OVERVIEW OF TECHNICAL MANUALS FOR THE NEW KORN FERRY ASSESSMENT SOLUTION

The Korn Ferry Assessment Solution (KfAS) offers a new and innovative process for assessing talent. Deployed on a technology platform that enables client self-service, the KfAS shifts the focus away from specific assessment products to solutions suitable for applications across the talent life cycle. Whether a need pertains to talent acquisition or talent management, the participant will experience a single, seamless assessment process. This process has three elements: a success profile, an assessment experience, and results reporting tailored to a variety of talent acquisition and talent management uses.

The success profile provides a research-based definition of “what good looks like” in a given role. Specifically, the success profile outlines the unique combination of capabilities, including competencies, traits, drivers, and cognitive abilities, that are associated with success in the role. These components are used to inform both the assessment experience and results reporting, which differ according to the solution, for both talent acquisition and talent management.

Whereas the KfAS is new, the assessment components are carried over from legacy Korn Ferry assessment products. The science, research, and psychometric-based information that are the foundation of these robust assessments remain relevant. Therefore, while we work to consolidate and refine technical manuals for the KfAS, we can use the existing technical manuals for KF4D Enterprise, Dimensions-KFLA, Aspects, and Elements as a bridge.

CONTINUES ON NEXT PAGE
CONTINUED FROM PREVIOUS PAGE

This technical manual provides background research and psychometric-based information regarding some of the assessments used in the KFAS. The following table summarizes the solutions available for talent acquisition offerings and where to find the relevant psychometric-based information in the technical manual. We would like to remind you that, in the talent acquisition context and absent a client-specific validation study, the assessment should always be used with other information (e.g., from interviews and résumés) to guide talent decisions.

The Korn Ferry Assessment Solution: Talent Acquisition

<table>
<thead>
<tr>
<th>SOLUTION/NEED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ENTRY LEVEL</strong></td>
</tr>
<tr>
<td><strong>Types of roles in solution offering</strong></td>
</tr>
<tr>
<td><strong>Assessment objects available</strong></td>
</tr>
<tr>
<td><strong>Technical manuals to reference</strong></td>
</tr>
</tbody>
</table>

The KFAS may also be used for a variety of talent management applications such as high potential identification, leadership development/succession, leadership selection, and professional development. The self-assessment components within these solutions include competencies, traits, and drivers, which may be used for diverse reporting needs, including learning agility and risk factors results. Technical information for these components is covered in the KF4D Enterprise technical manual. Multi-rater assessments may also be a part of talent management solutions.
FOREWORD

This Psychometric Review provides information relating to the development and use of Elements (Numerical, Verbal and Logical).

The information included focuses on the initial and ongoing development of the assessments, the background and rationale for their development, and their psychometric properties (for example, reliability, validity and group differences).

You will see how the assessments were developed and how they are unique to Korn Ferry. You will also learn more about the underlying models and their associated norm groups.
## CONTENTS

OVERVIEW OF TECHNICAL MANUALS FOR THE NEW KORN FERRY ASSESSMENT SOLUTION  
SECTION 1  
Background to Elements  
SECTION 2  
Rationale for Elements  
SECTION 3  
Developing Elements  
SECTION 4  
Elements and other ability tests  
SECTION 5  
Norms for Elements  
SECTION 6  
Reporting  
SECTION 7  
Reliability  
SECTION 8  
Validity  
SECTION 9  
Relationships between Elements and biographical data  
SECTION 10  
Completion times  
SECTION 11  
Relationships between Elements and Dimensions  
SECTION 12  
Language availability and cultural differences  
SECTION 13  
Summary  
SECTION 14  
Glossary  
SECTION 15  
References
BACKGROUND TO ELEMENTS

Elements is an innovative range of online adaptive ability tests measuring numerical, verbal and logical reasoning ability.

Elements adapts to the candidate's performance in real time affording a concise but rigorous assessment experience. Elements was developed by a team under the supervision of Roger Holdsworth. Roger was the main author of many other ability tests, namely SHL’s test of Verbal and Numerical Critical Reasoning (1987 onwards), which are used across many countries and industries. The first version of Elements was developed in 2007 and, following initial trials, the tests were revised later that year.

In the early years of psychometrics, beginning in the 1910s with work on IQ at Stanford University and the subsequent development of the American Army Alpha, the primary focus was on assessing abilities. As these methods developed and in particular with the application of factor analysis in the 1930s, interest in measuring a broader range of outcomes developed, with the widespread use of psychological testing and assessment centres deployed to support mobilisation during the Second World War.

Following this, the use of psychometrics to assess abilities has expanded considerably and has been intensively employed in high-volume assessment situations in particular in large organisations, for example, to support graduate recruitment.

Our approach to ability testing is firmly grounded in psychometric theory, and the Elements tests are designed to provide standardised, reliable, valid and fair quantitative measurement of abilities. Guilford's work in the 1960s has shaped ability testing generally and underpins our approach. His model proposed a non-hierarchical approach with specific abilities in particular areas such as verbal, numerical or mechanical reasoning. The model of specific abilities has been further expanded by Sternberg's triarchic model of intelligence (1985) and Gardner (1983) in his model of Multiple Intelligences. Subsequent meta-analytical validation of ability tests has demonstrated their efficacy in predicting job performance from frontline roles through to senior management positions (Schmidt and Hunter, 2004).

Elements tests how well a candidate can perform when they are trying their hardest (rather than measuring typical performance). The assessments also operate similarly to other cognitive ability tests in that they are norm-referenced, so results gain practical meaning and utility once they have been compared with an appropriate norm group (e.g. managerial and professional roles).
However, Elements is significantly different from most other ability tests in that all three tests are adaptive. Many ability tests were originally designed to be paper-based, with everyone being presented with the same set of questions which progressively increase in difficulty. Candidates would be given a set amount of time, usually about 30 minutes, to complete the tests. The number of questions the candidate gets right within the overall time limit then determines their raw score.

With the advent of the internet, it is now possible to take a more intelligent, efficient approach to test development known as adaptive testing. Within an adaptive test there is a large bank of questions (e.g. 300 or more) of varying levels of difficulty. The candidate starts the test with a question of medium difficulty. If they answer this question correctly, they are presented with one of a selection of harder questions. Alternatively, if they get a question wrong, the candidate is presented with a slightly easier question. This continues until they have completed a set number of questions and the combination of the right and wrong answers, along with respective difficulty levels of the questions, determines their final score.

The adaptive testing approach provides a number of significant advantages over traditional testing:

- **Secure**: The adaptive nature of the test means that individuals get questions which are unique to their level of ability, and which are randomly selected from a large bank of possible questions. This reduces the chance of questions and correct answers ‘leaking out’ and being published on the internet. The questions are randomised, which further reduces the chance that any two candidates will receive the same test. Verification tests are available for all candidates, which are short follow-up assessments completed under controlled conditions. The verification tests commence at the level an individual attained in the main test, and provide feedback concerning the similarity in scores between the two testing sessions. All of these features greatly improve the security of the test and the confidence clients can have in the results.

- **Fast**: Because candidates are quickly taken to questions appropriate to their level of ability, the testing process can be completed much more quickly (e.g. half the time) compared with traditional tests, with no compromise on scientific rigour.

- **One test system**: A much greater spread of questions (e.g. suitable for frontline roles all the way up to senior management) can be included in one testing system, removing the requirement for different tests for different roles. This also means that the tests are fair, with no assumptions being made regarding an individual’s education or work experience.

All three Elements tests (verbal, numerical and logical) are adaptive. They take a maximum of only 16 minutes per test to complete, half the time of equivalent traditional tests. A candidate’s score is then compared with different groups (e.g. senior managers, graduate, frontline candidates) to assess how well, relatively, a candidate has done on the assessment.
<table>
<thead>
<tr>
<th>Elements test</th>
<th>Ability measured</th>
<th>Average completion time (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical</td>
<td>Measures ability to reason with numerical information and data. Important for working with budgets or analysing business information. The questions contained within the item bank cover a wide range of calculations ranging from addition and subtraction, through to more complex percentage and ratio calculations. The complexity of the questions and their associated tables and charts also varies significantly. They range from simple tables through to a number of tables, with calculations frequently relying upon information taken from more than one table.</td>
<td>12 Minutes</td>
</tr>
<tr>
<td>Verbal</td>
<td>Measures ability to interpret verbal information and reach correct conclusions. Important for understanding business information and report writing. The questions contained within the item bank cover a range of question types. This includes questions requiring relatively simple comprehension capability through to more complex deduction and induction.</td>
<td>12 Minutes</td>
</tr>
<tr>
<td>Logical</td>
<td>Measures ability to analyse abstract information and apply this to determine outcomes and patterns – such as completing sequences of shapes. Used in a variety of roles requiring complex problem-solving, e.g. IT, engineering and scientists, as well as ‘culture free’ input to measuring potential. The range of questions within the item bank ensures that people at all levels have an opportunity to be challenged and to be able to respond positively to them. They range from the identification of simple patterns through to candidates being able to identify multiple rotations at varying degrees.</td>
<td>9 Minutes</td>
</tr>
</tbody>
</table>

Cognitive ability tests have consistently been shown to be strong predictors of job performance in a wide range of validity studies (Schmidt and Hunter, 2004).

The Elements verbal and numerical reasoning tests measure important workplace abilities: for example, interpreting and analysing written and financial reports, making deductions and inferences from information provided, and understanding and working with a variety of written and numerical information.

Logical reasoning assesses abstract problem-solving ability, which evidence suggests is also predictive of performance at work. It is particularly relevant for roles which are highly abstract and systematic, such as IT and engineering. Logical reasoning tests are somewhat different to verbal or numerical tests because the information analysed can be interpreted without reliance on other abilities, such as language or cultural understanding. In this way they provide a relatively ‘pure’ measure of reasoning ability that is independent of prior learning (e.g. language or mathematical skills). As such they are also helpful for assessing potential due to being less influenced by an individual’s educational and cultural background.
SECTION 2

RATIONALE FOR ELEMENTS

Elements is an online ability testing system which measures verbal, numerical and logical reasoning abilities.

It is also designed to be used across sectors, functions and, in particular, organisational levels. The design of the system draws on existing theories relating to ability testing, including classical test theory and item response theory. In developing the Elements tests, four key goals were kept in mind:

■ To develop a reliable and quick testing system for abilities using adaptive testing.
■ To ensure the system is secure, minimising opportunities for cheating.
■ To provide one testing system applicable across role levels, from frontline or customer service roles up to senior management - this is advantageous as it allows candidates to demonstrate their full ability (rather than applying a ceiling to the level of question difficulty that can be reached by pre-judging which version of a test to administer).
■ To ensure the system is both reliable and valid.
■ In the contemporary era of online assessment, employers are typically looking to balance the need for robust assessment with the desire for rapid and scaleable assessment of candidates. The latter is particularly an issue given the costs of attracting as well as selecting candidates. Additionally, employers look to ensure testing supports the employer brand, particularly given that with the growth of online testing, candidates are often completing assessments remotely. In this context, expecting candidates to sit lengthy tests is increasingly unrealistic and risks acting as a deterrent to applicants.

UNDERSTANDING THE STRATIFIED ADAPTIVE NATURE OF ELEMENTS

The development of Elements was focused very strongly on the need to balance robustness with speed to deliver a practical and effective solution.

■ In terms of test design, adaptive tests such as Elements are significantly different from conventional approaches to testing. In conventional tests, all candidates are presented with the same questions, normally in an ascending order of difficulty. In adaptive tests, the questions presented to a candidate depend on how the candidate scored on previous questions. Each time a candidate gets a question right within the time available, they are moved on to a more difficult one. If they get a question wrong, or fail to answer, they move to an easier question.
The diagram below illustrates how a candidate moves through an assessment based on the previous answers given.

**Step 1: Moving through one block of questions to another**

Elements also involves some practical differences from other adaptive tests. One approach to adaptive testing is to give a candidate a completely new question with new supporting information each time. However, this is in fact very time-consuming as each time a candidate is presented with a new question, they would also need to read a new passage of text, or review a new table of data. Conventional tests typically get round this challenge by presenting three or four questions for each ‘block’ of information (e.g. a passage or a table of numerical information) to save time.

In order to shorten the time required, Elements also works in small blocks of questions. For instance, at the beginning of the test the first question presented will relate to a particular table of information, such as a table of business-relevant numerical data. On completing the first question, if answered correctly, the candidate moves on to a harder question which still relates to the same table of information. They then get two further questions depending on whether their answer is right or wrong. Conversely, a similar logic applies if they get question one wrong and proceed to an easier question. Performance across a block of questions then determines the next question block which is presented to the candidate.

The diagram on the next page illustrates how a candidate might navigate a range of possible blocks of questions.
Step 2: Moving through a complete test

Within each block (where a candidate will complete three of the six available questions), the questions have item ‘difficulty levels’ which are used to determine the score attained so far in the test.

**SCORING ELEMENTS**

A candidate’s score on any passage or table is the sum of the difficulty levels of all questions that they got right, or zero if they got none right. In terms of how the tests work, getting a question wrong includes being timed out.

‘Difficulty level’ is the reverse of the percentage of candidates who got the question right in the trial versions of Elements: if 75% of candidates got a question right, that question’s difficulty level is 0.25 (25%). For any passage or table there are eight possible scores.

The aggregate score so far (based on difficulty levels of questions answered correctly) then determines the choice of the next block of questions from those within the Elements question database. This process continues to the end of the test.

The addition of time limitations on individual questions and a large number of response options per question means that these aspects are contained and controlled, allowing this approach to deliver reliable and valid measurement across a bank of questions with a wide span of difficulty levels. Item response theory approaches to adaptive testing include multiple parameters for assessing question discrimination, one of which relates to the probability of guessing. Candidates guessing answers within ability testing can have an effect on results and negatively influence reliability. Elements has therefore been designed to try to minimise this impact. Specifically, a large number of response options have been incorporated for each question so that the probability of candidates guessing correctly is much lower than in other tests. This is very helpful in minimising this problem and containing its impact on test effectiveness.
A candidate’s final score is the sum of the difficulty levels of all questions correctly answered. The more questions a candidate answers correctly, the higher the difficulty level of further questions, and thus the greater opportunity they have to achieve a very high total score.

All the questions are based on a passage of text (Elements Verbal) or on a table of statistical data (Elements Numerical). Each passage and each table has six possible questions, but any candidate is only presented with three of these. This also applies in principle to Elements Logical. Elements can be described as ‘stratified adaptive’ tests. In technical terms, this is an adapted or stratified version of a one-parameter IRT (item response theory) scoring approach known as Rasch scaling. Candidates move adaptively (according to whether they answer the question correctly, as explained above) from question to question within a passage or table, until they have been presented with three questions from that passage or table. They are then moved to another table or passage, according to their cumulative score at the time.

Based on our statistics, we can use the final score to determine a candidate’s percentile score in comparison with any given norm group. This can be our total database; high-potential norm; university graduates, managers and professionals; non-graduates such as customer service and administrative roles; or any group we are able to define. Because Korn Ferry Assessment Systems record all details of a candidate’s responses, and also their replies to background data questions, we can calculate many different norm statistics post hoc. As such we can create specific norm groups relevant to a particular population based on a client need.

For any candidate, it is also possible to conduct a seven-minute verification test. This is often used where candidates complete the main test unsupervised and remotely, e.g. from home. A verification test can be used to complete a short follow-up assessment on candidates who are brought forward to the next stage (such as an assessment centre) under controlled conditions on the client’s premises. The verification tests will commence at the level an individual attained in the main test, and provide feedback concerning the similarity in scores between the two testing sessions. For many clients, simply the possibility of conducting verification is very helpful as the risk of discovery strongly deters candidates who may be tempted to cheat, even if only a random sample of candidates actually undergo verification testing.

**SPEED OF TESTING**

Elements provides some key advantages in terms of ease of completion and test duration. Because the tests operate in blocks of three questions, it is not necessary for individuals to familiarise themselves with new background information for every question, enabling them to progress more quickly. Additionally, due to the use of adaptive testing, it is possible to test candidates reliably within a maximum of 16 minutes on each test, around half the time of traditional tests.

**RANDOMISATION**

As well as the adaptive nature of Elements, the tests are randomised with a wide selection of questions available for the system to choose from at each level of difficulty.

For instance, there are three different starter blocks for each test as well as a wide range of subsequent options available. This has significant benefits in terms of reducing the opportunity for cheating because there are so many different questions. It would be futile for candidates to share screenshots of their tests as every candidate is presented with a different selection of questions. This in combination with the verification testing approach outlined earlier ensures the tests are highly secure and significantly curtails the possibility of cheating by some candidates, even in the context of remote testing.

**APPLICABILITY ACROSS DIFFERENT ROLE LEVELS**

A final key benefit of the system is that due to the wide range of questions within each test, there is no need to decide which level of test to administer in advance as is the case with classical tests.

Rather, a candidate can simply start to complete Elements Numerical, for example, and the system will home in on the candidate’s ability level. This means the system is applicable across role levels, and the administrator can simply choose which norm group to compare the results with after a candidate has been tested.
SECTION 3

DEVELOPING ELEMENTS

The Elements tests were developed using a number of guiding principles and methods.

As previously mentioned, the idea was to create a testing system that would minimise opportunities for cheating and remove the need for a library of tests, instead enabling test users to have a one size fits all solution for everyone ranging from supervisors upwards. The final assessment also had to be cutting edge, innovative and relevant in the workplace.

The development of the Elements item banks came as the result of a number of inputs:

- A review of the existing validation literature to confirm which three areas of ability to focus on.
- A review of materials used across a variety of different workplaces to ensure that the final design was relevant and appropriate for the likely candidate groups.
- A consideration of advances in technology and psychometrics that could be harnessed, adapted and applied in the test system.

This phase of work resulted in a specification for the test system including the adaptive nature of the tests, the required psychometrics and also a detailed specification of exactly how many items at varying levels of difficulty and content type would be needed for each category of test. It also included an overview of the required number of answer options for each type of item.

After the inception of the initial design, item writing was shared amongst the development team. Items were then peer reviewed with a number of criteria in mind, namely:

- Appropriate level of difficulty
- Work relevance
- Face relevance
- How up to date they were
- Reading level (using accessible vocabulary)
- Free from ambiguity in passages/tables/items/answer options
- Checking the scoring key
Key international stakeholders were also asked to review the items for ease of translation, ensuring that there were no terms that would be difficult to translate and that the scenarios and content used could be applied across (or at least adapted to suit) a variety of cultures. This obviously led to a number of rewrites prior to the start of the first trial. Of course, as each test type involves different abilities, the checks were adapted to suit. For example, anyone reviewing the Elements Logical test was asked to find the correct answer and document the rules they applied to get to it. This was then compared to the original author’s documented rules about the item as a means of checking for ambiguity. Similarly, the Elements Verbal and Elements Numerical items were initially reviewed without a scoring key to check for any ambiguity amongst the answer options. A wider review was then carried out of the entire bank of items to ensure that there was a clear spread of difficulty and content.

The trial versions of the tests were fixed (i.e. non-adaptive) versions and each one consisted of 30 items. Each version was created with overlapping common items so that each item could be evaluated and calibrated appropriately. This design was optimal for the purposes of constructing the item banks. Trials were conducted on a combination of university students and employees from sponsoring organisations across a variety of sectors and industries, ultimately totalling in the region of 2,000 people.

Once trial data was obtained, the items were screened for acceptance into the item banks using a set of criteria:

- Item discrimination parameters were reviewed with items exhibiting low item-total correlation being rejected.
- Item difficulty parameters were reviewed with items exhibiting extreme values being rejected.
- Item distractors were also reviewed with item distractors exhibiting irregularities being rejected.

Items that did not meet the criteria set were excluded from the item bank. The final item bank incorporated a wide spread of item difficulties from 0.1 to 0.9 with the majority always falling into the 0.2 to 0.7 range. The group differences were in line with expectations and a more complete analysis of differences is provided in the ‘Relationship between Elements and biographical data’ section. The resulting banks of items were subjected to a final peer review for content and context coverage. The final banks of items were then configured to create the adaptive randomised tests.
SECTION 4

ELEMENTS AND OTHER ABILITY TESTS

Elements offers a number of advantages over and above other ability tests.

One of the main differences is that Elements is designed for use across all job levels and all industry sectors and that it uses stratified adaptive testing. Results are available as soon as the candidate has completed the test and are provided to the client in an easy-to-understand report.

Elements can only be completed online. Far from being a disadvantage, this has several benefits. Firstly, it means that we collect every piece of data, rather than having to rely on people sending data to us. We know exactly how each candidate has responded to each individual question, how long they took to respond, and even how many times they changed their mind. We are able to make very small amendments to questions imperceptibly, instead of having to print new editions. It also enables us to use the stratified adaptive testing approach.
SECTION 5

NORMS FOR ELEMENTS

Since any revisions (since 2007) to the content of Elements have been very minor, we are able to take all completions since mid-2008 as the basis for our norm group. The most recent version of the norms (2012) covers more than 145,000 completions of Elements Verbal and almost 110,000 Elements Numerical completions. As Elements Logical was released later, the database for this is smaller but still exceeds 40,000. The data used within the norms includes a variety of project types including those completing Elements in a recruitment, selection, development and restructure context, along with a series of research studies. Whilst it is recognised that this is an ‘opportunity sample’, the data used offers a wide representation from a variety of sectors, functions, age groups and of course a balance of gender.

The following norm groups are available:

- Professional, graduate and managerial (comprises people on our database who meet this description*)
- Customer service, administrative and similar (non-graduate – comprises people on our database who do not have a university degree*)
- High Potential (predominantly candidates from the Finance sector with above average scores on Elements)
- Composite

*These descriptions are intended to be of practical use, although we recognise that some managers and professionals may not have university degrees, and some people in customer service, administrative and similar roles may have degrees. Obviously, people should normally be compared with other applicants for the same type of job rather than with those who have the same educational level.

The sample size for the norm groups can be seen below:

<table>
<thead>
<tr>
<th>Norm group</th>
<th>Verbal</th>
<th>Numerical</th>
<th>Logical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional, graduate and managerial</td>
<td>94,388</td>
<td>83,754</td>
<td>29,921</td>
</tr>
<tr>
<td>Customer service, administrative and similar</td>
<td>26,659</td>
<td>24,502</td>
<td>11,977</td>
</tr>
<tr>
<td>HiPo</td>
<td>2744</td>
<td>2805</td>
<td>430</td>
</tr>
<tr>
<td>Composite</td>
<td>145,916</td>
<td>108,256</td>
<td>41,898</td>
</tr>
</tbody>
</table>

A further breakdown of each group is provided in the Norms and Reliability Supplement (which can be downloaded from the Korn Ferry Learning Centre).
SECTION 6

REPORTING

There are two main reports available: one for candidates and the other for the HR professional, psychologist or hiring manager.

The Elements scores are also included in the role match profile. A set of sample reports is available to download from the Talent Q website (www.talentqgroup.com).
SECTION 7

RELIABILITY

Reliability can be measured in a number of ways, either by looking at the consistency of a test over a period of time (test-retest reliability or parallel form) or by looking at the ‘internal consistency reliability’.

The internal consistency can be measured in a number of ways, the simplest being the split-half reliability (Spearman-Brown split-half method). This adds up the score on the items for half of the test (e.g. all the ‘even’ numbered items) and correlates this with the score on the items for the other half of the test (e.g. all the ‘odd’ numbered items). However, a more sophisticated approach is generally preferred by psychometricians and is usually measured by using Cronbach’s alpha coefficient.

To measure the internal consistency of a test, statisticians have developed complex equations which in practice identify the average level of correlation for all the possible ‘split-half’ combinations that could be made. As with other reliability coefficients, the alpha should typically be 0.7 or higher for any given assessment or scale within an assessment.

TEST-RETEST RELIABILITY

With regard to the test-retest reliability of Elements, a series of projects and research studies have allowed us to collect data on a large number of people who have completed at least one of the Elements tests twice. For all tests the majority of the participants were aged under 40 and most were qualified to degree level or above. The male/female split was similar for Numerical and Logical (roughly 70:30) but for Verbal it was closer to equal proportions of males and females.

The table below provides the correlations between the first and second attempts, the size of each of the groups and the average time (in days) between the two attempts.

<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>Numerical</th>
<th>Logical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.73</td>
<td>0.72</td>
<td>0.76</td>
</tr>
<tr>
<td>Sample size</td>
<td>114</td>
<td>437</td>
<td>409</td>
</tr>
<tr>
<td>Average time between tests</td>
<td>90</td>
<td>153</td>
<td>92</td>
</tr>
</tbody>
</table>
The industry accepted standards for reliability is 0.7, therefore all three tests are displaying good test-retest reliability. However, although these figures are perfectly acceptable there are important considerations in terms of why they may in fact be an underestimation of true reliability:

- As test takers in this research did not see the exact same test at the first and second testing session, the test-retest reliability coefficient can be seen as a conservative estimate since item sampling adds an extra degree of error beyond variation in a test taker’s performance.

- The text books refer to ‘coefficient of stability’ when the same test is given at a later time, say three months later, and ‘coefficient of equivalence’ when a parallel version is given at approximately the same time as the initial version. Because Elements tests are adaptive, the candidate has virtually no chance of being given the same version on the second testing occasion, so the reported correlations are actually coefficients of both stability and equivalence, which represents a much stricter test of their reliability.

- Further, in most published studies, the tests have been given under standardised, supervised conditions, whereas in this case testing was unsupervised and carried out remotely. It could well be that on one or other of the occasions, the candidate was distracted in some way. So, despite our best intentions, we cannot be sure that the same conditions applied to both occasions.

### INTERNAL RELIABILITY: CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

As adaptive tests such as Elements provide each test taker with a different set of questions, methods traditionally used to measure reliability are not appropriate. However, it was possible to consider the internal reliability of the fixed trial versions of the tests as one measure of reliability using Cronbach’s alpha (Cronbach, 1951). Using the trial versions, the alpha coefficient for each test was as follows:

<table>
<thead>
<tr>
<th>Internal consistency reliability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements Verbal</td>
<td>0.83</td>
</tr>
<tr>
<td>Elements Numerical</td>
<td>0.88</td>
</tr>
<tr>
<td>Elements Logical</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Obviously, test takers in this research were experiencing a wider breadth of questions than they actually see in the adaptive Elements tests so the figures above are regarded as a low estimate of the reliability of the shorter, adaptive versions.

The reliability of the adaptive version of the tests can be measured using a different measure of internal consistency known as ‘marginal reliability’. This is based on item response theory (IRT) and offers a better measure of precision (Green, Bock, Humphreys, Linn and Reckase, 1984).alpha (Cronbach, 1951). Using the trial versions, the alpha coefficient for each test was as follows:
In IRT, the reliability of a test varies along the range of the ability continuum. Marginal reliability is effectively the average reliability over the ability continuum. The values for marginal reliability range from 0 to 1 and their interpretation is analogous to Cronbach’s alpha. Marginal reliability provides a single value that summarises a scale’s overall precision (Thissen and Wainer, 2001)

The marginal reliability of each Elements test was estimated using IRT calibration (n=1000). The results are presented below. The results show that Elements ability tests are highly reliable.

<table>
<thead>
<tr>
<th>Internal consistency reliability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements Verbal</td>
<td>0.85</td>
</tr>
<tr>
<td>Elements Numerical</td>
<td>0.81</td>
</tr>
<tr>
<td>Elements Logical</td>
<td>0.82</td>
</tr>
</tbody>
</table>

It is worth noting that the reliability coefficients are equivalent to much longer non-adaptive ability tests used in employment settings.
SECTION 8

VALIDITY

The validity of a test can be measured in a number of different ways, namely face validity, construct validity and criterion validity.

FACE VALIDITY

Elements was developed by experts in the field of psychometrics with the intention that the questions would be as face valid as possible to the person completing the test. All questions are designed to be relevant to today’s working environments, such that test takers will ‘buy into’ the subject matter used rather than being put off by outdated or irrelevant content.

CONSTRUCT VALIDITY

Construct validity is a more theoretical but nevertheless important aspect of validity. A ‘construct’ in psychology is a term meaning a particular psychological characteristic, for instance numerical reasoning ability, or an aspect of personality (e.g. extraversion), values, motivation or performance. Construct validity is concerned with whether a test is measuring the construct it claims to measure. There are two ways of assessing construct validity:

- Convergent validity: if the test can be shown to correlate significantly with other tests measuring similar constructs, then it may be claimed to have convergent validity.
- Discriminant validity: if the test can be shown to not correlate significantly with other tests which are intended to measure different constructs, then discriminant or ‘divergent’ validity can be claimed.

As psychologists believe there is a concept known as general intelligence (or ‘g’), it is safe to assume that an ability test with good construct validity is likely to correlate with other tests (to some degree) due to this underlying factor of ‘g’.

The correlation observed between Elements Verbal and Numerical is 0.44 (based on a sample of more than 30,000 people who had completed both tests in their current versions). The correlation between Elements Verbal and Logical is 0.32 and the correlation between Elements Numerical and Logical is 0.46 (based on a sample of more than 15,000 people who had completed both tests in their current versions). This shows that there is a similar degree of overlap between the three ability assessments, but that verbal, numerical and logical reasoning are nevertheless separate attributes, worthy of being assessed separately.
Research comparing Elements with another similar series of tests also provides evidence of construct validity. 85 participants were asked to complete Elements and SHL Verify Verbal, Numerical and Inductive tests. Using the entire sample, the correlations* were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>0.85</td>
</tr>
<tr>
<td>Numerical</td>
<td>0.81</td>
</tr>
<tr>
<td>Logical/Inductive</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*The correlation coefficients were adjusted for the reliability of the tests.

Clearly, the numerical tests show the highest correlation, with the verbal tests slightly lower. Whilst there are going to be clear expected overlaps between the SHL and our Elements verbal tests, there are differences which will inevitably impact the size of the correlation between them. For example, the SHL tests offer candidates a choice of ‘True, False or Cannot Say’, whereas the Elements tests offer a series of possible responses and the candidate has to choose which is/are correct. Therefore, having a slightly lower correlation is in line with expectations.

Whilst we would also expect a correlation between the Elements Logical test and the SHL inductive test it is no surprise that they demonstrate the lowest correlation. Whilst the underlying constructs being measured are similar, there are differences in the tests. For example, Elements Logical offers more response options to choose from, candidates have more cells to choose between and interpret and also the underlying pattern that candidates need to uncover in the Elements test apply in a multidimensional manner (horizontal and vertical, not just horizontal).

Further evidence of construct validity can be sought by looking at the correlation between ability tests and related personality scales. Elements Verbal has significant positive correlations with Dimensions scales Conceptual and Flexible. Elements Numerical has significant positive correlations with Dimensions scale Analytical. Elements Logical has significant positive correlations with Dimensions scales Analytical (highest) and Flexible.

All of these results support the construct validity of Elements further.

**CRITERION (EMPIRICAL) VALIDITY**

Criterion validity looks at the relationship between the ability test and performance on the job. So far, we have collected nine studies of the empirical validity of Elements.
STUDY 1:

Telecoms company 1

Using the trial versions of Elements, against a ‘hard’ criterion of sales performance among 60 sales representatives in a telecoms company, there was a highly significant correlation with Elements Numerical (0.34). Sales representatives who performed better on Elements had higher sales performance figures. Since other factors than ability obviously influence performance in any job, a correlation of 0.40 is about the limit that a single test might be expected to reach.

STUDY 2:

Telecoms company 2

The sample consisted of 149 people in sales roles in a telecoms company in Asia. There were two main performance criteria for these people (although not everyone had all data). ‘Sales order value’ was available for 90 (of the 149) subjects, ‘revenue’ for 124. Some 14 subjects had neither performance value, 79 of them had both. Both criteria were expressed as percentages of target or budget.

All participants completed the tests in UK English. However, for a number of participants, English was not their first language. To remove the impact of language proficiency on the results, the data set was split into two groups: those who spoke English as their first language and those who did not.

For those people whose first language was not English, no significant relationships were found between the Verbal or Numerical tests and their performance criteria, suggesting that the completion language had affected results.

However, very clear and positive relationships were available for the validation group subjects who did claim English as their first language. We always recommend that people should only complete assessments in a language in which they are totally fluent, therefore the English-speaking subset is the only legitimate group to analyse.

Elements and performance

<table>
<thead>
<tr>
<th></th>
<th>Sale order value</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number in sub-group</td>
<td>47</td>
<td>60</td>
</tr>
<tr>
<td>Elements Verbal</td>
<td>0.66</td>
<td>0.05</td>
</tr>
<tr>
<td>Elements Numerical</td>
<td>0.97</td>
<td>0.33</td>
</tr>
</tbody>
</table>

The correlations for Elements Verbal and Numerical with the sales order value criterion are exceptionally high. These results should be treated with caution and may not be fully replicated in a study with another sample. However, they are very clearly indicative of the value of such tests in predicting sales performance.

Against the revenue criterion, the result for Elements Numerical is exactly at the level to be expected (from the collected body of all validity studies) for a relevant aptitude test when correlated with a legitimate performance criterion. It is worth repeating here that both performance criteria are objective measures so the results can hardly be spurious.
Summary

- The study confirms the value that Elements can have in identifying successful performers in sales roles.
- The study was very useful as we were able to examine the relationships between tests and objective (and highly tangible) criteria of performance.
- The above results held good for the subjects whose first language was English – but not for others, almost certainly because the latter group’s proficiency in English was a confounding variable which masked the participants’ true abilities.

STUDY 3:

Recruitment company

The sample for this study was 36 recruitment consultants based in Ireland. The group was a mixture of temporary/permanent/contract consultants across a mix of specialisms.

All participants completed Elements Verbal, Logical and Numerical and were each assigned a ranking based on their total profit within a fixed period.

This is of course a particularly small sample for a validation study and, as the performance criterion was a ranking, non-parametric correlations were used (Spearman’s rho) resulting in smaller correlation coefficients than one would usually see using parametric methods. However, taking this into account we can see that it still provides an indication of the relationship between the Elements tests and the performance criterion (profit). The correlations are presented below:

<table>
<thead>
<tr>
<th></th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements Verbal</td>
<td>0.11</td>
</tr>
<tr>
<td>Elements Numerical</td>
<td>0.27</td>
</tr>
<tr>
<td>Elements Logical</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*None of the correlations reached significance at the 0.05 level but this comes as no surprise with such a small sample.*

The data indicates that the Elements Numerical test goes some way towards predicting success in the recruitment consultant role. Whilst a higher correlation would give more credibility to this, we have to acknowledge this is a small group and therefore regard the finding with some caution. As highlighted above, numerical reasoning is only one part of the job so we would not expect to see a high correlation; however this is a promising start.
STUDY 4:

Online publishing company

The sample consisted of 58 people working across a variety of professional roles in an online publishing company. They all completed Elements Verbal, Numerical and Logical and were assigned to one of three performance categories, from high to low. This categorisation was based on their annual performance review results combined with a rating from the organisation’s CEO.

Elements Logical demonstrated a significant correlation with the performance measure (0.30, p<.05). This suggests that there is a relationship between someone’s ability to analyse and reason with abstract information, and performance in a variety of roles.

STUDY 5:

Industry standards body

The sample consisted of 94 people working as client managers for the organisation. A number of performance measures were provided but the key one that stood out was the breadth of the employee’s capability, as shown by a measure of the number of areas or ‘products’ in which an employee is trained to assess others. Whilst they completed all three ability tests, it was performance on Elements Verbal that correlated most highly with this measure of performance (0.23, p<.05).

This suggests that the higher someone’s verbal reasoning the more likely they will be to acquire multiple qualifications. A likely explanation here is that this illustrates a greater ability to learn based on higher levels of verbal cognitive ability.

STUDY 6:

Retail & Brewery company

The sample for this study was 34 employees who completed Elements Numerical and Elements Verbal as part of a wider development programme. Within the same programme they also took part in a 360-degree review project where their managers, peers and direct reports rated them on key organisational competencies (which were combined with the employees’ own self-ratings).

The key competencies that showed clear relationships with Elements Verbal and Elements Numerical were those relating to Persuading, Convincing and Influencing Others along with Thinking Commercially. Elements Verbal also correlated significantly with the ability to Deliver More for Customers and finally Elements Numerical with Delivering Results. An overall rating was created (based on the average of all competency ratings) and this showed moderate correlations with both Elements tests.

The results are presented in the table on the next page.
<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>Numerical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persuading, convincing and influencing others</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>Thinking commercially</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Deliver more for customers</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Delivering results</td>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td>Average 360 rating</td>
<td>0.22 (n.s.)</td>
<td>0.27(n.s.)</td>
</tr>
</tbody>
</table>

Unless stated, all correlations are significant at p<.05

This study suggests that Elements Verbal and Elements Numerical can go some way towards predicting overall performance (as measured by a 360-degree review process) and can also be linked with Persuading, Convincing and Influencing Others and Thinking Commercially too. Persuading, Convincing and Influencing Others in this project is measured by items that have an element of persuading using well-reasoned arguments, which explains why there are relatively strong correlations with verbal and numerical reasoning. Further, Thinking Commercially is measured by items that look at someone’s capacity for dealing with complexity, and considering the pros and cons of proposals. Again, it therefore comes as no surprise to see a significant correlation with verbal and numerical reasoning.

There is less clear evidence to explain the relationship between verbal reasoning and Deliver More for Customers or between numerical reasoning and Delivering Results.

No firm conclusions can be drawn from this study as the sample size is too small. However, the results go some way towards providing extra validation evidence for Elements.

**STUDY 7:**

**Automotive manufacturing and retail company -sales roles**

The sample for this study was 94 employees who completed Elements Numerical and Elements Verbal. The performance measure for the project was sales profit over a four month period. The role was a sales-based one working in car showrooms. The sales people’s scores on the Elements Verbal test correlated significantly with sales profit at 0.38. This suggests that the higher someone’s verbal reasoning the more success they have in relation to overall sales profit.
STUDY 8:

Internet bank

Data was collected on 38 employees who were taking part in a high potential scheme for an international bank. All 38 were assigned a rating on their overall performance by their manager and they were also required to complete all three Elements tests (data was only available for 35 people on Elements Logical). None of the tests correlated significantly with the performance rating, which comes as no surprise with such a small sample. The test approaching significance was Elements Numerical, which correlated with the manager’s rating at 0.35. More data is needed before any firm conclusions can be drawn.

STUDY 9:

Care & Social Care company

Within this study there were 102 operational managers working for a care organisation dedicated to helping people with disabilities to retain autonomy and independence. The participants were all given an annual performance review and from that a performance measure was provided that reflected their overall performance. Of the three tests, the one demonstrating the highest correlation was Elements Numerical which correlated at 0.3 (p<.001). This suggests that the better performing operational managers are those with the highest numerical reasoning capability.
SECTION 9

RELATIONSHIPS BETWEEN ELEMENTS AND BIOGRAPHICAL DATA

This part of the review for Elements Verbal is based on data from a sample of more than 32,000 people who completed Elements Verbal and for whom biodata was available. For Elements Numerical the sample was in the region of 30,000. Finally, the Logical sample was in the region of 15,000. As not everyone provided all biodata information, the sample sizes are smaller in some areas.

It should always be remembered that the remarks made in this part of the review relate to the samples concerned and may not be generalisable to other samples.

STATISTICAL NOTE

The respondents’ Elements scores have been compared with age, gender, education, function and ethnic origin. Some biodata variables lend themselves to correlations, others to differences between means. With such a large sample size, any correlation above 0.03 is flagged as statistically significant; therefore it is probably more sensible to consider correlations above 0.1 as the minimum for psychological significance. As regards significant differences between means, this would strictly depend on the sizes of the sub-samples involved; but we have taken 0.25 of a standard deviation (SD) as the critical value. Differences between 0.25 and 0.40 of a SD will be called significant, anything above 0.40 highly significant.
AGE

It is generally expected that reasoning ability (when measured by tests with time limits) will fall off gradually with increased age. This is less likely to be the case with tests such as Elements Verbal and Numerical because they are based on contemporary business-related subject matter.

To comply with UK guidance on collection of biographical data, we ask people to indicate in which age band they are, rather than exact age in years or date of birth. The table below presents the numbers falling into each age category; as can be seen there is a wide range of ages. The test correlating highest with age was Logical (–0.30), followed by Verbal (–0.12) and lastly Numerical (–0.10). Since all of these correlations are negative, these findings are consistent with industry expectations suggesting that reasoning ability decreases with age. Whilst this relationship is strongest for logical reasoning ability, all of these correlations are quite small, indicating that whilst people appear to do less well on ability tests as they get older, this relationship is not particularly strong.

![Age group distribution chart]

EDUCATION

The mean difference between graduates and non-graduates is in the region of 0.07 of a standard deviation for Elements Numerical and 0.18 of a standard deviation for Verbal.

This is a little less than one would expect and we need more data to explore these findings further. Interestingly the non-graduate group performed better on Elements Logical than the graduate group, with the mean difference being 0.15 of a standard deviation. This emphasises the fact that logical ability is less dependent on academic learning than verbal and numerical ability, meaning it may have less adverse impact than other ability tests.
Function

The number of respondents from the various functions for each ability test are as follows:

**Function distribution**

The highest scoring functions on Elements Verbal were legal, general management and marketing; the lowest were administrative/clerical and customer services. However, none of these differences were statistically significant.

The highest scoring function by far on Elements Numerical was finance; the lowest was administrative/clerical.

The highest scoring function on Elements Logical was IT, although this difference was not statistically significant. No group scored significantly lower on Elements Logical.

**ETHNIC ORIGIN**

In this research we had data on ethnic origin for 9,198 people who had completed Elements Verbal in their own language. Data was available for 7,449 people who had completed Elements Numerical in their own language.
These figures are presented below as percentage of overall completions below. The data reported on for Elements Logical includes all completions as this test is by design not language dependent. In this instance we had data on 6,750 people.

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>Elements Verbal</th>
<th>Elements Numerical</th>
<th>Elements Logical</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>78%</td>
<td>77%</td>
<td>67%</td>
</tr>
<tr>
<td>Chinese</td>
<td>7%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>Indian sub continent</td>
<td>10%</td>
<td>9%</td>
<td>14%</td>
</tr>
<tr>
<td>African and Afro Caribbean</td>
<td>5%</td>
<td>16%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Of people taking the Verbal test in their own language, those who said their ethnic group was ‘White’ made up 77%. In the Numerical test, this percentage was also 77%. Across the Logical test completers, the white group made up 67%. As a result, the other ethnic groups were somewhat smaller across all three tests, a factor we should consider in interpreting any statistics.

Some non-white ethnic groups are generally reported as performing less well than whites at tests of verbal and numerical reasoning, and many reasons have been put forward for this. If we find such differences, this does not necessarily mean that the tests are unfair. As far as this review goes, our data set is too small to give us a reliable answer to such questions.

The Chinese sample scored somewhat better than the other groups on the Numerical test, with the African and Caribbean group scoring somewhat less well.

The white sample scored better on Elements Verbal than the other groups, with the African and Caribbean group and the Indian sub-continent group scoring less well.

On Elements Logical, the Chinese group scored better than the other ethnic groups and the African and Caribbean group scored less well.

Generally speaking, in studies of ethnic differences on psychometric tests, it is often observed that the African and Caribbean groups perform less well than the white groups, whilst the east Asian groups frequently outperform whites. The west Asian groups only slightly underperform compared with the white groups generally speaking. As a profession we are seeking explanation for why these differences exist, but it is particularly difficult to tease out factors of educational background (or opportunity), social factors (socio-economic group, parental support, etc) and other cultural differences (level of aspiration, or fear of making mistakes, amongst many others).

In summary, we can say preliminarily that Elements ability tests produce similar patterns of differences between ethnic groups as have been found with other similar assessments. Any significant differences observed are between 0.25–0.5 of a standard deviation and therefore compare favourably to other ability tests.

ADDRESSING GROUP DIFFERENCES

Whilst the group differences outlined within this review are typically under half a standard deviation and therefore unlikely to have much of an impact on selection rates for the various groups, it is still important that users monitor results to ensure fairness.

With large volumes of use this can be carried out by looking at the actual scores of different groups. However, with smaller volumes this will involve making sure that the tests used continue to relate to the role and reviewing other validation studies where you are unable to conduct your own.
SECTION 10

COMPLETION TIMES

In Elements, there is a time limit for each question rather than for the test as a whole. The purpose of such a design is to constrain test takers to operate at a similar level of speed when answering the questions. This enables us to obtain purer and comparable estimates of individual ability.

The maximum time available (full time limit for all questions) is 16 minutes and 15 seconds for Verbal, and 16 minutes for Numerical. The average time taken is 11 minutes 37 seconds for Verbal, and 12 minutes 4 seconds for Numerical. Thus an average of 23 minutes 41 seconds for both tests together – which is less than half the time required for conventional tests of the same type.

The maximum time available for Elements Logical is 15 minutes. The average time taken in the sample of completions we have so far is 8 minutes 41 seconds.
SECTION 11

RELATIONSHIPS BETWEEN ELEMENTS AND DIMENSIONS

Correlations between Elements and Dimensions were as follows for a sample that had completed Dimensions and one or more of the Elements tests:

<table>
<thead>
<tr>
<th>Trait</th>
<th>Verbal</th>
<th>Numerical</th>
<th>Logical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicative</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Influential</td>
<td>0.05</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Socially Confident</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Supportive</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>Consultative</td>
<td>0.02</td>
<td>-0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>Analytical</td>
<td>0.12</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>Conceptual</td>
<td>0.15</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>Creative</td>
<td>0.03</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Methodical</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Conscientious</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.07</td>
</tr>
<tr>
<td>Relaxed</td>
<td>0.05</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>Resilient</td>
<td>0.02</td>
<td>0.03</td>
<td>0.12</td>
</tr>
<tr>
<td>Flexible</td>
<td>0.17</td>
<td>0.08</td>
<td>0.19</td>
</tr>
<tr>
<td>Decisive and Action-oriented</td>
<td>0.03</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>Achievement-oriented</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
</tbody>
</table>

This analysis was conducted on samples in the region of 10,000 (Logical) to 15,000 (Verbal and Numerical) people. As a result, all correlations could theoretically be classed as significant. We had previously hypothesised a positive correlation between Elements Numerical and Analytical, and between Elements Verbal and Conceptual, which are indeed two of the largest correlations for each of these two tests.

The findings support these hypotheses, therefore suggesting that those with a preference for being Analytical are also likely to show strong numerical ability. Similarly, those who have a preference for being Conceptual also demonstrate high verbal ability. Logical ability was also found to be positively correlated with Analytical. This supports the idea that people are likely to have a preference for aspects of work in which they also have an ability, although the direction of this relationship may be unknown. Other notable correlations were seen between being Flexible and having high verbal and logical ability which can likely be explained by the orientation towards learning new things and variety that are often related to problem solving capability.
A common question about psychometric assessments is whether they really translate to other languages and cultures.

Modern tests of verbal and numerical reasoning tend to be very up to date and job-relevant, being based on real or near-real situations. These will sometimes fail to translate literally; we have therefore adopted the following process:

- All assessments are originally developed with the intention that they will need to be translated and adapted to different languages and cultures.
- Translation is carried out by people with the target language as mother tongue, who are also fluent in English.
- The translation team includes occupational psychologists with up-to-date business experience in the target culture.
- Data continues to be collected until we have enough to create a local norm.
- Local norms are at first ‘modelled’ using a combination of local data and wider, international groups. When sufficient data is available, this data becomes the local norm basis.
- We carry out extensive cross-language and cross-cultural validation studies.
- We study cultural differences within same-language populations.

A wide programme of translation is currently under way. A list of available languages can be seen on the Talent Q website: www.talentqgroup.com
SECTION 13

SUMMARY

Talent Q has been founded on the basis of knowledge and insight gained from specialist occupational psychologists and practitioners.

This section provides some general conclusions surrounding Elements, plus an overview of the link with our online personality assessment, Dimensions. Dimensions measures 15 key personality traits and is often used alongside Elements.

The findings of this review illustrate the psychometric rigour of Elements and demonstrate its capability to assess candidates in a way which is fair and objective. In doing so, HR professionals and consultants can be reassured that by using Elements (and other Korn Ferry assessments), they are providing a reliable, fair and objective measure of ability (or personality in the case of Dimensions) which can be applied to recruitment and development decisions throughout the talent lifecycle.

We are committed to ongoing development and analysis. This review will be updated periodically to reflect international expansion in the use of Elements. We are always keen to talk to clients who would like to get involved in this process with us. Please contact your account manager for more information.

Specifically, this review has demonstrated the following:

THE RATIONALE AND DESIGN INTENTIONS UNDERLYING THE CONSTRUCTION OF ELEMENTS

This section of the review focuses on the need to create an adaptive system which would in turn create a fast, randomised suite of ability tests. The adaptive nature of Elements has many benefits, which have been discussed in detail in this review. For example, it offers the advantage of being able to measure people across all role levels with a single assessment.

ELEMENTS NORM GROUPS

Due to the adaptive nature of Elements, a range of norm groups for different roles and levels within an organisation is not needed. Instead, we offer three norm groups which cover most eventualities and can be tailored in exceptional circumstances.
RELIABILITY AND VALIDITY OF ELEMENTS

The sound reliability of Elements has been demonstrated through a plethora of research to establish it as a robust and stable measure of ability. Research has included a test-retest reliability study, split-half reliability analysis and internal consistency analysis. In addition, the validity of Elements has been indicated through the face validity of question content as well as several validation studies.

ELEMENTS AND BIOGRAPHICAL DATA

We have reviewed all Elements completions and looked at this in relation to biographical data provided by candidates. In doing so, it has been established that group differences are minimal when using Elements Verbal, Numerical and Logical ability tests. However, we maintain that client data should be regularly reviewed and cut-offs set accordingly to avoid any unnecessary adverse impact.

ELEMENTS AND DIMENSIONS

Finally, the review explored the relationship between Elements and Dimensions. As expected, a relationship was found between Elements Numerical and the Analytical scale in Dimensions, and between Elements Verbal and the Conceptual scale in Dimensions. These findings are intuitive and in line with expectations, in that we would expect to see people who have a preference for being analytical also having an ability in numerical reasoning. Similarly, those with high verbal reasoning ability are also likely to have a preference for being conceptual.
SECTION 14

GLOSSARY

Correlation

- A statistical way of assessing the strength of relationship between two variables (e.g. analytical reasoning scores on a test and job performance).

Reliability

- Refers to how accurate a test is. There are a number of different types of reliability, as described below.

Test-retest reliability

- This form of reliability assesses the extent to which an individual's score is likely to be similar if the test is taken at different points in time, e.g. a few weeks or months later.

Internal consistency reliability

- Understanding reliability in terms of internal consistency looks inside the test to assess whether the questions, or 'items', which make up the test are all measuring the same thing.

- Split-half reliability: The simplest way of assessing internal consistency is known as split-half reliability. This adds up the score on the items for half of the test (e.g. all the 'even' numbered items) and correlates this with the score on the items for the other half of the test (e.g. all the 'odd' numbered items).

- Cronbach's alpha: This still leaves the issue of how well related the different halves actually are, compared with other combinations. As a result, statisticians have developed more complex equations for assessing internal consistency, which in practice identify the average level or correlation for all possible 'split-half' combinations that could be made. One of the best known of these is Cronbach's alpha.

Validity

- Refers to the extent to which a test can be deemed to be effective in determining what we want to be able to predict, for instance future performance in a job. There are a number of types of validity, as described below.

Construct validity

- Concerned with whether a test is measuring the 'construct' it claims to measure. There are two ways of assessing construct validity:

  - Convergent validity: If the test can be shown to correlate significantly with other tests designed to measure the same or similar constructs, it may be claimed to have convergent validity.

  - Discriminant validity: If the test can be shown to not correlate significantly with other tests which are intended to measure different constructs, discriminant or 'divergent' validity can be claimed.

Face validity

- Related to whether people looking at the test, using their own judgement and experiences, think it does what it is supposed to do.
**Criterion validity**
- Tells us whether the results of a test predict something which is practically useful, such as subsequent performance in a role.

**Content validity**
- Looks at the validity of a test’s content, such as the questions and scenarios used. This requires that the questions in a test can be seen to directly relate to the behaviour or aspect of performance they are intended to measure.

**Computer adaptive testing (CAT)**
- CAT refers to a form of computer-based testing in which the difficulty level adapts depending on the candidate’s ability level.

**Item response theory (IRT)**
- IRT is a method employed in the design, analysis and scoring of assessments which draws on the application of mathematical models to testing data. IRT focuses on the question (or ‘item’) rather than the test level focus of classical test theory, by modelling the response of a candidate of a given ability to each item in the test.

**Classical test theory (CTT)**
- CTT is a method employed in the design, analysis and scoring of assessments. The essential idea behind CTT is that an observed assessment score is the sum of a ‘true’ score plus measurement error. The most important formula that lies at the core of classical test theory is defined as follows:

\[ X = T + E, \]

where
- \( X \) = the total score/observed score obtained
- \( T \) = the true score
- \( E \) = the error component

**Rasch scaling**
- This is a method of analysing assessment data to measure ability; it can also be used to measure attitudes and personality traits. It is theoretically similar to IRT, but uses a specific property that provides a criterion for successful measurement.

**Standard deviation**
- Refers to the amount of variation or ‘dispersion’ there is from the average (mean).
SECTION 15

REFERENCES


ABOUT KORN FERRY

Korn Ferry is the preeminent global people and organizational advisory firm. We help leaders, organizations, and societies succeed by releasing the full power and potential of people.

Through our Executive Search, Hay Group, and Futurestep divisions, our nearly 7,000 colleagues deliver services in the following areas:

- Strategy Execution and Organization Design
- Talent Strategy and Work Design
- Rewards and Benefits
- Assessment and Succession
- Executive Search and Recruitment
- Leadership Development

See how we help your organization rise UP at kornferry.com

Contact us:
http://infokf.kornferry.com/-PRO-contact-us.html