

Selector e-Profiler II

User Manual and Technical Description

May 2003

Copyright © 2003 Selector Limited

Contents

Part I	4
Administration Guidelines	5
Selector e-Profiler II Questionnaire	6
Selector e-Profiler II Report	7
Ability Assessment	7
Personal Styles.....	7
Work Preferences	9
Job Ideals	10
Behavioral Scale Summary	11
Report Interpretation Guidelines.....	12
Ability Assessment	12
Personal Styles and Work Preferences	13
Part II.....	14
Theoretical Basis and Construction.....	15
Ability Assessment	15
Personal Styles.....	15
Work Preferences	17
Introduction to Psychometric Concepts.....	18
Reliability	19
Test-Retest Reliability	19
Internal Consistency Reliability	20
Validity.....	23
Construct Validity.....	23
Criterion Related Validity	27
Criterion Related Validity: In the Workplace	31
Normative Base.....	36
References	38

Introduction

Selector e-Profiler II (eP2) is an online candidate screening and selection instrument developed for use in recruitment, team development, succession planning and training. Research consistently finds that the most effective method of selecting people for roles, other than work samples, is the psychometric profiling of ability and personality dimensions (Schmidt & Hunter, 1998). The robust psychometric characteristics of the eP2 scales coupled with the flexibility of online delivery make eP2 an indispensable part of your recruitment and organisational development solution.

This document is comprised of two parts:

Part I is the user manual for eP2. It contains guidelines for how the questionnaire should be administered and describes the various sections of the questionnaire. Details of each report scale are provided along with report interpretation guidelines.

Part II contains the technical description of eP2. It outlines eP2's development procedure and provides evidence for its reliability and validity. It describes the development sample and examines gender differences.

Part I

Selector e-Profiler II User Manual

Copyright © 2003 Selector Limited

Administration Guidelines

Ensure the candidate has a quiet working environment and will not be distracted by phones or other people. Inform the candidate of the following points:

- Selector e-Profiler II (eP2) is a 45-55 minute assessment that will provide an objective indication of:
 - How you relate to others
 - How you go about your work
 - What is important to you in your job
 - Your reasoning abilities
- It is important to be as honest as possible when answering questions.
- The questionnaire requires you to answer all questions.
- You will not be penalized for guessing where you are unsure of the correct answer during the abilities section.
- It is a good idea to have a pen and paper available to help work out problems.

Selector e-Profiler II Questionnaire

The ep2 assessment is web-based. This means that the candidate's computer must be connected to the Internet for the duration of the questionnaire. To protect from Internet connection or web browser failure, ep2 constantly keeps track of the last question completed. If the connection to an assessment is lost for any reason, the candidate can simply log back on and continue from where they were up to.

The questionnaire begins with a Welcome screen that summarizes the subsequent four sections.

Personal Details

This section collects personal data that is used to help ensure the assessment is not inadvertently discriminating against any group. Aside from the candidate's name and gender, none of the data collected appears in the report, nor does it affect the candidate's results.

My Personal Style

This section contains 53 questions that investigate the candidate's interpersonal, working and coping styles. Questions are posed in terms of how strongly one agrees or disagrees with various statements.

My Ideal Job

This section contains 64 questions related to work preferences comprise the My Ideal Job section. Candidates are asked how desirable various job characteristics are to them.

Ability Measure

The Ability Measure contains 30 questions that investigate the candidate's verbal, numeric and logical reasoning ability.

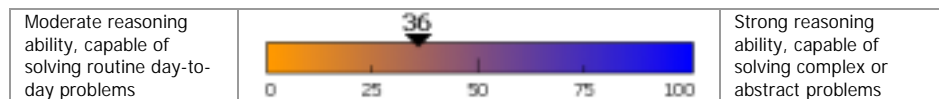
Selector e-Profiler II Report

The dimensions across which candidates are profiled have been selected for their relevance to screening and selection in employment settings, and are based on current psychological theory and best practice in applied psychology. This section describes the scales across which ep2 assesses a candidate. Traits or tendencies of those that score at the upper and lower end of each scale are provided.

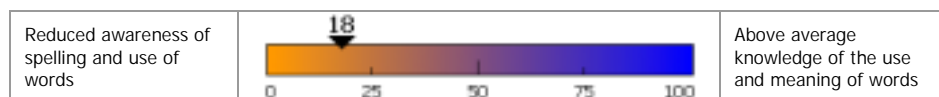
Ability Assessment

The ability section reports overall reasoning aptitude and the component verbal, numerical and logical reasoning scores.

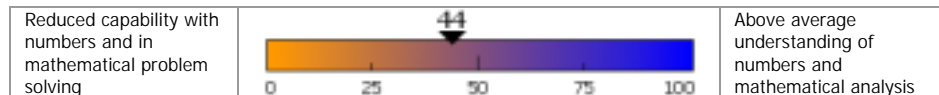
Overall Reasoning Aptitude



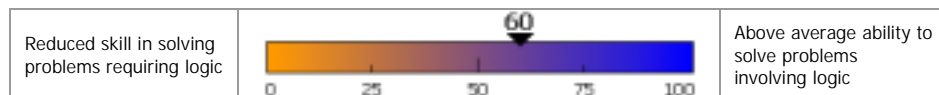
Verbal Reasoning



Numerical Reasoning



Logical Reasoning

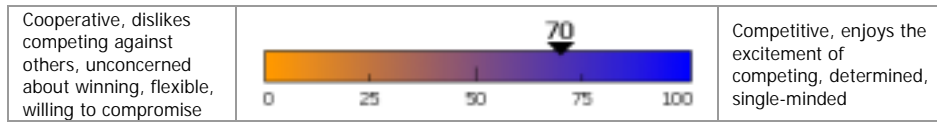


Personal Styles

This section presents the candidate across personality dimensions that are consistently proven to predict job performance when overall job performance is the criterion, and when more specific criteria are being predicted. The results of this section are particularly useful in assessing likely team fit.

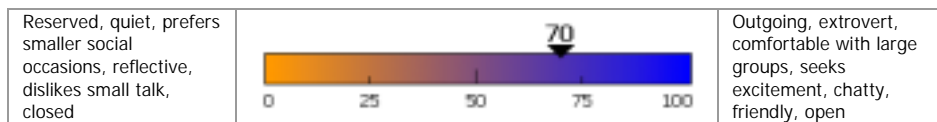
Competitiveness

Competitiveness is characterized by a need to compete with others, meet goals and can reflect a need to establish one's worth in relation to others.



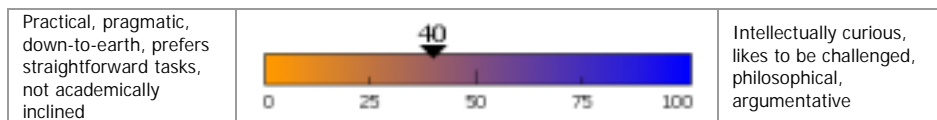
Extroversion

Extroversion is a reflection of the extent to which a person is energized by interacting with others. Extroverts are naturally outgoing, enjoy meeting and talking to people and feel comfortable socially. They seek out action and variety.



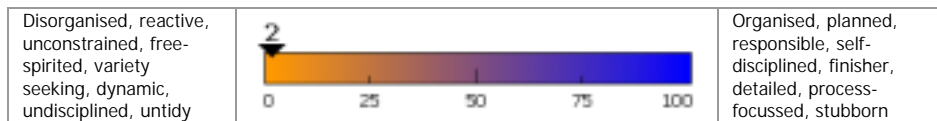
Openness to Ideas

Openness to ideas reflects interest in novel ideas and experiences. It is characterized by a natural sense of curiosity, an interest in concepts and theories, and a willingness to debate ideas and opinions.



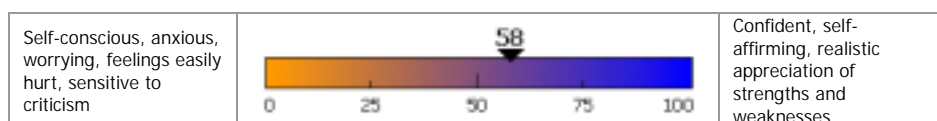
Orderliness

Orderliness measures the need for order and structure. It embodies reliability, responsibility, conscientiousness and constraint.



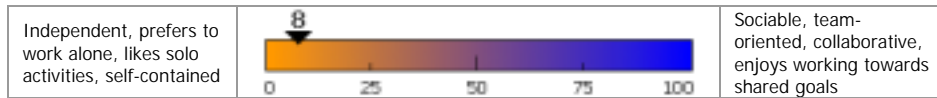
Self-Confidence

Self-confidence reflects the extent to which a person's sense of value or worth is based on the person's own views or on the opinion of others. Those who rely on their own judgment tend to be less anxious and more at ease than those who rely on others for their sense of self-worth.



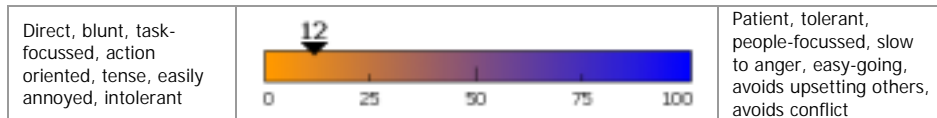
Teamwork

Teamwork indicates the need to work with others towards common goals and the willingness to fit in with the wishes and needs of others.



Tolerance

Tolerance is the ability to accept others as they are and is characterized by an ability to remain non-judgmental, maintain an even, patient manner in difficult or tense situations and avoid becoming angry or upset.

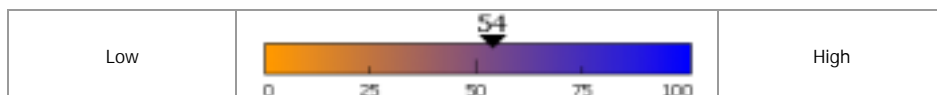


Work Preferences

This section presents the candidate's work preferences. Work preferences are important because people are more likely to excel in their work if they find it enjoyable. The results of this section can be used to assess the similarity between an individual's preferred work environment and the work environment of the position for which they are applying. As for the Personal Styles section, this section reports scales in rank order.

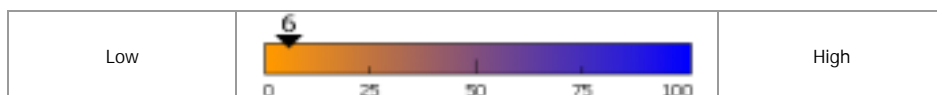
Autonomy

The Autonomy scale assesses the importance of working unsupervised and having direct influence over the course of work.



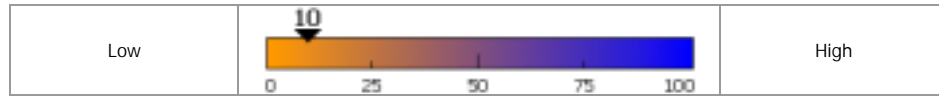
Complexity

The Complexity scale measures the importance of doing work that is mentally demanding and complex.



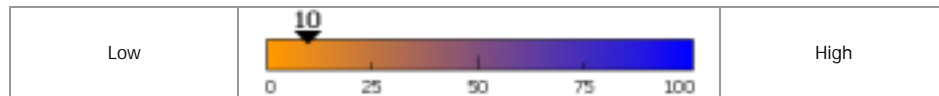
Interaction

The Interaction scale assesses the importance of interacting with and helping others in the work environment.



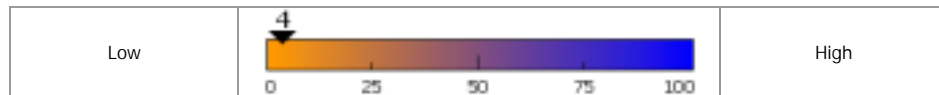
Physical

The Physical scale assesses the importance of working outside and being involved in physical work.



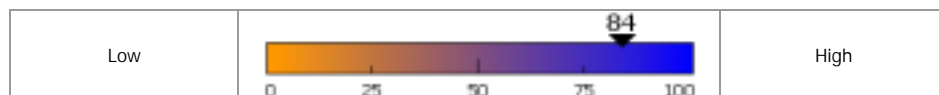
Predictability

The Predictability scale measures the importance of working in a stable, supportive, well-organised workplace with secure employment.



Pressure

The Pressure scale assesses the importance of doing work that requires a high level of effort and commitment.



Job Ideals

This section presents the candidate's responses to the My Ideal Job section of the questionnaire in two different ways.

First, it presents the candidate's raw responses. Job characteristics are grouped by what the candidate finds 'Absolutely essential' in a work role, right down to those that are 'Undesirable'. This presentation serves as a very convenient checklist for mapping what the candidate prefers with what a job role can offer.

Secondly, the job characteristics are presented in terms of how important they are to the candidate in comparison to others who have completed the questionnaire. This view can be used to temper the results of the previous view. For example, most candidates will answer positively to the 'offers good pay' characteristic. It is interesting to note however, whether compared to others, this characteristic is particularly important.

Behavioral Scale Summary

This section recasts each of the Personal Styles and Work Preference scales in a fixed order, making candidate comparison possible at a glance.

Report Interpretation Guidelines

The following guidelines are presented to help increase the effectiveness of your interpretation of the eP2 report.

Ability Assessment

Consider the importance of various abilities for the role concerned. If a measured ability is not critical to the role then the score obtained is less important.

Verbal Reasoning

Verbal reasoning is a measure of the level of competency a person has with written language, spelling and meaning of words.

Important questions to consider:

- Is the ability to accurately convey meaning and express oneself in writing important to the role?
- How important are spelling and grammatical correctness to the job under consideration?
- Will the person be writing critical memos or letters to important clients?

Numerical Reasoning

Numerical reasoning is a measure of the level of competency a person has with numbers, numerical problems and the relationships between numbers.

Important questions to consider:

- Is the ability to work comfortably and easily with numbers important to the role?
- Will the person be dealing with critical accounts or data entry?
- What are the potential risks if numerical errors are made?
- If calculations are important to the role the application of a specific skill based test that mirrors the requirements of the role may be appropriate.

Logical Reasoning

Logical reasoning measures the ability to make inferences and solve complex problems given information from which a solution can be derived.

Important questions to consider:

- Is the ability to solve complex problems important to the position?
- Will the person be required to provide accurate and verifiable solutions to complex problems?

Personal Styles and Work Preferences

eP2 reports the behavioral scales in rank order, from the candidate's highest score down to the lowest. An individual's strongest traits will thus feature early in the report and the order will differ from candidate to candidate.

Determine the traits or success factors that are critical to a role. This should occur through job analysis. If a job description exists, this is likely to be the basis by which the critical success and fail factors are identified. Consider also those characteristics that may be undesirable in a role or result in a higher risk of failure.

Consider the candidate's work preferences. The extremes will be the most pertinent. Look to confirm a potential 'fit' or to identify a mismatch between candidate and role. Consider whether the candidate will indeed receive the Job Ideals indicated in the report that will motivate them to perform.

Utilize eP2 as a tool to provide direction for further investigation. Use interviews and reference checks to clarify or confirm any possible concerns and situational examples to probe for behavior in the workplace.

Part II

Selector e-Profiler II Technical Description

Copyright © 2003 Selector Limited

Theoretical Basis and Construction

Ability Assessment

Theoretical basis

Research consistently shows that measures of cognitive ability are among the strongest predictors of future job performance (Schmidt & Hunter, 1998). e-P2 measures cognitive ability across the *Overall Reasoning Aptitude* scale and its component *Verbal Reasoning*, *Numerical Reasoning* and *Logical Reasoning* scales, for the purpose of employee screening in recruitment and selection and organisational development initiatives such as training and team building.

Development Methodology

Item Development

An item set was developed covering the three major aspects of general cognitive ability: verbal, numerical and logical reasoning.

Scale Development

The item set was administered to 503 participants. Responses were then analyzed using principle component factor analysis. The decision to use principle components was made on the basis of a number of considerations.

While the mean of 5.01 for the first factor emerging (logical reasoning) from the orthogonal rotation of principle component analysis was the lowest, the second (numerical) emerged with a higher mean than the third (verbal), suggesting that the structure is not solely the result of item difficulty.

The final arbiter of whether or not to use a principle components solution was whether or not the factor structure made sense. In light of the factor structure that emerged, we are confident of real world meaningfulness. Verbal, numerical and logical reasoning are routinely found in the factor analysis of ability; furthermore the solution reflected our hypothesized scale structure.

Three scales emerged from factor analysis of the ability data:

1. *Verbal Reasoning* measures verbal fluency and word knowledge.
2. *Numerical Reasoning* measures number awareness and numerical analysis.
3. *Logical Reasoning* measures logical analysis and the ability to solve problems given relevant information.

Personal Styles

Theoretical basis

Research consistently finds that there are certain aspects of personality that predict future job performance. In particular, conscientiousness predicts across all occupations when overall job performance is the criterion (Barrick, Mount & Judge 2002, Hunter & Schmidt 1998). Emotional stability is also

found to be a strong predictor when overall job performance is the criterion; however it is less stable as conscientiousness in terms of predicting more specific performance criteria.

The remaining factors of the big five model (extraversion, agreeableness and openness to experience) also predict performance, although these factors tend to do so for specific occupations and specific job performance criteria only. Taken together, the research described provides a compelling case for the use of personality testing in personnel selection and recruitment settings.

Development methodology

Item development

Items were developed based on the five factor model of personality and to cover aspects found to be critical to success and failure in the workplace from the experience of psychologist Keith McGregor.

Scale development

The items were administered to 503 participants and the results analyzed using principle components analysis. The factors that emerged after orthogonal rotation of seven factors (based on a scree test), and the dropping of ambiguous (double and triple loading) items and unrelated items (items loading less than 0.3) are as follows:

1. Extroversion
2. Orderliness
3. Openness to Experience
4. Teamwork
5. Tolerance
6. Competitiveness
7. Self-Confidence

Two of the factors are not typically considered personality factors under the five-factor model: *Competitiveness* and *Teamwork*. While we can be confident from the factor structure that they are measuring constructs independent of the other scales, at this stage, they are supported only by face validity, and the extent to which they are artifacts of the process of factor analysis will ultimately be determined through predictive validity studies investigating the relationship between test scores and job performance criteria.

Work Preferences

Theoretical basis

Preference inventories identify job characteristics that a candidate finds desirable and undesirable. The reason for including a preferences inventory is based on the premise that people perform better in roles they enjoy.

Development methodology

Item development

Items were developed based on relevant psychological theory and the experience of psychologist Keith McGregor.

Scale development

The items were administered to 503 participants and the results analyzed using principle components analysis. The scree test suggested a six factor solution:

1. Pressure
2. Autonomy
3. Interaction
4. Physical
5. Predictability
6. Complexity

Introduction to Psychometric Concepts

Psychometric assessments have demonstrated validity in selection settings (Schmidt & Hunter 1998). One of the fundamental advantages that psychometric testing has over other selection methodologies is that it adds standardization to the selection procedure. This means that all those in the selection process are exposed to the same selection procedure, which will help ensure equity in the selection process.

The demonstrated ability of selection instruments to add to the effectiveness of the selection process, coupled with the ability of psychometric assessments to provide insights into traits not easily assessable through other techniques, provide a compelling argument for the use of psychometrics in selection settings.

To fully capitalize on the benefits that psychometric testing can offer, it is important that the tests being used have sound psychometric properties. In psychometrics, there are two principle criteria that determine whether a psychometric test is appropriate in a given setting. These are the reliability of the assessment and the validity of the assessment.

The following two sections introduce these key concepts before applying them to ep2. For further reading on this topic we recommend consulting a book such as Kline's *Psychometrics Primer (2000)*.

Reliability

Test-Retest Reliability

If a test has high test-retest reliability, there is little chance that on a subsequent occasion a candidate will obtain a score that differs from their original score. It is crucial that test-retest reliability is high. If it is not high, either the test scales are unreliable, or the person has actually changed on the dimension in the period between the two testing occasions. In order to examine the test-retest reliability of a test, an assessment is made of the similarity between an individual's test scores over two occasions.

The test-retest reliability of a test is typically measured by a correlation coefficient, which varies between -1 and 1 . A coefficient of -1 indicates a strong negative relationship between the two test scores, while a coefficient of 1 would indicate a perfect positive relationship between the scores. The closer the test-retest reliability is to 1 the stronger the relationship between two test scores. The benchmark set for test-retest reliability of ep2 is 0.70 . This would indicate strong test-retest reliability.

The data presented below represent the test retest reliability of a subsection of 165 participants in the development sample over an interval of three months.

Ability assessment

The test-retest reliability coefficients for the eP2's *Overall Reasoning Aptitude* and its subscales is strong, with just *Numerical Reasoning* falling below our stated goal of 0.70 , and only by $.01$.

Scale	Reliability
Overall Reasoning	0.82
Verbal	0.70
Numerical	0.69
Logical	0.70

n = 165, all correlations significant at $p < .01$

Table 1: Test-retest reliability of the Ability Measure

Personal styles

The test-retest reliability data for the Personal Styles section is strong. With the exception of the *Self-Confidence*, all test-retest reliability coefficients are close to or in excess of 0.80 .

Scale	Reliability
Extroversion	0.88
Orderliness	0.82
Openness to Exp.	0.81
Teamwork	0.80
Tolerance	0.79
Competitiveness	0.79
Self-Confidence	0.69

N = 165, all correlations significant at $p < .01$

Table 2: Test-retest reliability of Personal Styles

Work preferences

The test-retest reliability data of the Work Preferences scales is strong. All scales show test retest reliability coefficients of 0.70 or greater.

Scale	Reliability
Physical	0.82
Predictability	0.79
Pressure	0.79
Autonomy	0.70
Complexity	0.81
Interaction	0.84

N = 165, all correlations significant at $p < .01$

Table 3: Test-retest reliability of Work Preferences

Internal Consistency Reliability

An alternative indicator of the stability of a scale is split-half reliability. This splits the test up into two equivalent halves and assesses the relation between the two halves. The most common measure of split-half reliability is Kuder-Richardson KR-20. This can be proven mathematically to be the mean of all the possible split half reliabilities of a given test. Kuder Richardson KR-20 deals with the reliability of right-wrong response items, such as in the case of ability tests. A generalized formula altered to deal with multiple choice personality questionnaire items is alpha.

Because it is the average of all possible split-half combinations it is referred to as a measure of internal consistency of the test. The stronger the positive relationship among scale items and between items and the scale, the higher the internal consistency of the test, and the closer the coefficient alpha is to the test-retest reliability coefficient. Alpha can be thought of as a ratio of true variance to error variance. An alpha that is too high indicates that there could be redundancy in the scale; however an alpha that is too low indicates that the items in the scale are not measuring the same trait. Accordingly, an alpha level of between 0.7 and 0.9 is the standard that we aimed to achieve.

The standard error of measurement

Essential to evaluating the appropriateness of any statistical test is the concept of the standard error of measurement. This is a band that is placed around the score an individual obtains, and indicates that due to the non-perfect reliability of a scale, an individual's score may actually fall either side of the observed score. The smaller the standard error of measurement of a

scale the more confident we can be of the accuracy of the measurement. The standard error of measurement is provided for all eP2 scales.

The internal consistency and standard error of measurement data presented below was calculated from a sample of 755 job applicants. The data was collected over the year from March 2002 to March 2003. The demographic data for these 755 people is presented in the norms section of this manual.

Ability Measure

The *Overall Reasoning Aptitude* has good internal consistency reliability. The subscales on which we report have moderate reliabilities. Given that the alpha coefficient is a function of both internal consistency and scale length, it is not surprising that 10 item sub scales have moderate internal consistency reliability. Accordingly it is recommended that the total score be the arbiter of the selection decision, and the subscale scores be considered as indicative. Further work identified for the eP2 ability measure includes development to increase subscale internal consistency reliability.

Scale	No Items	Mean	Std dev	Alpha	Std error
Overall Reasoning	30	17.88	5.29	0.80	2.36
Verbal	10	6.36	2.18	0.60	1.38
Numerical	10	5.60	2.00	0.56	1.33
Logical	10	4.45	2.11	0.57	1.38

Table 4: Internal consistency reliability of the Ability Measure

Personal Styles

The internal consistency reliability of the Personal Styles section is strong. With the exception of the *Self-Confidence*, all reliabilities are over 0.7, indicating that the items within the scales themselves are measuring the same construct.

Scale	No Items	Mean	Std dev	Alpha	Std error
Extroversion	7	29.28	5.08	0.84	2.03
Orderliness	9	40.96	6.00	0.85	2.29
Openness to Experience	8	34.09	4.27	0.72	2.24
Teamwork	6	25.13	3.66	0.75	1.81
Tolerance	8	34.27	4.44	0.70	2.45
Competitiveness	8	29.37	4.89	0.72	2.58
Self-Confidence*	7	23.32	3.54	0.53	2.42

Table 5: Internal consistency reliability of Personal Styles

*Due to the lower internal consistency reliability of the Self-Confidence scale, it is recommended that the scale be used only as an indicator that must be backed up through exploration of information from another source, such as interview or referee check. Selector will continue to develop and refine this scale.

Work Preferences

The internal consistency reliability of the Work Preferences section is strong, providing evidence that the scales are measuring homogenous constructs.

Scale	No Items	Mean	Std dev	Alpha	Std error
Physical	5	14.74	5.31	0.83	2.20
Predictability	5	22.76	4.51	0.83	1.87
Pressure	5	16.75	4.34	0.75	2.18
Autonomy	5	23.71	2.96	0.66	1.73
Complexity	5	21.72	3.76	0.72	1.98
Interaction	5	22.39	3.49	0.68	1.98

Table 6: Internal consistency reliability of Work Preferences

Validity

Kline (2000) described validity as the extent to which a test measures what it purports to measure. Clearly validity is an important characteristic of psychometric tests. The question being asked when we investigate the validity of a test is whether or not the instrument is suitable for the use we intend. There are a number of approaches to assessing the validity of a psychometric test. The lowest level of validity is known as face validity – to answer this question we simply ask whether or not, at a surface level, the test appears appropriate for its intended use. This is insufficient justification for assessment of a test's appropriateness, and all good tests will have evidence of construct validity and criterion related validity.

Construct Validity

To demonstrate that a test has construct validity, we must first show that the test has interpretable factors or scales. By interpretable, we mean that the scales of a test are measuring separate constructs. Once we demonstrate that our scales can account for where one psychological construct ends, and a new one begins, we have the basis of construct validity. We demonstrate that we have interpretable factors in eP2 through the statistical procedure of factor analysis. Factor structures demonstrate scale independence, and add to the evidence for the construct validity of assessments.

Ability Measure

Scale	1	2	3
Logical Reasoning 1	0.63		
Logical Reasoning 2	0.56		
Logical Reasoning 3	0.56		
Logical Reasoning 4	0.52		
Logical Reasoning 5	0.47		
Logical Reasoning 6	0.45		
Logical Reasoning 7	0.41		
Logical Reasoning 8	0.39		
Logical Reasoning 9	0.36	0.30	
Logical Reasoning 10	0.33		
Numerical Reasoning 1		0.65	
Numerical Reasoning 2		0.51	
Numerical Reasoning 3		0.50	
Numerical Reasoning 4		0.46	
Numerical Reasoning 5		0.44	
Numerical Reasoning 6		0.43	
Numerical Reasoning 7		0.41	
Numerical Reasoning 8		0.41	
Numerical Reasoning 9		0.36	
Numerical Reasoning 10		0.36	0.32
Verbal Reasoning 1			0.54
Verbal Reasoning 2			0.50
Verbal Reasoning 3			0.49
Verbal Reasoning 4			0.49
Verbal Reasoning 5			0.44
Verbal Reasoning 6			0.44
Verbal Reasoning 7			0.41
Verbal Reasoning 8			0.36
Verbal Reasoning 9			0.36
Verbal Reasoning 10			0.33
% Variance explained	8.87	8.74	7.87

Table 7: Factor structure of the Ability Measure

Ability Measure subscale inter-correlations

Further support for the independence of the scales is provided by the subscale inter-correlations, presented in table 8 below.

	Total	Logical	Numerical	Verbal
Overall Reasoning	-	-	-	-
Logical	0.83	-	-	-
Numerical	0.78	0.49	-	-
Verbal	0.71	0.36	0.34	-

**All correlations significant at $p < .05$

Table 8: Correlations between Ability Measure subscales

Personal Styles

Scale	1	2	3	4	5	6	7
Extroversion 1	0.79						
Extroversion 2	-0.77						
Extroversion 3	0.77						
Extroversion 4	-0.73						
Extroversion 5	0.69						
Extroversion 6	-0.61						
Extroversion 7	0.57						
Orderliness 1		-0.73					
Orderliness 2		-0.71					
Orderliness 3		0.67					
Orderliness 4		0.67					
Orderliness 5		-0.66					
Orderliness 6		0.60					
Orderliness 7		0.59					
Orderliness 8		0.57					
Orderliness 9		-0.54					
Openness to Exp. 1			0.67				
Openness to Exp. 2			0.66				
Openness to Exp. 3			0.61				
Openness to Exp. 4			0.60				
Openness to Exp. 5			0.57				
Openness to Exp. 6			0.54				
Openness to Exp. 7			-0.53				
Openness to Exp. 8			0.46				
Teamwork 1				0.84			
Teamwork 2				0.74			
Teamwork 3				0.71			
Teamwork 4				0.66			
Teamwork 5				-0.64			
Teamwork 6				-0.58			
Tolerance 1					0.71		
Tolerance 2					0.67		
Tolerance 3					-0.66		
Tolerance 4					0.63		
Tolerance 5					0.60		
Tolerance 6					0.50		
Tolerance 7					0.41		
Tolerance 8					0.31		
Competitiveness 1						0.76	
Competitiveness 2						0.69	
Competitiveness 3						-0.66	
Competitiveness 4						-0.60	
Competitiveness 5						-0.48	
Competitiveness 6						0.43	0.30
Competitiveness 7						0.39	
Competitiveness 8						0.39	
Self-Confidence 1							0.57
Self-Confidence 2							0.55
Self-Confidence 3							0.55
Self-Confidence 4							0.49
Self-Confidence 5							0.47
Self-Confidence 6							0.45
Self-Confidence 7							0.44
% Variance explained	7.85	7.71	6.76	6.43	6.15	5.91	5.02

Table 9: Factor structure of Personal Styles

Work Preferences

Scale	1	2	3	4	5	6
Physical 1	0.82					
Physical 2	0.77					
Physical 3	0.74					
Physical 4	0.72					
Physical 5	0.70					
Predictability 1		0.76				
Predictability 2		0.73				
Predictability 3		0.72				
Predictability 4		0.72				
Predictability 5		0.67				
Pressure 1			0.74			
Pressure 2			0.73			
Pressure 3			0.68			
Pressure 4			0.67			
Pressure 5			0.63			
Autonomy 1				0.74		
Autonomy 2				0.74		
Autonomy 3				0.72		
Autonomy 4				0.59		
Autonomy 5				0.51		
Complexity 1					0.78	
Complexity 2					0.73	
Complexity 3					0.63	
Complexity 4					0.63	
Complexity 5					0.60	
Interaction 1						0.72
Interaction 2						0.65
Interaction 3						0.61
Interaction 4			0.32			0.59
Interaction 5						0.55
% Variance explained	10.30	9.66	9.59	8.68	8.57	7.96

Table 10: Factor structure of Work Preferences

Criterion Related Validity

Criterion related validity assesses the degree to which a test relates to appropriately selected criteria. These criteria may be other tests known to be effective measures of the construct being measured, or appropriately selected real world criteria.

We can ensure we have a valid assessment by ensuring that the test scales reflect the current state of theory, measure independent constructs, correlate with real world criteria and other measures known to reflect the construct, and appear relevant to inspection by laypeople.

Concurrent validity is the most common form of criterion related validity. It is more common than predictive validity because, as the name implies, measurements on both the predictor and the criterion are taken at the same time; obviating the need to wait for long periods of time before having criterion related data.

Ability Measure

Having demonstrated a basis for the construct validity of the Ability Measure, including the independence of the scales and the theoretical support for the proposed structure, it is important to demonstrate criterion related validity for the instrument. Accordingly, a study was undertaken to assess the level of correlation between the Ability Measure and educational achievement, a criterion known to correlate strongly with general cognitive ability.

The educational qualification of 503 participants was rated on a 6 point-scale, ranging from no formal qualifications (rating of 1) through to a doctoral qualification (rating 6). Of the 503 initial participants, 30 indicated unspecified professional or vocational qualifications, or other statements or certificates of achievement. Due to difficulty ascertaining the nature of these qualifications, these participants' data were removed from the sample. This left 470 participants, with educational qualifications rated on the scale presented in Table 11 below.

Educational Qualification	Rating
Doctorate or PhD	6
Masters or postgraduate (with or without honours)	5
Bachelors or postgraduate (with or without honours)	4
Trade certificate	3
Secondary school qualifications	2
No formal qualifications	1

Table 11: Classification system for educational achievement

The correlation between educational qualification and scores on the *Overall Ability Measure* scale and ability subscales was then calculated. Because of the multiple comparisons being made, Bonferoni adjustments were made to keep the overall significance level at a 0.05 level of significance. All correlations presented in the table below are significant at $p = 0.0125$.

Scale	Correlation	Corrected
Overall Reasoning Aptitude	0.35	0.39
Verbal Reasoning	0.29	0.38
Numerical Reasoning	0.18	0.22
Logical Reasoning	0.33	0.40

n = 470, all correlations significant at p < .01

Table 12 Correlations between the Ability Measure and educational achievement

The correlations in the corrected column of Table 12 have been corrected for unreliability. The correlation of 0.39 between *Overall Reasoning Aptitude* and educational qualification provides strong support for the predictive validity of the *Overall Reasoning Aptitude* scale. The correlations between the subscales also provide sound evidence of the practical significance of scoring well in the Ability measure. The benchmark for such coefficients to be of practical significance is 0.3 (Kline 2000).

Personal Styles and Work Preferences

In deciding what other tests to correlate the eP2 scales with, we focused on ensuring that the scales used as baseline measures had construct level meaning. In much personality research, disagreement between scales scores on two tests is the result of a scale name that does not reflect the underlying dimensions being assessed. Indeed, this was much of the reason behind the early inconclusiveness of studies into the effectiveness of personality testing for the prediction of job performance in personnel selection situations.

Following the work of Hunter and Schmidt (1998) and more recently Barrick, Mount & Judge (2002) there is general agreement that the big 5 personality dimensions can predict job performance. The strongest predictor of performance is found to be conscientiousness; followed by emotional stability. The remaining personality dimensions of extroversion, agreeableness and openness are found to predict well, but for more specific measures of job performance than overall performance (IPIP, 2001).

For this reason we selected the 50 item big-5 factor markers from the International Personality Project, which can be found on the web at http://ipip.ori.org/ipip/new_home.htm. This site contains the items that make up the marker factors used in the current study, their scale characteristics, and correlations with tests such as the NEO of Costa and McCrae, one of the most widely known and replicated classification frameworks of personality in the world today. The table below contains our hypothesized direction of correlations that would exist between eP2's scales and the big 5 factors.

Hypothesized Correlations

eP2 Scales	Hypothesised Correlations with Big 5
Extroversion	Extroversion +
Orderliness	Conscientiousness +
Openness to Exp.	Openness +
Tolerance	Agreeableness +, Emotional Stability +
Self-Confidence	Emotional Stability +
Teamwork	Extroversion +
Competitiveness	Extroversion +
Physical	Openness -
Predictability	Neuroticism +
Pressure	Conscientiousness +
Autonomy	Conscientiousness + openness+
Complexity	Openness +
Interaction	Extroversion + Agreeableness +
Numerical	Openness +
Verbal	Openness +
Logical	Openness +
Overall Reasoning	Openness +

Table 13: Hypothesized correlations with Big 5 Marker Factors

Actual Correlations

Table 14 below presents the correlations observed between the eP2 scales and the big-five marker factors. Correlations approaching one, even prior to the standard correction for scale unreliability, indicate strong relationships between the two scale sets.

eP2 Scales	Correlation	Corrected	Big Five
Extroversion	0.89	1.00	Extroversion
Orderliness	0.81	1.00	Conscientiousness
Openness to Exp.	0.64	0.78	Openness
Tolerance	0.28	0.35	Agreeableness
Tolerance	0.58	0.70	Emotional Stability
Self-Confidence	0.39	0.51	Emotional Stability
Teamwork	0.42	0.50	Extroversion
Competitiveness	0.31	0.37	Extroversion
<i>Physical</i>	<i>-0.01</i>	<i>-0.01</i>	<i>Openness</i>
<i>Predictability</i>	<i>-0.09</i>	<i>-0.11</i>	<i>Emotional Stability</i>
<i>Pressure</i>	<i>0.00</i>	<i>0.00</i>	<i>Conscientiousness</i>
<i>Autonomy</i>	<i>0.04</i>	<i>0.05</i>	<i>Conscientiousness</i>
Autonomy	0.44	0.57	Openness
Complexity	0.32	0.39	Openness
Interaction	0.55	0.64	Extroversion
Interaction	0.40	0.48	Agreeableness
<i>Numerical</i>	<i>0.10</i>	<i>0.13</i>	<i>Openness</i>
Verbal	0.30	0.39	Openness
<i>Logical</i>	<i>0.17</i>	<i>0.22</i>	<i>Openness</i>
Overall Reasoning	0.24	0.29	Openness

n =165, all correlations significant at $p < .01$ unless italicized

Table 14: Observed correlations with Big 5 marker factors

The majority of the hypothesized correlations between eP2's scales and the big 5 factor markers presented in Table 13 were found to exist in the hypothesized direction. Table 14 above contains the actual and corrected correlations between eP2's scales and corresponding big 5 factor markers.

Clearly there is a strong relationship between the big 5 factor markers and the eP2 scales. In particular, conscientiousness and extroversion stand out as

being strongly related between the two scale sets. Research literature indicates conscientiousness and emotional stability are the most consistent predictors of subsequent job performance; followed by extroversion, openness and agreeableness with regard to more specific criteria. Table 14 above indicates that eP2's scales measure similar constructs to the five-factor model of personality, providing strong support for the use of eP2 in selection and recruitment settings. More specifically, strong correlations were found to exist between extroversion, conscientiousness, openness and emotional stability.

The hypothesized negative correlation between the *Physical* scale and Openness was not observed, indicating these two scales are measuring different factors. Similarly, hypothesized correlations between the *Pressure* and *Autonomy* scales and the big 5 factor of conscientiousness were not observed in the study. Moderate correlations were found between the *Overall Reasoning Aptitude* and reasoning subscales and openness, a factor commonly found to correlate with intelligence.

Criterion Related Validity: In the Workplace

Expected Results

Given the known relationship between performance and cognitive ability (e.g. Schmidt & Hunter, 1998), we expected to observe strong correlations between the eP2 ability measure and the ability subscales and both (a) manager ratings of performance against corporate competencies and (b) managerial ratings of behavioural traits (on 5 point versions of the eP2 scales), to the extent that these traits are job relevant.

Further, we expected to observe correlations above 0.2 between ability and job performance. The table below, taken from the O-net Testing and Assessment Guidelines, interprets the meaning of various value ranges of the validity coefficient, and highlights the rationale for the target value of an r-value of 0.2.

Validity Coefficient	Value Interpretation
> 0.35	Very Beneficial
0.20 – 0.35	Likely to be Useful
0.11 – 0.20	Depends on Circumstances
< 0.11	Unlikely to be Useful

Table 15: O-Net Interpretations of Validity Coefficient Size

It is important to keep in mind that significance is a function of the size of the effect and the sample size. The minimum significant correlation detectable for example, with a sample size of 65 (the size of the sample in this study) is 0.24 at $p < .05$, and .31 at $p < .01$. With this in mind we have drawn attention to significant relationships as well as substantive non-significant correlations that subject numbers did not permit the detection of correlation at a significant p level.

Study One:

Relationships between eP2 and Performance in a call centre environment.

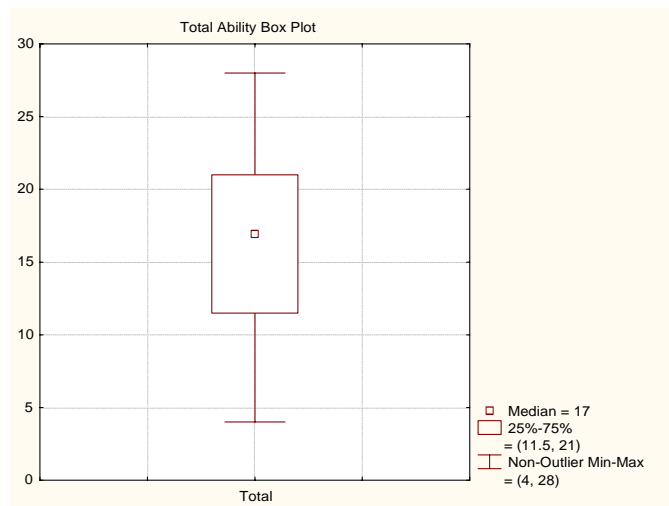
The difficulty of the eP2 ability measure for call centre applicants

Analysis indicated the level of difficulty of the eP2 ability measure is suitable for call centre selection. In particular, evidence that eP2 ability measures are of an appropriate level come from three sources.

- There are no floor or ceiling effects in the data for overall ability (nobody scored zero out of 30 and nobody scored 30 out of 30).
- The measures of central tendency are all very close to the centre of the possible range (i.e. 15 out of 30). There was a mean of 16.57, and a median of 17. While there are multiple modes, these are all presently above the mean, and we expect the mode to converge on the mean as the number of subjects who complete the test increases;
- Subscale analysis indicates that the means of numerical and verbal reasoning are both slightly above the centre of the possible range of scores, with just logical having a mean slightly below the middle of the range.

	N	Average	Median	Minimum	Maximum	Std dev
Total	56	16.57	17.00	4	28	5.57
Numerical	56	6.86	7.50	1	10	2.67
Verbal	56	5.82	6.00	1	10	2.10
Logical	56	3.89	4.00	0	9	2.01

Table 16 Descriptive Statistics for Ability Scores of Call Centre Staff



Box plot of total ability scores from a call centre environment

Relationships between call centre performance and eP2

Criterion	Best Predictor	r	p
Average Handling Time	Orderliness	-0.28	0.08
Call Coaching	Physical	-0.44	0.00
Communication	Predictability	-0.30	0.02
Customer Service	Predictability	-0.23	0.08
Detail Consciousness	Verbal	0.25	0.06
Flexibility	Work Pressure	0.29	0.03
Self Management	Work Pressure	0.20	0.13
Tardiness	Total Ability	0.29	0.03
Total Rating**	Predictability	-0.30	0.03
Leave Behaviour***	Interaction	0.20	0.14

Table 17: Relationships between performance criteria and eP2 scale

Average Handling Time

Orderliness is negatively correlated with call handling times ($r = -0.28$, $p=0.08$). Orderliness measures the need for order and structure. It embodies reliability, responsibility, conscientiousness and constraint. The higher the individual's orderliness score the more likely they are to have shorter handling times. All other things being equal, in particular, provided there is no problem with issue resolution, low call times are desirable. We took three measures of call handling time, the average for each of the three previous months. We could have used each of the three months as the criterion, or an aggregate such as the mean, or the best of the three. The relationships below are presented for the best of the three months. However, similar relationships are found for the remaining individual months and for the mean average handling time across three months.

Call Coaching

There is strong evidence that scores on the physical scale is negatively related to call coaching scores ($r = -0.44$, $p=0.00$). The Physical scale assesses the importance of working outside and being involved in physical work. As is the case with the handling time criterion, we have used the best coaching score over the three months as the criterion. However, similar patterns are evident across the three individual months and the average of the three months.

Leave and Absenteeism Data

We had a number of performance indicators in this area. The Interaction scale was positively associated with leave behaviour (the strongest relationship was $r=0.20$, $p = 0.14$). The Interaction scale assesses the importance of interacting with and helping others in the work environment. Observed correlations against leave data fall into the category of substantive but non-significant, that is, the sample size was inadequate to detect correlations of this magnitude at a significance level of 5%. However, we can note that across three of the four measures of leave behaviour, the Interaction demonstrated substantive correlations.

eP2-Contact center competency relationships

Communication is conveying to, seeking & receiving from others, information in a clear, positive & sensitive manner. The eP2 scale of Predictability is negatively correlated with communication ($r=-0.30$, $p =$

0.02). The Predictability scale measures the importance of working in a stable, supportive, well-organised workplace with secure employment.

Customer Service Commitment is discovering and meeting internal and external customer's needs and offering fit for purpose solutions. The eP2 scale of Predictability is negatively correlated with customer service commitment ($r=-0.23$, $p=0.08$). The Predictability scale measures the importance of working in a stable, supportive, well-organised workplace with secure employment.

Detail Consciousness is attending to the detail and order/correct procedure of a task and ensuring accuracy & completion. The eP2 scale of Verbal Reasoning is positively correlated with Team Leader Detail Consciousness ratings ($r=0.25$, $p=0.06$). Verbal reasoning is a measure of the level of competency a person has with written language, spelling and meaning of words.

Flexibility is being prepared to modify your thinking and supporting the business by accepting changes. The eP2 scale work pressure scale is correlated positively with flexibility ($r=0.29$, $p = 0.03$), indicating the higher the work pressure score, the higher the managerial flexibility rating. The Pressure scale assesses the importance of doing work that requires a high level of effort and commitment.

Self-Management is about knowing, anticipating and managing personal behaviour and emotions in all situations. The eP2 Work Pressure scale is correlated positively with managerial ratings of Self Management ($r=-0.20$, $p = 0.13$). The work pressure scale assesses the importance of doing work that requires a high level of effort and commitment.

Tardiness is being late for work or back from a break. The Verbal, Numerical and Logical scales were all positively correlated with tardiness, and as a result, so is overall ability ($r=0.29$, $p=0.03$). The higher your total ability scores the higher your tardiness rating (and the less tardy you are!).

Total rating is a composite that was the sum of all of the individual competency ratings. Predictability correlated negatively with this rating ($r=-0.30$, $p = 0.03$). The Predictability scale measures the importance of working in a stable, supportive, well-organised workplace with secure employment.

Employees who perform well in call center environments score higher on ability scales, are more orderly, enjoy work that requires effort & commitment, have a lower preference for outdoor physical work, and a show lower need for stability, security and predictability at work.

Study Two: Job Performance

An analysis was conducted of the job descriptions of 65 staff from 9 job categories in a contact center environment of a large corporation. Key performance indicators and key job responsibilities were derived from each and analysed by industrial and organisational psychologists. The results were grouped under 13 competency headings.

Five-point behaviourally anchored rating scales were then developed for each of the 13 competencies identified in table x below. The employees were rated by their immediate managers on the five point scales with regard to how they performed against each of the 13 competencies. The competencies identified through the job description analysis and against which staff members were rated are presented in table 15 below.

Corporate Competencies
Total Performance
Promoting the corporate brand
Being supportive of other team members
Maintaining an achievement orientation
Being sales focused
Providing outstanding customer service
Accuracy
Efficiency
Timeliness
Problem solving ability
Negotiation skills
Business understanding
Proactive & uses initiative
Conscientiousness

Table 18: Corporate competencies for Criterion Related Validation

Spearman rank order correlations were computed between the manager ratings of job performance on the 13 competencies and employee profiles on eP2. The results are presented and discussed in the tables below.

Competency	Best Predictor	r	p
Total Performance	Need for Interaction	-0.18	0.16
Promoting the Corporate Brand	Physical	-0.34	0.01
Collegial Support	Tolerance	-0.19	0.13
Achievement Orientation	Numerical	0.25	0.04
Customer Service	Physical	-0.24	0.05
Accuracy	Self Confidence	0.22	0.07
Efficiency	Interaction	-0.20	0.11
Timeliness	Physical	-0.22	0.08
Problem Solving	Logical	0.35	0.00
Negotiation Skills	Logical	0.20	0.11
Business Understanding	Total Ability	0.34	0.01
Proactive	Physical	-0.36	0.00
Conscientiousness	Physical	-0.20	0.12

Table 19: Correlations between scale scores and corporate competencies

Study Two: Behavioural Characteristics

This study investigated whether behavioural patterns identified on eP2 were identifiable by managers. Managers rated these workers on each of the eP2 constructs using corresponding five point likert scales. For example, a five point rating scale was constructed for extroversion, and managers rated staff on extroversion, one of the seven scales in the My Personal Styles section of eP2. Spearman rank order correlations were calculated between the management ratings and staff scores on the respective eP2 scales.

Spearman rank order correlations were computed between the manager ratings of job performance on the 13 competencies and employee profiles on

eP2. The results are presented and discussed in the tables below. What we expected to see are positive

Scale	R	N	p
Extroversion	0.33	64	0.01
Orderliness	0.12	64	0.34
Openness	0.11	64	0.37
Teamwork	0.14	64	0.28
Tolerance	0.20	64	0.12
Competitiveness	0.20	64	0.11
Self Confidence	-0.06	64	0.62
Physical	0.22	63	0.09
Security	-0.02	64	0.89
Work Pressure	-0.01	64	0.96
Job Autonomy	0.03	64	0.82
Work Complexity	0.02	64	0.90
Interaction	0.07	64	0.58

Table 20: Correlations between scale scores and managerial behaviour rating

Normative Base

Norms are comparison groups by way of which we interpret a score. They are important because the nature of psychometric assessment means that we have no way of interpreting what a score on a test scale means without reference to a comparison group. For example, what does a score of 15 out of 30 mean with regard to ability? What does a score of 0 or 30 mean? By itself such data is of limited use. We give them meaning by saying what score represents relative to other people who have completed the test.

These comparison groups are known as norm groups. Clearly it is more meaningful when individuals are compared against norm groups comprised of individuals of similar ability to the individual completing the test. For example, you would not compare an engineer against a norm group of school children. It would, however, be very useful to compare the score of an engineer against other tertiary qualified individuals. It would be even more useful to compare an engineer against a norm group of engineers. Selector Assessment Manager (SAM) has the functionality to facilitate this process.

In assessing the suitability of the norm group, take into account the similarity between the individual's background and the backgrounds of those in the comparison group, and, more importantly, consider whether the background of the individuals comprising the norm group is similar to the background of people in similar roles to the role in question.

The SAM (Selector Assessment Manager) system is a web-based system that allows you to collect and maintain your own norms, which will help ensure they are as representative as possible.

Norm Characteristics

The default norm group in the SAM system consists of 755 individuals who completed eP2 for job applications or organizational development over the period March 2002 to March 2003. Of the 755 individuals 332 (44%) were male and 423 (56%) were female; 302 (40%) had a highest qualification at tertiary level, and 453 (60%) had a highest qualification was lower than tertiary level. These data were collected over the period March 2002 to March 2003.

Development sample

The sample used in the development of eP2 was comprised of 503 participants. The factor tables presented in this manual are from the development sample. Of the 503 participants, 267 were male and 236 were female. English was the first language of 96% of the sample. The average age of the sample was 37.

References

- Barrick, M. R., Mount, M. K. & Judge T, A. (2001) Personality and Performance at the Beginning Of The New Millennium: What Do We Know And Where Do We Go Next? *International Journal of Selection and Assessment*, 9, 9-30.
- International Personality Item Pool (2001). A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences (<http://ipip.ori.org/>). Internet Web Site.
- Kline, P. (2000) *A Psychometrics Primer*. Free Assn Books.
- Schmidt, F. L. & Hunter, J. E. (1998) The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, Vol. 124, No. 2, 262-274.