

TECHNICAL MANUAL & USER GUIDE

REASONING TEST 20'

This manual has been created by the psychologists from the Research and Innovation department of Central Test.

The assessment REASONING TEST 20' and the Central Test logo are registered trademarks.

No reproduction, modification or adaptation is permitted without the written consent of the publisher Central Test.

*Copyright © 2020 by Central Test - Empowering Talent.
All rights reserved.*

Table of Contents

Introduction	3
1. General Description and Applicability	4
1.1. Factor Descriptions	4
1.2. Main Areas of Application	5
1.3. Administration and Questionnaire Structure	5
1.4. Technical and Security Issues	5
2. Psychometric Properties	7
2.1. Sample Descriptives	7
2.2. Norms	9
2.3. Reliability	14
2.4. Validity	14
2.5. REASONING TEST 20' and Equal Opportunities	16
3. Additional Information and Considerations	22
References	23

Introduction

Wechsler describes intelligence as “the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment.”

The three capacities, as mentioned by Wechsler, are of extreme importance in a corporate environment wherein executives are expected to take stable and reasonable measures, and to accommodate to frequent market changes. In order to facilitate organisations in the measurement of these requisites, the **REASONING TEST 20'** assesses the overall and specific intellectual capacities of corporate individuals across three areas, namely, *Figural Reasoning*, *Verbal Reasoning* and *Numerical Reasoning*. It is deemed that these three areas are important as they encompass skills that are relevant in the corporate world such as communication, planning, organizing, calculating and foreseeing outcomes.

1. General Description and Applicability

Intelligence is a multifaceted and multilayered concept. Among all the factors considered to make up one's intelligence, researchers found three principle factors, namely, numerical, verbal and logical reasoning skills to be most predictive of General Intelligence or Spearman's "g-factor". The IQ (Intelligence Quotient) obtained on any intelligence test is a numerical representation of this "g-factor". It offers a global picture of an individual's capacity to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, learn from experience and adapt to most situations.

Therefore, the **REASONING TEST 20'** is designed to assess these three essential factors of Intelligence – *Figural Reasoning*, *Verbal Reasoning* and *Numerical Reasoning*.

The **REASONING TEST 20'** Report includes the following:

- An Intelligence Quotient to reflect an individual's overall performance
- The number of correct responses obtained by the individual
- A graphical representation of scores on each of the three factors – *Figural Reasoning*, *Verbal Reasoning* and *Numerical Reasoning*
- A detailed description of each factor
- Solutions along with the individual's responses to the correct answers

1.1. Factor Descriptions

The **REASONING TEST 20'** measures the three important components of intelligence, i.e. *Figural Reasoning*, *Verbal Reasoning* and *Numerical Reasoning*.

Figural Reasoning - The *Figural Reasoning* factor refers to the ability to make connections between various elements and discover existing relationships within sets of varying complexity.

This type of reasoning is involved in virtually all aspects of everyday life. Whenever a person has to discover the relationships between different parts of a set, logical faculties are stimulated. It is the ability most strongly related to general intelligence.

In professional life, many activities involve figural reasoning. For example, information technology, law and medicine are occupations in which logic is especially important.

Verbal Reasoning - The *Verbal Reasoning* factor corresponds to the ability to understand language subtleties and to be at ease with verbal information.

Situations that draw on this factor are numerous and varied: speaking diplomatically, playing word games, etc.

The written word and language play a crucial role in our daily lives, enabling us to understand and process the mass of information that surrounds us as well as contribute our part.

Moreover, verbal communication has even become a form of knowledge transmission. Therefore being able to effectively communicate is of utmost importance for surviving as well as for thriving in today's world.

Numerical Reasoning – The *Numerical Reasoning* factor represents the ability to reason successfully and accurately with numbers.

Real-life situations where numerical skills will be needed are numerous and varied: economics, finance, accounting, sales negotiations, quantitative data analysis, as well as doing mental calculations, etc.

1.2. Main Areas of Application

The **REASONING TEST 20'** is recommended for professionals such as sales persons, entrepreneurs, managers and other office executives.

The **REASONING TEST 20'** is especially reserved for professionals in the Human Resource departments of organizations, schools and colleges, as well as trainers.

1.3. Administration and Questionnaire Structure

Remote Administration of the REASONING TEST 20'

The **REASONING TEST 20'** can be administered under non-supervised settings; however, appropriate care should be taken by the test administrator to minimize the possibility of getting external aid on the part of the test-taker.

Format of the REASONING TEST 20'

- Number of Questions : 21
- Assessment Format : Multiple Choice Questionnaire
- Duration : 20 minutes (Timed)
- Administration : Online

1.4. Technical and Security Issues

The following points serve as guidelines for securely administering the **REASONING TEST 20'**. These are useful for ensuring that the **REASONING TEST 20'** is administered in an objective and fair manner. Security specifications are also important for protecting the candidate's information, both during administration and after the completion of the assessment.

Respecting the Ethical Charter

As a publisher of online assessment tools, *Central Test* respects the legal conventions of French labour law and the regulatory provisions resulting from the law of 6 January 1978 pertaining to information technology, files and freedoms: Declaration N°1834067. Additionally, as of 25th May, 2018, *Central Test* complies with the GDPR (General Data Protection Regulation) regulations made by the European Parliament & Council.

Test results and personal data cannot be shared with third parties without explicit consent from the candidate. Candidates have the right to access their results and the assessor must grant them access

when requested. Furthermore, if the candidate requests it, all data must be permanently deleted; to do so the candidate must contact either the organisation or *Central Test* directly.

Confidentiality and Security

Central Test guarantees the security and confidentiality of data generated during the administration of online tests: access codes are encrypted in the database and cannot be accessed by anyone. Passwords are computationally generated and not visible to companies. A URL rewriting module (secure link generator) has been installed and has transparent parameter passing for the user.

Browsing and data exchanged via the *Central Test* application can be secured and represented by an icon in the form of a padlock in the internet browser. The *Central Test* application has a “GeoTrust SSL TRUE BusinessID” digital signature, issued and certified by GeoTrust®. This certificate protects all *Central Test* assessments from phishing attacks, and serves as proof of our dedication to providing a secure platform.

Invitations to take tests are made via secure links that are of limited use (in terms of numbers and duration). The tests are timed and IP addresses are tracked.

2. Psychometric Properties

Central Test ensures that the data provided from our assessments and reported in our manuals are objective, and do not discriminate people based on factors such as gender, age, country of origin as well as study level. Such is also the case with this **REASONING TEST 20'** Manual as it was made for an English-speaking sample from across the globe in order to demonstrate that the assessment is reliable and valid.

The sample used comprises of 7,088 English-speaking professionals across 60 countries.

2.1. Sample Descriptives

Table 1. Distribution based on Gender.

Gender	Percentage (%)	N
Female	40.05	2839
Male	59.95	4249

Table 2. Distribution based on Age Group.

Age Group	Percentage (%)	N
Not Defined	3.60	255
18 – 25	52.78	3741
26 – 33	32.14	2278
34 – 41	9.14	648
42 – 49	1.79	127
50+	0.55	39

Table 3. Distribution based on Country of Origin.

Country of Origin	Percentage (%)	N
India	47.43	3362
Nigeria	46.53	3298
Others	6.04	428

Note: Countries of Origin in the "Others" category has been grouped together as so as they correspond to a number that is not sufficient to be represented. The countries grouped as so are Afghanistan, Algeria, Angola, Australia, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Cyprus, Czech Republic, Egypt, France, Gabon, Ghana, Guadeloupe, Hong Kong, Iceland, Iran, Iraq, Israel, Italy, Japan, Jordan, Kenya, Kuwait, Lebanon, Malaysia, Mauritius, Mexico, Mozambique, Nepal, New Zealand, Niger, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Qatar, Russia, Rwanda, Saint Martin (Netherlands), Saudi Arabia, Singapore, South Africa, Spain, State of Palestine, Switzerland, Taiwan, Tunisia, United Arab Emirates, United Kingdom, United States of America and Zambia.

Table 4. Distribution based on Study Level.

Study Level	Percentage (%)	N
Level 1: Master's Degree, PhD, Engineering or Business School, MBA	11.06	784
Level 2: Bachelor's Degree, First-year Masters	54.95	3895
Level 3: 2 years of Vocational or University Studies	5.76	408
Level 4: High School Diploma	5.06	359
Level 5: Trades Training, No High School Diploma	0.78	55
Level 6: Students	13.83	980
Not Defined	8.56	607

2.2. Norms

Norms help us to compare an individual's score on a test with the scores of others who complete the same assessment, usually under similar conditions. These individuals are representative of the population for which an assessment is intended, and are therefore referred to as the norm group. The closer the match between the characteristics of the sample and the target population, the more accurate the distribution will be as a ranking guide. With norms, test developers and psychologists are able to establish comparisons between individuals and make more objective decisions. Norm groups enable us to attribute meaning to test scores and quantify an individual's performance, level of an attribute or trait relative to his/her peers.

Tables 5 to 9 detail the Mean, Standard Deviation (SD), Median, Minimum (Min), Maximum (Max), Range and Standard Error of Mean (SE) scores for the English-speaking sample on the factors of the **REASONING TEST 20'**. These tables are divided into categories based on gender, age group, country of origin and study level.

Table 5. Factor Score Norms for the Whole Sample.

Factor	Mean	SD	Median	Min	Max	Range	SE
Intelligence Quotient	11.47	3.34	11.00	0.00	20.00	20.00	0.04
Figural Reasoning	6.04	2.33	5.70	0.00	10.00	10.00	0.03
Numerical Reasoning	5.38	2.57	5.70	0.00	10.00	10.00	0.03
Verbal Reasoning	5.88	1.95	5.70	0.00	10.00	10.00	0.02

Table 6. Factor Score Norms based on Gender.

Factor	Mean	SD	Median	Min	Max	Range	SE
Females							
Intelligence Quotient	11.41	3.35	11.00	0.00	20.00	20.00	0.06
Figural Reasoning	6.01	2.31	5.70	0.00	10.00	10.00	0.04
Numerical Reasoning	5.17	2.59	5.70	0.00	10.00	10.00	0.05
Verbal Reasoning	6.04	1.95	5.70	0.00	10.00	10.00	0.04
Factor	Mean	SD	Median	Min	Max	Range	SE

Males							
Intelligence Quotient	11.50	3.33	11.00	1.00	20.00	19.00	0.05
Figural Reasoning	6.06	2.34	5.70	0.00	10.00	10.00	0.04
Numerical Reasoning	5.52	2.55	5.70	0.00	10.00	10.00	0.04
Verbal Reasoning	5.78	1.95	5.70	0.00	10.00	10.00	0.03

Table 7. Factor Score Norms based on Age Group.

Factor	Mean	SD	Median	Min	Max	Range	SE
18 – 25							
Intelligence Quotient	11.48	3.38	11.00	1.00	20.00	19.00	0.06
Figural Reasoning	6.13	2.33	5.70	0.00	10.00	10.00	0.04
Numerical Reasoning	5.53	2.55	5.70	0.00	10.00	10.00	0.04
Verbal Reasoning	5.67	1.98	5.70	0.00	10.00	10.00	0.03

Factor	Mean	SD	Median	Min	Max	Range	SE
26 – 33							
Intelligence Quotient	11.51	3.33	11.00	0.00	20.00	20.00	0.07
Figural Reasoning	5.97	2.34	5.70	0.00	10.00	10.00	0.05
Numerical Reasoning	5.23	2.61	5.70	0.00	10.00	10.00	0.05
Verbal Reasoning	6.16	1.88	5.70	0.00	10.00	10.00	0.04

Factor	Mean	SD	Median	Min	Max	Range	SE
34 – 41							
Intelligence Quotient	11.53	3.15	11.00	4.00	20.00	16.00	0.12
Figural Reasoning	5.99	2.31	5.70	0.00	10.00	10.00	0.09

Numerical Reasoning	5.17	2.49	5.70	0.00	10.00	10.00	0.10
Verbal Reasoning	6.25	1.82	5.70	1.40	10.00	8.60	0.07

Factor	Mean	SD	Median	Min	Max	Range	SE
---------------	-------------	-----------	---------------	------------	------------	--------------	-----------

42 – 49

Intelligence Quotient	11.28	3.12	11.00	5.00	19.00	14.00	0.28
Figural Reasoning	5.59	2.20	5.70	0.00	10.00	10.00	0.20
Numerical Reasoning	4.98	2.53	4.30	0.00	10.00	10.00	0.22
Verbal Reasoning	6.44	1.86	5.70	1.40	10.00	8.60	0.17

Factor	Mean	SD	Median	Min	Max	Range	SE
---------------	-------------	-----------	---------------	------------	------------	--------------	-----------

50+

Intelligence Quotient	12.15	4.04	13.00	4.00	19.00	15.00	0.65
Figural Reasoning	5.86	2.74	5.70	0.00	10.00	10.00	0.44
Numerical Reasoning	5.34	3.05	5.70	0.00	10.00	10.00	0.49
Verbal Reasoning	7.14	2.05	7.10	2.90	10.00	7.10	0.33

Table 8. Factor Score Norm based on Country of Origin.

Factor	Mean	SD	Median	Min	Max	Range	SE
---------------	-------------	-----------	---------------	------------	------------	--------------	-----------

India

Intelligence Quotient	11.00	3.31	11.00	1.00	20.00	19.00	0.06
Figural Reasoning	5.94	2.31	5.70	0.00	10.00	10.00	0.04
Numerical Reasoning	5.36	2.51	5.70	0.00	10.00	10.00	0.04
Verbal Reasoning	5.27	1.93	5.70	0.00	10.00	10.00	0.03

Factor	Mean	SD	Median	Min	Max	Range	SE
Nigeria							
Intelligence Quotient	11.67	3.19	11.00	0.00	20.00	20.00	0.06
Figural Reasoning	6.02	2.32	5.70	0.00	10.00	10.00	0.04
Numerical Reasoning	5.25	2.58	5.70	0.00	10.00	10.00	0.04
Verbal Reasoning	6.36	1.76	6.40	0.00	10.00	10.00	0.03

Factor	Mean	SD	Median	Min	Max	Range	SE
Others							
Intelligence Quotient	13.58	3.69	14.00	4.00	20.00	16.00	0.18
Figural Reasoning	6.98	2.31	7.10	1.40	10.00	8.60	0.11
Numerical Reasoning	6.54	2.72	7.10	0.00	10.00	10.00	0.13
Verbal Reasoning	7.02	2.03	7.10	1.40	10.00	8.60	0.10

Table 9. Factor Score Norms based on Study Level.

Factor	Mean	SD	Median	Min	Max	Range	SE
Level 1: Master's Degree, PhD, Engineering or Business School, MBA							
Intelligence Quotient	12.15	3.48	12.00	3.00	20.00	17.00	0.12
Figural Reasoning	6.39	2.41	7.10	0.00	10.00	10.00	0.09
Numerical Reasoning	5.76	2.68	5.70	0.00	10.00	10.00	0.10
Verbal Reasoning	6.22	1.96	5.70	0.00	10.00	10.00	0.07

Factor	Mean	SD	Median	Min	Max	Range	SE
Level 2: Bachelor's Degree, First-year Masters							
Intelligence Quotient	11.64	3.29	11.00	2.00	20.00	18.00	0.05
Figural Reasoning	6.12	2.32	5.70	0.00	10.00	10.00	0.04

Numerical Reasoning	5.40	2.55	5.70	0.00	10.00	10.00	0.04
Verbal Reasoning	6.06	1.91	5.70	0.00	10.00	10.00	0.03

Factor	Mean	SD	Median	Min	Max	Range	SE
--------	------	----	--------	-----	-----	-------	----

Level 3: 2 years of Vocational or University Studies

Intelligence Quotient	10.73	3.36	10.00	0.00	20.00	20.00	0.17
Figural Reasoning	5.53	2.34	5.70	0.00	10.00	10.00	0.12
Numerical Reasoning	4.98	2.61	4.30	0.00	10.00	10.00	0.13
Verbal Reasoning	5.59	1.98	5.70	0.00	10.00	10.00	0.10

Factor	Mean	SD	Median	Min	Max	Range	SE
--------	------	----	--------	-----	-----	-------	----

Level 4: High School Diploma

Intelligence Quotient	10.54	3.33	10.00	2.00	19.00	17.00	0.18
Figural Reasoning	5.53	2.35	5.70	0.00	10.00	10.00	0.12
Numerical Reasoning	4.73	2.52	4.30	0.00	10.00	10.00	0.13
Verbal Reasoning	5.55	2.01	5.70	0.00	10.00	10.00	0.11

Factor	Mean	SD	Median	Min	Max	Range	SE
--------	------	----	--------	-----	-----	-------	----

Level 5: Trades Training, No High School Diploma

Intelligence Quotient	8.91	2.82	9.00	2.00	16.00	14.00	0.38
Figural Reasoning	4.37	1.98	4.30	0.00	8.60	8.60	0.27
Numerical Reasoning	4.05	2.40	4.30	0.00	10.00	10.00	0.32
Verbal Reasoning	4.76	1.86	4.30	1.40	10.00	8.60	0.25

Factor	Mean	SD	Median	Min	Max	Range	SE
Level 6: Students							
Intelligence Quotient	10.89	3.18	10.00	2.00	20.00	18.00	0.10
Figural Reasoning	5.84	2.25	5.70	0.00	10.00	10.00	0.07
Numerical Reasoning	5.25	2.49	5.70	0.00	10.00	10.00	0.08
Verbal Reasoning	5.31	1.91	5.70	0.00	10.00	10.00	0.06

2.3. Reliability

Reliability refers to the extent to which an assessment is free from error and the degree to which items on a scale are measuring the same construct. For a measure to be reliable, the same results must be consistently generated over time. Reliability is important for an assessment as it enables us to have confidence that scores on a test are really typical of an individual's behavior, aptitude or attitude. A primary concern when developing and using tests, is the extent to which random measurement errors influence test performance. If a test is unreliable, it is impossible for it to be valid as accuracy of results (validity) depends on the consistency with which the same results are produced. High reliability of an assessment helps to guard against biases that may be caused by errors in human judgment during the recruitment and selection process.

Reliability values vary from 0 to 1, with higher values being more desirable. As a rule of thumb, a reliability of 0.70 or higher is recommended (Nunnally, 1978), with values greater than 0.9 (excellent); 0.7 to 0.9 (good); 0.6 to 0.7 (acceptable); 0.5 to 0.6 (poor) and values less than 0.5 (unacceptable).

Internal reliability measures the extent to which items on a scale are internally consistent and all measure the same construct. For the **REASONING TEST 20'**, a powerful and lesser-known estimator of internal reliability based on the assumptions of Classical Test Theory, called the Greatest Lower Bound (GLB) was used. GLB produces better results over alpha (Oliver et al., 2014) and together over alpha and omega (Wilcox et al., 2014), in cases where the data is not normally distributed.

For the **REASONING TEST 20'**, a greatest lower bound to reliability of 0.8 was obtained, indicating that the scale has a good level of consistency and is reliable.

2.4. Validity

Validity refers to the extent to which an assessment measures what it claims to measure, and that this is well-founded and accurately corresponds to actual behaviour. In psychometrics, validity relates to the degree to which practical evidence and scientific theory support the interpretations of test scores. Whilst the reliability of an assessment relates to the degree to which results are consistent over time, validity tells us the degree to which results of an assessment are accurate and enable us to confidently make predictions about a target outcome. Therefore, a test can be reliable but not valid (one can consistently get the wrong answer) but it cannot be valid without being reliable, as confidence in the accuracy of results depends on the consistency with which we obtain

the same results. Established validity protects practitioners from faulty, costly, unethical and legally risky decisions.

Construct validity measures the degree of the appropriateness of inferences made on the basis of observation or measurement. For work-based assessments, construct validity is particularly important as it: a) adds legitimacy and acceptability to the standardised measurement of employee behaviours, b) facilitates understanding of organisational implications of findings, and c) enables organisations and individuals to tailor interventions to meet desired performance targets. For **GAAT**, construct validity was established using two approaches: factor correlations and confirmatory factor analysis.

The factors in the **REASONING TEST 20'** measure reasoning skills. However, each factor measures a type of reasoning skill and therefore, the individual is provided with 3 separate scores for each type of skill. Therefore, if someone scores highly on one factor, it does not mean that they will score highly on the other 2 factors.

The correlation coefficient is a numerical way to quantify the relationship between the variables, and ranges between -1 and 1. The closer the value to 1, the stronger the relationship. The correlation coefficient does not imply causality; it merely signifies the association between variables. Traditionally, factor correlations are measured using Pearson (1895) correlation.

Table 10. Inter-factor Correlations.

	Figural Reasoning	Numerical Reasoning
Numerical Reasoning	0.53	
Verbal Reasoning	0.31	0.29

The results displayed in Table 10 shows that a positive relationship exists between the 3 factors. Between *Figural Reasoning* and *Numerical Reasoning*, there exists a moderate uphill (positive) relationship. Between *Verbal Reasoning* and *Figural Reasoning* as well as between *Verbal Reasoning* and *Numerical Reasoning*, a weak uphill (positive) linear relationship exists.

A positive correlation is to be expected in an intelligence assessment as all three factors, *Numerical Reasoning*, *Verbal Reasoning* and *Figural Reasoning*, contribute towards the measurement of a second-order intelligence quotient factor or in other words, an overall intellectual capacity.

Confirmatory Factor Analysis

To investigate the theoretical constructs or factors that may be represented by a set of items as well as the quality of individual items, a confirmatory factor analysis (CFA) is usually performed. CFA is used to determine how well the model fits the data, and therefore produces fit statistics that are indicative of the test. Several statistical tests are used to determine how well the model fits the data in CFA and SEM. The closer fit indices meet statistical requirements, the better the model. The Chi Square (X^2) statistic is considered the primary index of goodness-of-fit. However, due to problems associated with reliance on the X^2 statistic's association with sample size (the statistic increases with the number of items in the model regardless of fit), it is recommended to report measures of *relative*

fit alongside it. The X^2 is a measure of absolute fit, and provides a fundamental indication of how well the proposed theory fits the data and does not rely on comparison with a baseline model (Jöreskog & Sörbom, 1993).

For the validation of the **REASONING TEST 20'**, the Root Mean Square Error of Approximation (RMSEA, Steiger & Lind, 1980), the Bentler Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI) were chosen as additional measures of fit, as they are largely unaffected by sample size (Fan, Thompson & Wang, 1999). Furthermore, different indices are necessary as each index reflects an aspect of model fit (Hooper, Coughlan & Mullen, 2008). Like the X^2 , the RMSEA is a measure of absolute fit that assesses the error associated with a proposed model. The lower the RMSEA, the more robust the model; indices of .05 or below are considered to be a good fit (Hu & Bentler, 1999). RMSEA is calculated using the model (i.e. data that does not work with the fitted model) and therefore provides a measure of error in the model. Both the CFI and TLI are *incremental* or *relative* fit indices that assess overall improvement of the theorized model compared with an independence model (i.e. a model where we expect variation but no correlation between items) where observed variables are uncorrelated (Hooper et al., 2008; Holtzman & Vezzu, 2011). While higher values of the CFI and TLI (.9 and above) indicate an excellent fit, it should be noted that a CFI or TLI of .90 or higher may not necessarily correspond to a RMSEA of .05 or less (Hu & Bentler, 1999).

Table 11. The Goodness of Fit Indices and Model Results.

N	X ²	df	X ² /df	P	RMSEA	P<=.05	CFI	TLI
7088	1273.020	186	6.844	0.000	.029	1.000	.942	.934

The results displayed in Table 11, shows an RMSEA of .029, CFI of .942 and TLI of .934, indicating that the **REASONING TEST 20'** has an excellent model fit and that the factors exist in the defined structure.

2.4. REASONING TEST 20' and Equal Opportunities

Organisations are increasingly under pressure to ensure that tools used to inform employee recruitment, selection, development, promotion and performance measurement decisions are fair, objective and accurate. This pressure stems from a progressively knowledgeable applicant pool, a more sophisticated labour force and stringent local and international employment legislation and testing standards. Equal opportunities therefore represent a critical consideration for any modern day work assessment, and an assessment is only considered to be truly valid when it is appropriate for use with different groups. The **REASONING TEST 20'** is a fair and accurate measure of critical thinking, and HR practitioners and psychologists can be confident of its application.

Measurement Equivalence

Measurement equivalence (ME) refers to the consistency of measurement across some specified group demarcation such as nationality, ethnicity and gender. The focus of ME is to ensure that individuals from different groups with the same standing on an attribute have the same set of observations on an assessment (Drasgow, 1987; McLeary & Cruise, 2012). ME helps to determine the extent to which respondents are interpreting items similarly, and identified differences across groups

reflect actual differences in the trait being measured (Drasgow, 1987). One approach to assessing measurement equivalence is differential item functioning analysis.

Differential Item Functioning (DIF)

An item functions differently, or displays DIF, if respondents from different groups have different probabilities of endorsing the item, despite having the same level of the attribute or trait. When a different probability of endorsing an item remains constant across different levels of the attribute or trait, the DIF is said to be uniform. Here, the item of interest consistently gives one group an advantage across all levels of the attribute or trait. Conversely, when a different probability of endorsing an item varies along the level of the attribute or trait, the DIF is characterised as non-uniform (Clauser & Mazor, 1998; Hanson, 1998). Rather than a consistent advantage being given to the reference group across the ability continuum, an advantage to one group over another exists at different levels of the trait. For instance, an item may give the reference group an advantage at the higher end of the continuum, while a disadvantage at the lower end.

DIF Analysis: Gender

DIF analysis was performed across gender to ensure that there was not a bias towards either males or females, and therefore that the **REASONING TEST 20'** is a fair test for both genders. For this analysis, the norm sample was divided according to gender; 2839 females and 4249 males.

Table 12. Coefficients for Significant DIF and Effect Size for Dif Items.

Factor	Coefficient	R ²	Effect size
Figural Reasoning 1	6.710	0.000	Negligible
Figural Reasoning 2	11.349	0.000	Negligible
Numerical Reasoning 1	6.510	0.000	Negligible
Numerical Reasoning 2	11.596	0.000	Negligible
Verbal Reasoning 1	6.159	0.000	Negligible

The analysis flagged 5 of the 21 items for DIF, based on gender. Table 12 shows the DIF coefficient as well as the R² effect size. According to Zumbo and Thomas' (1997) classification of effect size, any R² under 0.13 is considered negligible. This is because an item is expected to have a statistically significant DIF statistic if the sample is large – as it is highly unlikely that demographic groups from a large sample, will have the same level on every item – and the R² gives an indication of how much an item is causing this difference as opposed to whether there is a difference. Out of the 5 items flagged for DIF, none showed an R² ≥ 0.13. Therefore, it can be considered that the **REASONING TEST 20'** does not contain any item that functions differently across genders.

DIF Analysis: Age

DIF analysis was performed across age to ensure that there was not a bias towards either those individuals below the age of 40 and those at and above 40, and therefore that the **REASONING TEST 20'** is a fair test for both categories. For this analysis, the norm sample was divided between the 2 categories; 6835 individuals below the age of 40 and 253 individuals at or above the age of 40.

Table 13. Coefficients for Significant DIF and Effect Size for Dif Items.

Factor	Coefficient	R ²	Effect size
Figural Reasoning 1	6.435	0.000	Negligible
Numerical Reasoning 1	22.969	0.000	Negligible
Numerical Reasoning 2	10.389	0.000	Negligible
Verbal Reasoning 1	12.928	0.000	Negligible
Verbal Reasoning 2	10.161	0.000	Negligible
Verbal Reasoning 3	7.172	0.000	Negligible
Verbal Reasoning 4	11.564	0.000	Negligible

The analysis flagged 7 of the 21 items for DIF, based on the age categories. Table 13 shows the DIF coefficient as well as the R² effect size. According to Zumbo and Thomas' (1997) classification of effect size, any R² under 0.13 is considered negligible. This is because an item is expected to have a statistically significant DIF statistic if the sample is large – as it is highly unlikely that demographic groups from a large sample, will have the same level on every item – and the R² gives an indication of how much an item is causing this difference as opposed to whether there is a difference. Out of the 7 items flagged for DIF, none showed an R² ≥ 0.13. Therefore, it can be considered that the **REASONING TEST 20'** does not contain any item that functions differently across individuals who are below the age of 40 and those who are at or above the age of 40.

T-tests and ANOVAS

Although 6 items were flagged with DIF, their impact on the overall assessment is negligible, meaning that the assessment does not bias towards gender at the item level. Therefore, any significant difference in the assessment's factors can be attributed to a difference in groups from the population from which the sample was drawn, rather than biases of the assessment. To further explore the differences found between the groups, an independent sample t-test was performed between the groups; t-tests and ANOVAS were performed between genders and age groups (see Tables 14 to 17).

Gender

In order to check for differences in scores between genders, an independent sample t-test was conducted to compare all the **REASONING TEST 20'** factors between females (n = 2839) and males (n = 4249) (see Table 14).

Table 14. T-tests For Factors by Gender.

Factor	Female	Male	T.	df	P
Figural Reasoning	6.01	6.06	-0.85	7086	0.39
Numerical Reasoning	5.17	5.52	-5.61	7086	0.00
Verbal Reasoning	6.04	5.78	5.39	7086	0.00

The t-tests found that there were statistically significant differences between the groups, with males scoring higher than females on *Numerical Reasoning* and females scoring higher than males on *Verbal Reasoning*, though, in both cases, the difference is small.

Age Group

In order to check for differences in scores between age group, an ANOVA was conducted to compare all **REASONING TEST 20'** factors between five age groups: 18-25 (n = 3741), 26-33 (n = 2278), 34-41 (n = 648), 42-49 (n = 127) and 50+ (n = 39) (see Table 15).

Table 15. ANOVA by Age Group.

Factor	18-25	26-33	34-41	42-49	50+	F	df	p-value
Figural Reasoning	6.13	5.97	5.99	5.59	5.86	2.348×10^{35}	4	0.00
Numerical Reasoning	5.53	5.23	5.17	4.98	5.34	1.021×10^{36}	4	0.00
Verbal Reasoning	5.67	6.16	6.26	6.44	7.14	6.18×10^{35}	4	0.00

The ANOVA found that there were significant differences for all the factors. Post-hoc analysis was conducted using pair-wise t-tests (Bonferroni correction) in order to see where these differences occurred. Individuals in the age category 18-25 (M = 5.53) scored higher than those in the age categories 26-33 (M = 5.23, $p < 0.001$) and 34-41 (M = 5.17, $p = 0.01391$) on *Numerical Reasoning*. Individuals in the age category 18-25 (M = 5.67) scored lower than those in the age categories 26-33 (M = 6.16, $p < 0.001$), 34-41 (M = 6.26, $p < 0.001$), 42-49 (M = 6.44, $p < 0.001$) and 50+ (M = 7.14, $p < 0.001$); and those in age category 26-33 (M = 6.16) scored lower than those in 50+ (M = 7.14, $p = 0.02576$) on *Verbal Reasoning*.

Country of Origin

In order to check for differences in scores between countries of origin, an ANOVA was conducted to compare all **REASONING TEST 20'** factors between three 3 categories: India (n = 3362), Nigeria (n = 3298) and Others (n = 428) (see Table 16).

Table 16. ANOVA by Country of Origin.

Factor	India	Nigeria	Others	F	df	p-value
Figural Reasoning	5.94	6.02	6.98	38.67	2	0.00
Numerical Reasoning	5.36	5.25	6.54	48.87	32	0.00
Verbal Reasoning	5.27	6.36	7.02	367.10	2	0.00

The ANOVA found that there were significant differences for all the factors. Post-hoc analysis was conducted using pair-wise t-tests (Bonferroni correction) in order to see where these differences occurred. Individuals in the category Others (M = 6.98) scored higher than those in the categories India (M = 5.94, $p < 0.001$) and Nigeria (M = 6.02, $p < 0.001$) on *Figural Reasoning*. Individuals in the category Others (M = 6.54) scored higher than those in the categories India (M = 5.36, $p < 0.001$) and Nigeria (M = 5.25, $p < 0.001$) on *Numerical Reasoning*. Individuals in the category India (M = 5.27) scored lower than those in the categories Nigeria (M = 6.36, $p < 0.001$) and Others (M = 7.02, $p < 0.001$); and those in category Nigeria (M = 6.36) scored lower than those in Others (M = 7.02, $p < 0.001$) on *Verbal Reasoning*.

Study Level

In order to check for differences in scores between study level, an ANOVA was conducted to compare all **REASONING TEST 20'** factors between six study levels: Level 1 - Master's Degree, PhD, Engineering or Business School, MBA (n = 784); Level 2 - Bachelor's Degree, First-year Master's (n = 3895); Level 3 - 2 years of Vocational or University Study (n = 408); Level 4 - High School diploma (n = 359); Level 5 - Trades Training, No High School Diploma (n = 55); and Level 6 - Students (n = 980) (see Table 17).

Table 17. ANOVA by Study Level.

Factor	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	F	df	p-value
Figural Reasoning	6.39	6.12	5.53	5.53	4.37	5.84	16.67	6	0.00
Numerical Reasoning	5.76	5.40	4.98	4.73	4.05	5.25	13.08	6	0.00
Verbal Reasoning	6.22	6.06	5.59	5.55	4.76	5.31	31.58	6	0.00

The ANOVA found that there were significant differences for all the factors. Post-hoc analysis was conducted using pair-wise t-tests (Bonferroni correction) in order to see where these differences occurred.

On Figural Reasoning,

individuals in Level 1 ($M = 6.39$) scored higher than those in the Level 3 ($M = 5.53$, $p < 0.001$), Level 4 ($M = 5.53$, $p < 0.001$), Level 5 ($M = 4.37$, $p < 0.001$) and Level 6 ($M = 5.84$, $p < 0.001$). Individuals in Level 2 ($M = 6.12$) scored higher than those in the Level 3 ($M = 5.53$, $p < 0.001$), Level 4 ($M = 5.53$, $p < 0.001$), Level 5 ($M = 4.37$, $p < 0.001$) and Level 6 ($M = 5.84$, $p = 0.01024$). Individuals in Level 5 ($M = 4.37$) scored lower than those in Level 3 ($M = 5.53$, $p = 0.01050$), Level 4 ($M = 5.53$, $p = 0.01164$) and Level 6 ($M = 5.84$, $p < 0.001$).

On Numerical Reasoning,

individuals in Level 1 ($M = 5.76$) scored higher than those in the Level 2 ($M = 5.40$, $p = 0.00747$), Level 3 ($M = 4.98$, $p < 0.001$), Level 4 ($M = 4.73$, $p < 0.001$), Level 5 ($M = 4.05$, $p < 0.001$) and Level 6 ($M = 5.84$, $p < 0.001$). Individuals in Level 2 ($M = 5.40$) scored higher than those in the Level 3 ($M = 4.98$, $p = 0.03510$), Level 4 ($M = 4.73$, $p = 0.00004$) and Level 5 ($M = 4.05$, $p = 0.00226$). Individuals in Level 6 ($M = 5.25$) scored higher than those in Level 4 ($M = 4.73$, $p = 0.02061$) and Level 5 ($M = 4.05$, $p = 0.01505$).

On Verbal Reasoning,

individuals in Level 1 ($M = 6.22$) scored higher than those in the Level 3 ($M = 5.59$, $p < 0.001$), Level 4 ($M = 5.55$, $p < 0.001$), Level 5 ($M = 4.76$, $p = 0.00000$) and Level 6 ($M = 5.31$, $p < 0.001$). Individuals in Level 2 ($M = 6.06$) scored higher than those in the Level 3 ($M = 5.59$, $p < 0.001$), Level 4 ($M = 5.55$, $p < 0.001$), Level 5 ($M = 4.76$, $p < 0.001$) and Level 6 ($M = 5.31$, $p < 0.001$).

3. Additional Information and Considerations

The **REASONING TEST 20'** was developed primarily to guide Human Resource decisions for recruitment but its usage is manifold for organisational settings. It is important that the appropriate norms, scores and interpretations are used in order to aid selection, career development, coaching, and performance management. Accurate interpretation of psychometric tests helps both administrators and candidates to: a) put the assessment into context; b) effectively compare candidates across target capabilities; c) accurately identify areas for development; d) maximise fit between job requirements and candidate potentials; and e) promote the organisation as a fair and objective place to work. The current chapter gives guidelines for Human Resource practitioners to understand and successfully interpret candidate scores of the **REASONING TEST 20'** for different organisational objectives.

Report generation software for the **REASONING TEST 20'** provides graphs and descriptions for all the three factors that relate to the candidate's score. The report should only be interpreted by professionals who have completed *Central Test's* accredited training course and are familiar with this manual. Training in and knowledge of the assessment process are vital as scores should be interpreted in light of information regarding reliability, validity, standard error of measurement, and any accommodations to the assessment or assessment session that were made. Scores should also be presented within the context of the assessment and in a clear, meaningful and appropriate way for the target audience. It is also important to note that all the assessments by *Central Test* have been developed for Human Resource purposes only. Therefore, no clinical conclusions can be drawn from the results – not even from extreme scores.

Recruitment and selection is an important organisational function which can be very time-consuming and expensive. Furthermore, bad recruitment decisions have additional effects on organisational productivity, capability and sustainability. Therefore, it is recommended that prior to administering the **REASONING TEST 20'**, recruiters have a detailed job description and person specifications based on a job analysis.

Furthermore, the **REASONING TEST 20'** can be used in line with performance evaluations, which are a key component of employee development and success. Performance reviews should be fair and balanced assessments of an employee's achievements and capabilities, and should give the opportunity to discuss expectations and accomplishments.

Whilst the **REASONING TEST 20'** is effective in identifying areas for development, it should not be used to directly track an employee's performance progress. We therefore recommend the use of objective performance assessments or other evaluations in conjunction with the **REASONING TEST 20'**.

References

- Clauser, B. E., & Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice*, 17(1), 31-44. <http://doi.org/10.1111/j.1745-3992.1998.tb00619.x>
- Dewey, J. (1910/1933). *How we think: A restatement of the reflective thinking to the educative process*. Boston: Heath.
- Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72(1), 19-29. <http://doi.org/10.1037/0021-9010.72.1.19>
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56-83.
- Hanson, B. A. (1998). Uniform DIF and DIF Defined by Differences in Item Response Functions. *Journal of Educational and Behavioral Statistics*, 23(3), 244-253. <http://doi.org/10.3102/10769986023003244>
- Holtzman, S., & Vezzu, S. (2011). *Confirmatory Factor Analysis and Structural Equation Modeling of Noncognitive Assessments using PROC CALIS*. <http://www.lexjansen.com/nesug/nesug11/sa/sa07.pdf>
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi.org/10.1080/10705519909540118
- Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. *Scientific Software International*.
- Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E., (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context*, 6(1), 29-36.
- McLeary, C., & Cruise, P. (2012). Modeling Organisational Trust: Addressing Challenges of Measurement Invariance. In *Proceedings of the 11th European Conference on Research Methods* (p. 470). Academic Conferences Limited.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Pearson, K. (1895). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London: Mathematical, Physical and Engineering Sciences*, 186, 343-414. <http://doi.org/10.1098/rsta.1895.0010>

- Schmidt, F. L., & Hunter, J. E. (1996). Intelligence and Job Performance: Economic and Social Implications. *Psychology Public Policy and Law*, Vol. 2, 447-472.
- Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. In annual meeting of the Psychometric Society, Iowa City, IA, 758.
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times: Learning for life in our times*. San Francisco: John Wiley & Sons.
- Wilcox, S., Schoffman, D. E., Dowda, M., & Sharpe, P. A. (2014). Psychometric properties of the 8-item english arthritis self-efficacy scale in a diverse sample. *Arthritis*.
- Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada: Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia.

CONTACT US

Client Service: **customer@centraltest.com**

Training and Consulting: **consulting@centraltest.com**



www.centraltest.com

© 2020 by Central Test – Empowering Talent.
All rights reserved.