the qualification to this statement is 'when used well' which requires that the abilities assessed by the test are related to the skills required for the job performance (see page 15) and that the administration is clear, standardised and suitable for all test takers (see Section Two). Helping test takers to prepare for the testing session, for example by sending out the Test Taker's Guide (see page 21) or giving access to other approved practice materials, also helps to give everyone an understanding of what is involved and increases the chance of them demonstrating their abilities – by analogy giving everyone some 'training and route experience' before they begin the 'race'.

Psychometric tests further contribute to effective selection and development decisions by explicitly recognising the potential for error in test scores. All assessments (e.g. educational qualifications, ratings from assessment centres or interviews) are subject to error, but this error is rarely acknowledged (see pages 50 and 51 for further discussion of test error). Recognising that test scores contain a degree of error and making this explicit, allows the band of error to be taken into account when making decisions based on test scores.

The relationship between test scores and subsequent job performance or success on training courses has been touched on above. To be defensible as a selection method, links between test scores and subsequent job or training performance have to be established. When this link is established, a test or other selection method is said to have 'validity' or to be 'fit for the purpose'. Showing a test has validity is also important as it is the basis for showing a selection process to be defensible from a legal perspective.

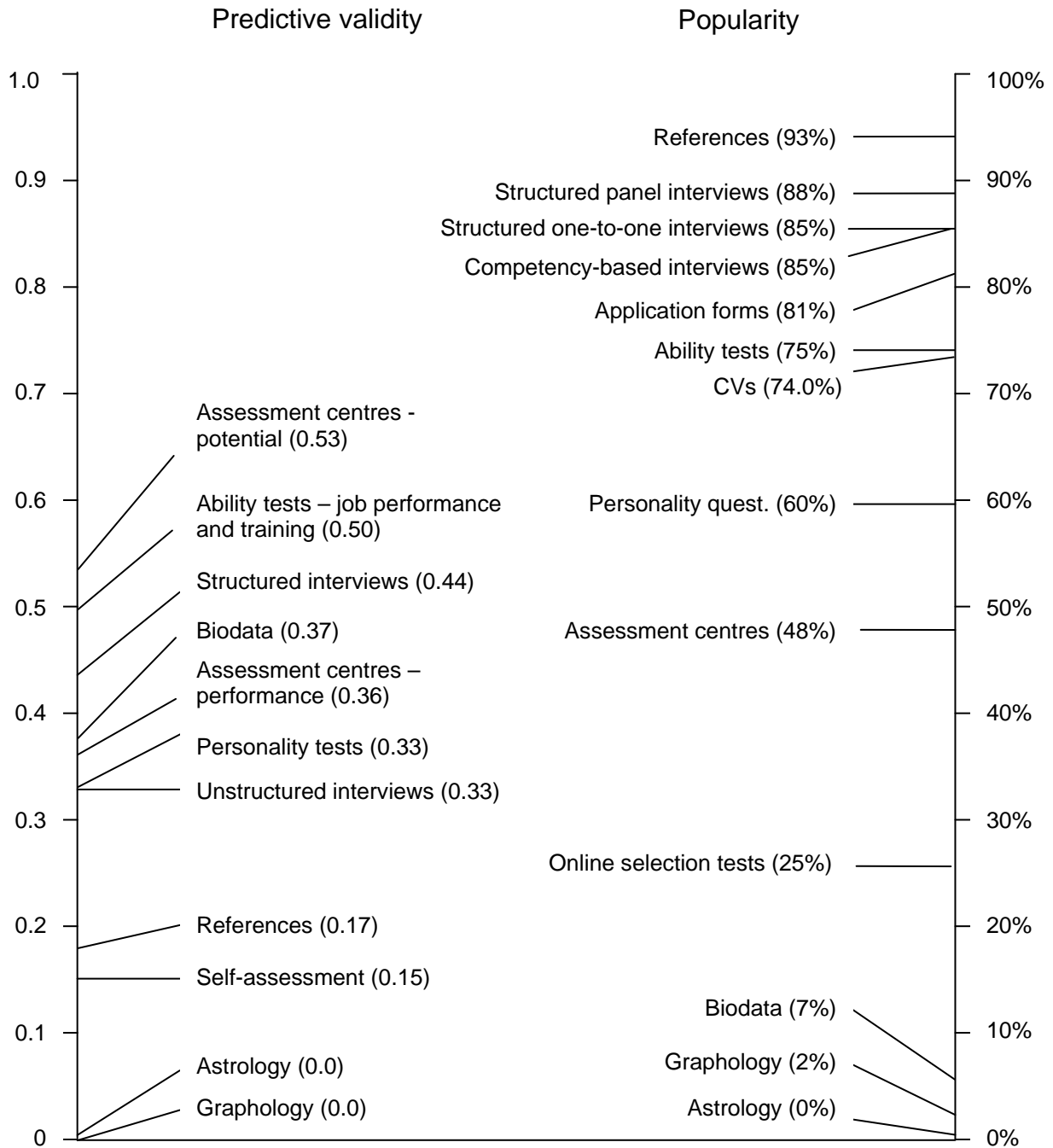
Early research on the links between test scores and subsequent job performance produced mixed results, often due to the limitations of the research itself. More rigorous research methods have since identified a considerable relationship between performance on the family of tests represented by the PfS Reasoning Tests and job performance (e.g. Bertua, Anderson and Salgado, 2005; Schmidt and Hunter, 1998).

**Figure 1** summarises the findings from a number of sources on the predictive validity and popularity of a variety of assessment methods. From this it can be seen that ability tests are amongst the better predictors of job performance and job related training success, and are one of the most frequently used assessment methods after interviews and references.

In a meta-analysis of validity data, Schmidt and Hunter (1988) showed tests of general mental ability to have a predictive validity of 0.51. Recent work using validity studies from the UK produced figures of 0.48 for the relationship between general mental ability and job performance, and 0.50 for the relationship with job related training success (Bertua *et al*, 2005). These figures may appear high until it is remembered that they have been 'corrected for criterion unreliability and range restriction'. Assessment centres have a slightly higher predictive validity but this needs to be understood in the light of the fact that reasoning tests are often part of these assessment centres. The incremental validity of assessment centres once ability tests have been allowed for is quite modest, with some estimates suggesting that, at best, they add no more than 0.1 to the correlation with job performance (Schmidt and Hunter, 1998).

A further finding of note from Bertua *et al* (2005) was the relationship between predictive validity and different occupational groups. Tests of general mental ability showed higher validities for the 'managerial' and 'professional' categories as well as such areas as engineering and sales but lower for clerical and certain kinds of operators. This suggests that such tests have greater importance for the prediction of more complex and cognitively demanding roles – as we might expect. As the authors note, this finding contradicts the assumption held by some that ability tests have less validity for more senior appointments.

Figure 1: The predictive validity and popularity of different assessment methods



Notes: figures for predictive validity taken from Bertua, Anderson and Salgado (2005), Gaugler, Rosenthal, Thornton and Bentson (1987), Hunter and Hunter (1982), McDaniel, Whetzel, Schmidt and Maurer (1994), Reilly and Chao (1982), Robertson and Kinder (1993), Schmidt and Hunter (1998). Figures for popularity based on British organisations and taken from CIPD (2000; 2006) and Shackleton and Newell (1991) and indicate use by organisations for at least some appointments

## Reasons for developing the PfS Reasoning Tests

The PfS Reasoning Tests were primarily developed to:

- meet the demand for new materials, due to the increase in psychometric testing;
- offer test users the advantages of computer-based assessments;
- give users the options of 'open' and 'closed' versions of tests assessing the same constructs to address the need for both supervised and unsupervised assessment; and
- deal with common issues with psychometric testing – issues that have been identified from extensive experience of using and training people to use psychometric assessments. These include having sufficient choice to select tests of appropriate level of difficulty according to the ability range of the individuals being tested and to have a consistent format across these levels in order to reduce the learning curve (sometimes needed since tests can have very different formats) for both administrator and candidate/testee.

The advantages of computerised testing have been recognised for some time, particularly in the areas of administration, scoring and generating reports (e.g. Kline, 2000). Within the fields of occupational, clinical and educational assessment, computer-based testing is now widely accepted. Test developers are starting to go beyond computerised versions of pencil-and-paper assessments and explore how technology can be used to create innovative and engaging new forms of assessment.

When developing the PfS Reasoning Tests, one of the goals was to make the potential benefits of testing more accessible. Using the internet as a method of delivery means that the psychometric assessments can be used more flexibly and with a wider range of people. With computer-based testing it is also easier to make tests such as the PfS Reasoning Tests visually appealing. This is important when assessing people who may have lower levels of motivation for completing the tests, as it makes the testing experience different from traditional educational assessments.

The PfS Reasoning Tests also meet the need for administrators to be able to set up and monitor the assessment process, to have control of the data and how it is used, and to generate informative reports from the test results. Through the PfS online system, administrators can control which tests are made available to test takers, which norm groups the results are compared to, what types of report are generated and who receives the reports. Security is guaranteed by the use of passwords. Computerised testing makes scoring fast and accurate.



The output from the PfS Reasoning Tests provides some added value information which goes beyond the usual raw and standardised test scores. They provide an analysis of speed and accuracy (see **Section Three**) to enable the interpreter to consider potential reasons for low or high scores which may have to do with strategy rather than just ability. These statistics are combined into reports that give development suggestions and interview prompts, through which reviewers can explore with test takers the meaning and implications of their results. These analysis and reporting facilities mean that all test takers can receive valuable, personalised feedback, regardless of the outcome of the test results. This makes the PfS Reasoning Tests truly developmental. By using the data entry and scoring facilities of the PfS online assessment system, users of the paper-based tests can also benefit from the features of the automated test report (see Section 3).

Two related challenges often faced by test users are:

- selecting the appropriate ability level of psychometric tests; and
- familiarising themselves with the formats of different tests.

Both of these issues are addressed by the PfS Reasoning Tests.

It is not usually possible to adequately cover a wide ability range with a single test. Tests provide maximum information when the mean score is close to the middle of the possible score range. A single test capable of assessing accurately across the full ability range would take too long to administer to be practical in organisational settings and would be frustrating for many test takers: able test takers would become bored with many simple questions and less able test takers frustrated at the number of questions they found too difficult. To deal with the issue of ability, there are four levels of the Verbal, Numerical and Abstract Reasoning Tests, spanning school-leavers to people with postgraduate qualifications and considerable professional experience.

Each of the three tests uses a common format. This means that once users have familiarised themselves with one level of a test, they will be equally familiar with all levels of both the closed and open versions. Users of the PfS Reasoning Tests therefore no longer have to become familiar with different test formats for different populations – simplifying the administration process. The same formats are also used for scoring and reporting, reducing the possibility of errors and making the interpretation and review of test results easier.

Test users often see identifying the appropriate test level as a major challenge, particularly when the test is to be used by a diverse group of people. The PfS Reasoning Tests address this issue through suggesting how the tests can be used as 'locator' tests. By administering one of the tests to an existing group of employees, the results can be used to determine an IRT score (see Section 6) which can then be used to suggest how they would perform on each of the four test levels and the most appropriate can then be identified for the position in question. The use and interpretation of locator tests is simplified if the computer-based versions are used.

In areas such as graduate and managerial selection and development, the use of psychometrics is well-established. As more organisations use psychometrics there is a risk that the tests become over-exposed, with applicants in some cases taking the same test more than once, so giving them an unfair advantage over others.

All new tests offer a short-term solution to the problem of over-exposure, though this has become an increasingly important issue with the advent of unsupervised testing over the internet. The PfS Reasoning Tests have also been developed with the goal of addressing this in the long-term. The open and closed versions of the Verbal, Numerical and Abstract Reasoning Tests have been developed to give users confidence in the security of the closed tests whilst retaining the option of unsupervised internet assessment using the open versions. Further parallel versions of the Verbal, Numerical and Abstract Reasoning Tests are already under development and there is also the option for bespoke assessments consisting of unique series of items to be developed for clients on request. The PfS Reasoning Tests therefore offer organisations the opportunity to avoid the problems associated with the over-exposure of tests.

Another way to address the issue of over exposure is by the use of adaptive testing – i.e. a different set of items is administered to each candidate (chosen from a bank and based on whether they got the previous item right or wrong). Whilst this has some nice features it introduces a reliance on statistical indices to provide equivalence between all the different tests – and statistical indices suffer from being rather sample dependent. A second issue is that it still does not address the key issue with online unsupervised assessment which is called 'authenticity' i.e. we still do not know whether the person who was supposed to complete the test did so and without help? Therefore the PfS method for addressing these issues is to use an open test for screening and a closed test for re-test at a later stage in the recruitment process and to investigate when there is a significant discrepancy between the two results.

To summarise the PfS Reasoning Tests available:

- Levels 1 to 4 closed tests cover the areas of Verbal, Numerical and Abstract Reasoning and are intended to be used for secure testing situations which are either supervised or where they are administered online to known test takers.
- Each level contains a unique set of items and levels are broadly tied to educational stages: Level 1 for test takers in the last years of compulsory education (years 10 and 11), Level 2 for those in further education, Level 3 for undergraduates and Level 4 for postgraduates and experienced professionals.
- Levels 1 and 2 of the open tests are intended for use under less secure conditions (e.g. during an initial sift where results are collected remotely).
- As with the closed tests, each level contains a unique set of items. Level 1 of the open tests covers the same ability range as Levels 1 and 2 of the closed tests and Level 2 of the open tests the same range as Levels 3 and 4 of the closed tests.

## **How to use the PfS tests for making comparisons**

### **The traditional approach**

When using any of the PfS tests the first stage in getting the results is to calculate the Raw Score – this is the number of items/questions that a person gets correct. However, to be told that someone got 25 right out of a possible 40 gives very little information. The test could be very hard or very easy – and the concept of hard or easy is relative. Are we talking about a group of Nobel prize winners or a group of 5 year olds?

For this reason, the interpretation of test scores is usually done against a norm group – a group of people that are a suitable basis for comparison. This approach is perfectly possible with the PfS tests and a large number of comparison groups are presented in Appendix 4. However, multiple norm groups can become confusing. Examining different samples such as 'Managers and Professionals' or 'Call Centre Staff' soon reveals that they can be a very mixed group with different levels depending on sectors, ages, seniority etc.

### The PfS approach

The PfS approach simplifies all of this by going back to basics. It is assumed that all the tests from a particular domain (Verbal, Numerical or Abstract) measure the same attribute but are targeted at different levels. This assumes that there is an underlying scale and each test is designed for a part of that scale in the same way as different thermometers can be designed for different purposes such as taking a person’s temperature or measuring the temperature of liquid gases. The principle can be portrayed as follows:

Underlying IRT scale	Test 1	Test 2
106		
105	7	
104	6	
103	5	
102	4	
101	3	7
100	2	6
99	1	5
98		4
97		3
96		2
95		1
94		
93		

The PfS tests use a methodology based on Item Response Theory (IRT) to calculate this underlying scale<sup>1</sup> (this is explained in more technical detail in Section Six). This means that scores from any test can be communicated using the same scale and hence they become, to some degree, inter-changeable. Of course, just as with the thermometer, there is still a need to select a test at the right level for the intended purpose. However, the practical value is that any test can provide an estimate of performance against any norm group – provided it is within range.

<sup>1</sup> This underlying scale will be referred to as the IRT scale throughout this User Guide



## **Section Two: Selecting and administering the Reasoning Tests**

### **Introduction**

For any test to play a valuable role in the decision-making process, it has to be matched to the abilities and competencies required by the job role. The first part of this section provides an overview of how to identify appropriate tests and introduces the facilities in the PfS Reasoning Tests series that allow the most suitable level of each test to be selected.

Good administration, whether the tests are being taken in pencil-and-paper format or using a computer, is the key to achieving reliable and valid test results. When administering the test in person, a well-defined procedure is to be followed. However, computer administration offers test takers the opportunity to complete tests in their own time, at a location of their choosing, without an administrator being present. Under these conditions the administration procedure may not be as closely controlled, but it is still possible for clear guidelines to be established. The second part of this section outlines the procedure for supervised test administration and goes on to offer guidelines for organisations on how to develop procedures for unsupervised testing.

### **Selecting appropriate tests**

The information provided by the PfS Reasoning Tests should be valuable in the decision-making process. To make sure this is the case, the abilities being assessed by the tests must relate to core job competencies. The starting point for any selection or development process is either a detailed job analysis focussing on the competencies and personal characteristics that employees need in order to perform successfully in the first role OR a careful competency analysis which identifies the kinds of skills a person would need if they were to progress further into the organisation. This relates to a basic recruitment philosophy – are you appointing someone to a particular job role or as an asset/talent for the organisation in the longer term. Since job roles and organisational structures are becoming ever more fluid there is a trend towards less emphasis on the immediate job role and more on the identification of the competencies needed for those who work in evolving work environments (and changing roles).

It is important to remember that reasoning tests can provide valuable information, but are rarely sufficient on their own. Tests should be seen as only one part of an overall assessment package. As with any form of assessment, both their strengths and weaknesses need to be acknowledged. Through drawing on the strengths of a variety of assessment methods and carefully integrating the information from them, it is possible to reach far more valid and defensible decisions that are also more likely to be viewed as fair within the framework of employment law.

In order to provide the best possible information about individuals, it is important that the correct level of each Reasoning Test is selected. If tests are not at the correct level for the group in question, their ability to differentiate between people is lowered and they may have a de-motivating effect on those who take them. It is important to recognise that selecting more difficult tests will not result in the raising of standards within an organisation. Tests give most information when scores are spread around the mid-point of the distribution; if they are too easy or too hard, scores will be more bunched together so making it difficult to reliably differentiate between the test takers. This is moderated by whether the tests are being used to screen out (give an easier test in order to identify those who cannot even show the skills at that level) or screen in (give a hard test so that only those who are genuinely at the top of the distribution can be identified). The availability of appropriate norm groups is another factor in determining test selection and also indicates for which ability levels or groups tests are suitable.

Currently, there are four levels of each of the closed Reasoning Tests and two levels of the open Reasoning Tests (referred to in the online PFS assessment system as 'Reasoning Skills Tests' to differentiate them from the closed tests). Each level has been developed to correspond to a broad ability band, as shown in **Table 1 overleaf**.

These bands should be considered as a starting point for test selection.

Level for the Closed Reasoning Tests	Approximate educational level of the norm group	Level for the Open Reasoning Skills Tests
Level 1	This covers the top 95% of the population and is broadly representative of the general population.	Level 1 and Combined Reasoning Skills Test
Level 2	This covers the top 60% of the population and is broadly representative of people who have studied for A/AS Levels, GNVQ Advanced, NVQ Level 3 and professional qualifications below degree level	
Level 3	This covers the top 40% of the population and is broadly representative of the population who study for a degree at a British University or for the BTEC Higher National Diploma/Certificate, NVQ Level 4 and other professional qualifications at degree level	Level 2
Level 4	This covers the top 10% of the population and is broadly representative of the population who have a postgraduate qualification, NVQ Level 5 and other professional qualifications above degree level	

*Table 1: Correspondence between the PfS Reasoning Tests and level of ability as indicated by the level of educational attainment*

Remember that when choosing a test, the level is determined by the challenge of the task/role rather than the level of the applicants. For example, if the challenge of the role requires cognitive abilities at graduate level then it is perfectly acceptable for applicants who have only studied to A-level to complete a graduate level test. And we know that many A-Level students could easily have been graduates if they had chosen that path. Therefore test selection is not based on the level of the individual unless the purpose is for personal guidance and development. **Table 1** should be used to give a rough initial guide based on an understanding of the cognitive challenge in the job. Deciding on the appropriate level can be more difficult when the tests are being used for multi-faceted jobs and/or with less homogenous group, (i.e. when some are graduates and others have considerable work experience but limited academic qualifications). When the most suitable test level is not immediately apparent, users may consider identifying the appropriate level by using a 'locator' test. A description of how to use a locator test is given below.



## Appendix Three: Comparison tables – IRT Scores to Raw Scores

All the PfS tests have been linked using Item Response Theory (see Section Six above). This means that the raw score on any test can be converted into an IRT score. This is done by first identifying the table for the type of test (Verbal, Numerical or Abstract) and then locating the column for the type and level of test used. The next step is to find the person's raw score in this column and reading off the IRT score in the far left-hand column. In the same row will be the raw score that the person would be estimated to obtain had they completed a different level test. Since such scores are estimates there is not always a one-to-one correspondence and there is often a range of corresponding scores. For example a raw score on Verbal CL1 of 8 results in an IRT score of 57 to 60 and the estimated raw score on Verbal CL3 is between 6 and 7. This information can then be used to estimate the person's performance against any of the norm groups available for Verbal Closed Level 3 as well as the norm groups for Verbal Closed Level 1 (and by the same process for Level2, Level4 and Open Level 1 and 2). However, the user is reminded that such scores are estimates. The errors of estimate increase when there is a mismatch between the person's ability and the test's difficulty (i.e. when the raw scores are either very low or very high).

Verbal Tests – IRT to Raw Score conversions						
IRT score	Closed (Critical Reasoning) Tests				Open (Reasoning Skills) Tests	
	level 1	level 2	level 3	level 4	Level 1	Level 2
0	0				0	
1	0				0	
2	0				0	
3	0				0	
4	0		0	0	0	
5	0		0	0	0	
6	0		0	0	0	
7	0		0	0	0	0
8	0		0	0	0	0
9	1		0	0	0	0
10	1		0	0	0	0
11	1		0	0	0	0
12	1		0	0	0	0
13	1	0	0	0	1	0
14	1	0	0	0	1	0
15	1	0	0	0	1	0
16	1	0	0	0	1	0
17	1	0	1	0	1	0
18	1	0	1	1	1	0
19	1	0	1	1	1	0
20	1	0	1	1	1	1
21	1	0	1	1	1	1
22	1	0	1	1	1	1
23	2	0	1	1	1	1
24	2	0	1	1	1	1

## Core Norm tables available for the PfS Reasoning Tests

### Core Norms for Closed Tests

Test	Level	Norm Group	Sample Size
Verbal	1	GCSE Students and students in their first year of courses at FE Institutions	210
	2	FE students studying a range of vocational and academic courses.	303
	3	Undergraduate students from a range of universities across the UK.	1322
	4	Postgraduate students studying for an M.Sc., Ph.D. or MBA.	1131
Numerical	1	GCSE students and students in their first year of courses at FE institutions	250
	2	FE students studying a range of vocational and academic courses.	337
	3	Undergraduate students from a range of universities across the UK.	1609
	4	Postgraduate students studying for an M.Sc., Ph.D. or MBA.	1510
Abstract	1	GCSE Students and students in their first year of courses at FE Institutions	156
	2	FE students studying a range of vocational and academic courses.	242
	3	Undergraduate students from a range of universities across the UK.	860
	4	Postgraduate students studying for an M.Sc., Ph.D. or MBA.	881

## Appendix Five: Norm table descriptions for Open tests

### Description of Norms for Open Tests and Combined Reasoning Tests

Test	Level	Norm Group	Sample Size
Verbal	1	General population	3815
		Students aged 17 to 19	762
		Internal applicants from a public service organisation completing the test as preparation for an internal selection process (compiled 2007).	768
	2	University Undergraduates	18,554
Postgraduate Students		1203	
Numerical	1	General population	4168
		Students aged 17 to 19	843
		Internal applicants from a public service organisation completing the test as preparation for an internal selection process (compiled 2007).	689
	2	University Undergraduates	23756
Postgraduate Students		2012	
Abstract	1	General population	2348
		Students aged 17 to 19	850
	2	University Undergraduates	22,034
		Postgraduate Students	733
Combined Reasoning Test		Students aged 15 to 17 (General Population)	1375
		Year 10 to 12 Students in compulsory education from non-selective UK schools	345
		Year 10 to 12 Students in compulsory education from selective UK schools	290

## Appendix Six: Norm table supplements for the Closed tests

Description of Norms table supplements for the Closed Tests

Test	Level	Norm Group	Sample Size(s)
Verbal	3	Graduate applicants to a leading financial services institution (compiled 2005).	420
Numerical	3	Graduate applicants to a leading financial services institution (compiled 2005).	435
Abstract	3	Graduate applicants to a leading financial services institution (compiled 2005).	415
Verbal	4	Postgraduate students at a business school of a UK London University (Compiled 2005).	894
Numerical	4	Postgraduate students at a business school of a UK London University (Compiled 2005).	1211
Abstract	4	Postgraduate students at a business school of a UK London University (Compiled 2005).	573
Verbal	4	British MENSA members (Compiled 2005)	193
Numerical	4	British MENSA members (Compiled 2005)	220
Abstract		British MENSA members (Compiled 2005)	144
Verbal	2	Graduate applicants to 'Operations' roles in a multi-national Financial Services Organisation (Investments & Fund management – 2006 to 2012)	1942
Numerical	3	Graduate applicants to a multi-national Financial Services Organisation (Investments & Fund management – 2006 to 2012)	2390
Verbal, Numerical & Abstract	3	Graduate applicants to graduate positions in a multi-national Insurance Company (compiled 2005)	465/483/462
Verbal, Numerical & Abstract	4	Applicants with degree level qualifications applying to a leading UK Business School (2003-2010)	995/1408/646
Verbal, Numerical & Abstract	4	Senior professionals and administrators applying to a regulatory body (2011)	595/589/588

We hope you have enjoyed exploring this free introductory version of the Reasoning Tests User's Guide. We would be delighted to get your feedback and to discuss your thoughts and reactions.

In addition, we hope that you have got a flavour of what the Reasoning Tests are all about. We realise that not all the features that make this approach innovative and practical have been explained, but you will get some idea by referring to the Table of Contents and the full version will give you more detail about:

1. how the whole suite of reasoning tests were developed
2. the practicalities of administering in order to engage the person in the process
3. requirements for both the software and/or the paper versions
4. ideas about conducting a review (feedback) session
5. the technical psychometric details related to construction, reliability and validity
6. how the tests are all linked using IRT (Item Response Theory) which enables comparisons to be made across different versions of the tests and which provides access to more than 40 different norm tables

If you are interested in the complete version, (normally provided to our affiliated partners and clients as part of our training courses) or in discussing your interest further, we welcome you to get in touch by filling in the [Contact Form](#)