

TECHNICAL MANUAL

&

USER GUIDE

AVATAR

This manual has been created by the psychologists from the Research and Development Department of Central Test.

The assessment AVATAR and the Central Test logo are registered trademarks.

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

Table of Contents

Introduction	3
1. General Description and Applicability	5
1.1. Exploration Security	5
1.1.1. Exploration Factors	5
1.1.2. Security Factors	6
1.2. Control Adaptation	6
1.2.1. Control Factors	7
1.2.2. <i>Adaptation Factors</i>	7
1.3. Goal-focus People-focus	7
1.3.1. <i>Goal-focus Factors</i>	8
1.3.2. <i>People-focus Factors</i>	8
1.4. Main Areas of Application	9
1.5. Questionnaire Structure	9
1.6. Mode of Administration, Languages and Special Requirements	10
1.7. Social Desirability	11
1.8. Administration Instructions	12
1.9. Technical and Security Issues	13
1.9.1. <i>Respecting the Ethical Charter</i>	13
1.9.2. <i>Confidentiality and Security of User Data</i>	13
1.9.3. <i>Hardware Architecture and Server Availability</i>	14
1.9.4. <i>Computer-generated Reports</i>	14
1.10. Supply, Conditions and Availability	14
2. Conceptual and Practical Scale Development	16
2.1. Theoretical Background	16
2.1.1. <i>The History and Value of Psychometric Testing to Personnel</i>	16
2.1.2. <i>Different Approaches to Personality: Type vs. Trait</i>	17
2.1.3. <i>Personality Theories used in AVATAR</i>	17
2.1.3.a. <i>Theory of Psychological Types - Carl G. Jung (1921)</i>	17
2.1.3.b. <i>Five Factor Model - Costa and McCrae (1987)</i>	18
2.1.3.c. <i>HEXACO Model of Personality - Lee and Ashton (2004)</i>	19
2.1.3.d. <i>DISC Model - William M. Marston (1928)</i>	19
2.1.3.e. <i>Myers-Briggs Type Indicator - Briggs & Briggs-Myers (1980)</i>	21
2.1.4. <i>Emotional Intelligence</i>	22

2.1.5. <i>The Relationship between Personality and Emotional Intelligence (EQ)</i>	22
2.1.6. <i>Constructual Mapping of AVATAR Factors</i>	23
2.2. Development of AVATAR	25
2.2.1. <i>Dimensions, Factors and Item Construction</i>	25
2.2.2. <i>Profile Calculation Method</i>	26
2.2.3. <i>Ipsative Response Format</i>	26
2.2.4. <i>Likert Response Format</i>	27
3. Scoring	29
3.1. The Relationship between Raw Scores and Norms	29
3.2. Transforming Raw Scores to Standardised Scores	31
3.3. Standard Error of Measurement	31
3.4. Choosing an Appropriate Norm Group	32
3.5. Creating In-house Norms	32
3.6. Link between Scores and Report Comments	33
4. Sample Descriptives	34
5. Norms	36
6. Reliability	54
6.1. Internal Reliability	54
6.2. Test-retest Reliability	55
7. Validity	58
7.1. Concurrent Validity	58
7.1.1. <i>The NEO-PI-R Scale</i>	59
7.2. Construct Validity	62
7.2.1. <i>Inter-factor Correlations</i>	62
7.2.2. <i>Confirmatory Factor Analysis</i>	64
7.2.3. <i>Consequential Validity</i>	66
8. AVATAR and Equal Opportunities	67
8.1. Measurement Equivalence	67
8.1.1. <i>Differential Item Functioning (DIF)</i>	68
8.1.1.a. <i>DIF Analysis: Gender</i>	69
8.1.1.a. <i>DIF Analysis: Age</i>	71
8.1.1.c. <i>T-tests and ANOVAs</i>	72
References	77

Introduction

AVATAR is a typological personality assessment designed to assess an individual's personality across a spectrum of characteristics. The test's theory is derived from William Marston's DISC Model (1928), Carl Jung's (1921) Theory of Personality Types, and the MBTI model (Briggs-Myers & Myers, 1980). **AVATAR** gives the user a reliable classification of their personality type across three dimensions; *Exploration - Security, Adaptation - Control*, and *People-focus - Goal-focus*. On top of this, **AVATAR** provides the user with a score on a subscale measuring *Emotional Stability*.

Today's labour market has changed dramatically over the course of the past several decades. Little more than a generation ago, lower skill level jobs did not frequently require workers to have academic or social competencies. Today, however, that premise is not easily defensible. During the past generation, the workplace has changed dramatically. The job opportunities that, stereotypically, require little proficiency are rapidly changing with the shift away from a 'production driven' economy toward a 'service-driven' economy. This shift has created 'new' workplace demands, which now requires that prospective employees be more adept than those of prior generations (Thompson, 2001). Thus, employers now deem it essential that the personality traits, motivations and values, and skills they seek in employees exist.

As a result, the objective of **AVATAR** was to create a typological personality assessment that is accessible to both individuals for personal use, and companies in achieving their employment and self-development procedures. **AVATAR** serves a dual purpose:

- I. For organisations, this assessment provides a quick and comprehensive profile of their applicants or employees
- II. For individuals, it marks the beginning of their journey towards understanding themselves better

Some benefits of **AVATAR** that best distinguish it from other personality tools in the market are as follows:

- I. Typological personality resulting in an **AVATAR** 'type' presented on the report
- II. More contemporary items to reflect relevant personality characteristics
- III. Average completion time of 10-12 minutes
- IV. Developed using psychometric techniques such as ipsative and normative scaling items to guard against social desirability and impression management
- V. User-friendly and customisable reports that can easily be tailored by the administrator
- VI. Practical and easy-to-understand questionnaire and report

This Technical Manual and User Guide enables the user to understand the theoretical foundations on which **AVATAR** is based, the factors and dimensions that comprise it, and the manner in which the questionnaire is designed and structured. This manual and guide help the user familiarise themselves with the stages of test administration, report interpretation and applications in various organisational contexts. Chapters 1, 2, 3 and 8 are specifically tailored for HR practitioners, while chapters 4 to 7 comprise detailed technical and psychometric information.

1. General Description and Applicability

AVATAR classifies the test-taker's personality type within three opposing dimensions of personality: **Exploration vs. Security**, **Control vs. Adaptation**, as well as **Goal-focus vs. People-focus**. The assessment identifies an individual's behavioural inclinations within each of the three dimensions, and assigns them a psychological 'type' after taking the assessment. These dimensions then each encompass three opposing factors, as shown in Table 1.

Table 1. AVATAR Dimensions and their Corresponding Factors.

EXPLORATION	vs	SECURITY	CONTROL	vs	ADAPTATION
Spontaneous		Organised	Persuasive		Impressionable
Novelty-seeking		Traditional	Directive		Participative
Experimental		Cautious	Driving		Easy-going

GOAL-FOCUS	vs	PEOPLE-FOCUS
Individualistic		Humanistic
Vigilant		Trusting
Autonomous		Affiliative

The individual's 'type' will be determined by the dimensions they endorse. A detailed description of the dimensions, along with their factors, is provided below.

1.1. Exploration | Security

This dimension centres around the amount of change a person seeks, or enjoys, in their life. The explorative individual, on the one hand, wants to explore the environment around them. With a desire to experience new situations, they are likely to want to test new things out and involve themselves in events that may have uncertain outcomes. The security-seeking person, on the other hand, may prefer the safety, comfort and assurance that comes from being in situations that offer familiarity. Rather than seeking change, they look for the security that is provided from structured or stable environments, where there is less room for things to go wrong.

1.1.1. Exploration Factors

- I. **Spontaneous** - This factor refers to the level of planning an individual seeks. The spontaneous person is not likely to want to plan or make a schedule for their day-

to-day life. Instead, they have a tendency towards creating space to improvise, and feel that rigid plans might cause them to miss out on opportunities.

- II. **Novelty-seeking** - This factor refers to the level of novelty or familiarity that a person looks for. The novelty-seeking individual tends to seek out new experiences and unfamiliar environments. They welcome change, and are likely to prefer being around a variety of people and exposed to situations that differ from what they already know.
- III. **Experimental** - This factor refers to the willingness to experiment with unfamiliar ways of doing things. The experimental person tends to question the way things are usually done and strives to find better ways of doing things. As a result, they are likely to take risks in order to test new methods and they welcome change and innovation for the better.

1.1.2. Security Factors

- I. **Organised** - This factor refers to the level of planning an individual seeks. The organised person is likely to plan in advance, and create a structured agenda for their day-to-day life. They prefer to have stability and certainty over their schedule, and may feel that if they do not plan ahead, things may not go well.
- II. **Traditional** - This factor refers to the level of novelty or familiarity that a person looks for. The traditional individual tends to stay within their comfort zone, seeking out familiar environments without too much disruption from variety or new experiences.
- III. **Cautious** - This factor refers to the willingness to experiment with unfamiliar ways of doing things. The cautious person is likely to stay away from taking unnecessary risks, and instead stick to methods that have been tested and proven to be effective. They are likely to be consistent, preferring certainty over risk and experimentation.

1.2. Control | Adaptation

This dimension focuses on a person's tendency to want to either control or adapt to the environment around them. An individual that is high on control will have a strong desire to influence their environment and shape events around them. They are fast-paced individuals and tend to want to orient situations, whether this be through taking charge, persuading others or navigating events towards the outcome that they look for. On the other hand, an individual that is high on adaptation is likely to comply and give in to their environment. This means that they are more inclined to either let events follow their natural course, take a more passive approach to avoid confrontation, or let others take a role of authority.

1.2.1. Control Factors

- I. **Persuasive** - This factor refers to the extent to which people will want to convince others of their ways. The persuasive person tends to modify their speech and behaviour according to who they are talking to, in order to convince them of something. They are likely to want to persuade others of their viewpoints instead of adjusting their own opinions to avoid disagreement.
- II. **Directive** - This factor refers to the extent to which an individual seeks to direct other people. The directive individual has a tendency to take charge and impose their methods on others. They are likely to believe that their way is the best way and can sometimes seem authoritative. As a result, they prefer to be leaders than followers, where they are in a position to instruct others.
- III. **Driving** - This factor refers to the extent to which an individual involves themselves in situations around them. The driving individual can be described as fast-paced; they want to get things moving, and have a tendency to check that things are going the 'right' way, and to push things in the direction that they desire.

1.2.2. Adaptation Factors

- I. **Impressionable** - This factor refers to the extent to which people will want to convince others of their ways. The impressionable person tends to be flexible with their opinions, and may be predisposed to adjusting their point of view in the face of disagreement, in order to avoid conflict. As a result, they are more likely to be persuaded than to try to convince others.
- II. **Participative** - This factor refers to the extent to which an individual seeks to direct other people. The participative individual has a tendency to be obliging and compliant in their interactions, and avoid directing or imposing their methods on others. They are likely to prefer to follow the people in charge and let another individual lead.
- III. **Easy-going** - This factor refers to the extent to which an individual involves themselves in situations around them. The easy-going individual is not solely focused on making things happen quickly or in a particular way, and they tend to let things work themselves out naturally. They avoid being heavily involved in everything around them, and instead let things take their natural course.

1.3. Goal-focus | People-focus

The dimension proposes that a person possesses a natural orientation towards either goals or people. A goal-focused individual will direct their attention towards themselves and their objectives. Likely to hold a realistic outlook and be wary of the environment

around them, they tend to act independently and orient their behaviour around the achievement of their goals. The people-focused individual on the other hand, is likely to behave both for and in response to the people around them. With an idealistic and sociable outlook on life, they trust and see the good in the people and environment around them. As a result, their behaviour tends to stem largely from how their actions may affect others.

1.3.1. Goal-focus Factors

- I. **Individualistic** - This factor refers to how people place themselves in relation to others. Individuals high on individualism tend to have a realistic outlook on life, and therefore invest their time and attention into reaching their own goals and looking after their own needs. They are driven to find the most efficient route to reaching their goals, even if it goes against other people's interests.
- II. **Vigilant** - This factor refers to the level that people trust their environment and the people around them. Individuals high on vigilance tend to perceive their environment as unfavourable, and may be wary of the intentions and motives of the people around them. As a result, they are likely to steer their focus away from people and instead concentrate on their personal goals and targets.
- III. **Autonomous** - This factor refers to how much people like to invest in social relationships. Individuals with an autonomous personality tend to seek solitude and therefore, opt for a more independent way of living that is less focused on forming bonds with other people, and more focused on concrete goals. Sometimes described as self-reliant, autonomous individuals do not have a great desire to spend a lot of time socialising with others.

1.3.2. People-focus Factors

- I. **Humanistic** - This factor refers to how people place themselves in relation to others. Individuals that are high on humanism tend to have an idealistic outlook on life, and want to put others first. Helping others is their most important value, even if it means overlooking their own interests. With a strong desire to improve human welfare, their actions tend to revolve around the needs of others.
- II. **Trusting** - This factor refers to the level that people trust their environment and the people around them. Individuals high on trust are likely to perceive their environment as favourable, having faith in the people around them and seeing the good in their intentions and motivations.
- III. **Affiliative** - This factor refers to how much people like to invest in social relationships. Individuals high on affiliation have a strong interest in forming bonds and getting to know others. They like to meet new people, and may be

described as 'warm' due to their friendliness and the ease at which they can build relationships and make others feel comfortable around them.

1.4. Main Areas of Application

AVATAR is intended for the general public, as well as HR professionals and career counsellors. Its classification of personality types is intended to be used for personal assessment with regards to career selection, recruitment, and personal development. In the same way, it can be used by recruiters for applicant selection, as well as by employers to aid decision-making in career progression or development of employees.

AVATAR can be combined with other personality questionnaires such as *Central Tests VOCATION* and **PROFESSIONAL PROFILE 2**, in order to acquire a more precise and comprehensive profile of the individual as determinants for employment.

Table 2. Possible Applications of AVATAR with Other Assessments.

Additional test	Objectives	Advantages
PROFESSIONAL PROFILE 2	Career profiling / development / orientation	In-depth information about career personality fit
VOCATION	Career orientation	In-depth evaluation of career interests

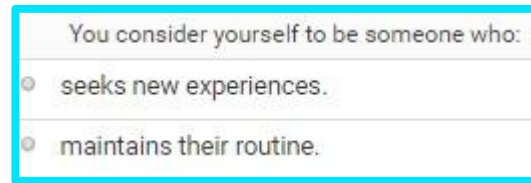
1.5. Questionnaire Structure

AVATAR consists of 65 core questions and an additional 8 questions measuring the candidate's emotional quotient (EQ). The core questions are presented in a forced choice format, whereby each question consists of two general statements that require the candidate to choose one option that best represents their behavioural orientation. Each statement assesses a different component of the three dimensions measured in the assessment. This means that each factor is measured against its opposing factor seven times. The core questions are paired in such a way that each response option assesses a different factor, which results in a more realistic measurement of human personality. Such an item structure also reduces the likelihood of respondent fatigue and response bias; forcing respondents to choose one of two options safeguards against impression management and the candidate's attempts to fake results.

Alternatively, the EQ items use a Likert scale response type, allowing respondents to position themselves in relation to the statement. On the 4-point Likert scale, respondents need to specify their level of agreement or disagreement with each item (1 = Strongly

Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree). Figures 1 and 2 depicts the questionnaire answering structure of **AVATAR**.

Example of a Core Item



The screenshot shows a questionnaire item titled "You consider yourself to be someone who:". Below the title are two radio button options: "seeks new experiences." and "maintains their routine." The entire interface is enclosed in a light blue border.

Figure 1. Questionnaire Structure of AVATAR.

Example of an Emotional Stability Item



The screenshot shows a questionnaire item titled "Generally, you do not lose your temper easily.". Below the title are four radio button options: "Strongly Agree", "Agree", "Disagree", and "Strongly Disagree". The entire interface is enclosed in a light blue border.

Figure 2. Questionnaire Structure of AVATAR.

1.6. Mode of Administration, Languages and Special Requirements

AVATAR is administered online and works with all modern computer operating systems including Windows, MAC-OS and Linux. A PDF version of the questionnaire is available, if required.

AVATAR can be administered remotely and without supervision. However, should standardised group assessments be required, this can easily be accommodated. Reports are solely generated online. Therefore, if the test has been administered in paper-format, an administrator must manually enter candidate responses into the online platform of the questionnaire to generate candidate reports.

The assessment is currently available in English and will soon be available in French and Spanish. For more information on language availability, please contact the *Central Test Consulting Service* (consulting@centraltest.com).

AVATAR was designed to be easily used by individuals from diverse backgrounds and by those with a range of special needs. All test-takers are able to adjust visual screen specifications based on individual requirements. For example, respondents with visual

impairments or those requiring a larger font size for comfort, may increase the font size by enlarging the computer screen's display (by pressing "Ctrl" and "+"). Similarly, respondents are able to access test instructions throughout the assessment. It is also possible to navigate through questions, or change answers, as desired by selecting on-screen prompts. Alternatively, for blind candidates or those with severe visual disabilities, items may be read aloud by a trained administrator. Individuals with hearing impairments are not excluded from taking the questionnaire since all content is presented visually. Although the average time for completion is 10-12 minutes, **AVATAR** has no time limit. Therefore, individuals with learning disabilities are able to take the time they need. Administrators are encouraged to consider whether a candidate who needed an unusual amount of time to take the test may have special needs. Please contact the test publisher for guidance on administering the questionnaire to individuals or groups with special requirements.

1.7. Social Desirability

Social desirability is the tendency to present oneself in a manner that will be viewed as favourable by others. When done unintentionally, it is referred to as self-deception. When done intentionally, it is referred to as impression management (Tett & Simonet, 2011). As people are social by nature, they are generally inclined to seek some degree of social acceptance. Unlike aptitude tests, the results of questionnaires like **AVATAR** are affected by impression management. An individual's attempt to present himself/herself in a favourable light negatively impacts test results, whether the questionnaire is completed online or in a paper-based format (Dodou & de Winter, 2014).

Candidates are usually tempted to answer in a way that would impress the recruiter even though there is no right or wrong answer. Candidates may provide responses according to what they feel others may expect to fulfil a role. As a result, social desirability is a significant risk to the validity of assessments (Mueller-Hanson, Heggstad, & Thornton III, 2003). If social desirability was equivalent across all candidates, then test administrators would be able to factor it into the scoring and interpretation of results. However, faking differs across individuals, resulting in distortion of scores.

AVATAR actively guards against social desirability with the use of an ipsative (forced choice) questionnaire format. Research has found ipsative measures to be more resistant to distorted responses compared to more commonly used Likert scales (Bowen, Martin, & Hunt, 2002; Kluger, Reilly, & Russell, 1991). Not only does the ipsative format reduce the candidates' ability to inflate their scores, but items are also combined in such a manner that makes it difficult for candidates to conceive a response strategy, as items are paired with equally appealing response options.

1.8. Administration Instructions

In most cases, **AVATAR** will be taken remotely through online access, and without supervision. However, in the case of taking the assessment within a supervised setting, the following guidelines should be applied.

Although instructions for completing **AVATAR** are presented on-screen, it is important that administrators communicate with the candidates. Prior to starting the assessment, instructions should be read aloud by the administrator, and candidates should be given the opportunity to ask questions before taking the assessment. Administrators should verify that candidates understand what they are required to do prior to starting the assessment.

Test instructions are the same for every individual. Standardised instructions ensure that all participants are exposed to similar conditions during the assessment process to ensure that it is fair and objective, and that assessment conditions do not directly or indirectly give any individual or group an advantage.

Further tips:

- If candidates have questions after the assessment session has begun, the administrator should try to answer them without disturbing other candidates.
- If a candidate does not understand a question, the administrator should answer carefully and avoid explaining the meaning of words or questions in his/her own words in order to not influence the candidate.
- If an administrator suspects that a candidate's lack of understanding is due to language proficiency, the administrator should ask the candidate to take the assessment in the language that they are most comfortable with.
- If an administrator suspects that a candidate is being over cautious when responding, the candidate should be encouraged to endorse the response they feel is most honest to them.

Candidates receive a notification to take the questionnaire via email with an embedded web link and login details. Web links to the questionnaire are password protected and unique to the candidate's account. It is possible to set expiration dates in order to ensure that candidates complete the questionnaire within the specified time-frame. It is also possible to configure the invitation in a way that ensures the assessment will be completed at the same IP address once the questionnaire has been accessed from a particular IP address.

AVATAR has no time restriction. However, candidates generally take about 10-12 minutes to complete the questionnaire. For monitoring purposes, completion times are recorded to allow administrators to see if a candidate rushed through the questionnaire or needed an unusual amount of time to complete the test. Candidates are encouraged to answer

questions spontaneously, without rushing, in order to avoid spending too much time thinking about each statement as there is no right or wrong answer.

1.9. Technical and Security Issues

This chapter details guidelines governing the security of administering **AVATAR**. These guidelines ensure that every administration of **AVATAR** is a pleasant, objective and fair process. Security specifications ensure that candidate information is protected both during administration and after assessment completion.

1.9.1. 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 may not be forwarded to a third person without the explicit consent of the test candidate. A candidate has the right to access their results. The assessor must grant them access automatically or when the candidate requests it.

The candidate has the right to request that all his/her personal data be permanently deleted. To do this, he/she must contact either the organisation or *Central Test* directly, who will delete his/her data.

1.9.2. Confidentiality and Security of User Data

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 are not visible to companies. A URL rewriting module (secure link generator) is installed and has a transparent parameter passing for the user.

Browsing and data exchange via the *Central Test* application can be secured and encrypted using SSL. Secure browsing is guaranteed by the https protocol, which is 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 is an irrefutable evidence of our serious attitude towards security and protects all *Central Test* assessments from phishing attacks. 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. Access to questionnaires may be limited in terms of numbers and duration.

1.9.3. Hardware Architecture and Server Availability

Central Tests architecture consists of a network of 10 physical dedicated servers providing high-quality service.

Each server is housed in a specialised data centre featuring:

- Dedicated lines
- A guaranteed bandwidth of 100Mbps (per server)
- Redundant energy sources
- 4-hour Equipment Restoration time
- 2-hour Incident Intervention time
- Monitoring and assistance 24/7/365

Access to the server administration is restricted to an IP address and is password-secured. Data exchanges are secure.

Central Tests application has been tested on and is compatible with the following browsers:

- Internet Explorer 8 or above
- Firefox 21 or above
- Google Chrome 30 or above
- Safari 7 or above

1.9.4. Computer-generated Reports

Reports can be generated within minutes of candidates completing the questionnaire. The **AVATAR** report is user-friendly and customisable. Report content can be tailored depending on the needs of the administrator. Separate individual reports may be configured for the candidate and the company, recruiter, coach or psychologist. The reports are also available in PDF format.

1.10. Supply, Conditions and Availability

Conditions of supply and availability for **AVATAR** vary from country to country. However, the following points generally apply:

- **AVATAR** is available for Psychologists, HR practitioners and coaches.
- Training is compulsory prior to use and distribution.
- **AVATAR** can be ordered online, or by contacting any *Central Test* customer service representative or local partner worldwide (customer@centraltest.com).

- For information on **AVATAR** prices and related products and services, contact any *Central Test* customer service representative or local partner worldwide.
- For more information on the training, please contact *Central Test's* customer service or one of *Central Test's* partners.

2. Conceptual and Practical Scale Development

AVATAR was developed using robust theoretical concepts, practical and relevant work requirement theories, as well as advanced scientific analyses. The nine opposing factors assessed in **AVATAR** are the result of a deductive test construction method, based on previously established theories such as Jung's Psychological Types (1921) and the Big Five Model (Costa and McCrae, 1985). This method resulted in the development of a short, administration-friendly, work-relevant, and valid assessment. All items in the item bank were piloted and pretested to ensure that **AVATAR** meets the highest standards.

2.1. Theoretical Background

2.1.1. The History and Value of Psychometric Testing to Personnel Assessment

Historically, the Latin origin of the word "personality" (persona) referred to masks worn to represent roles that actors interpreted. Today, the term covers psychological characteristics common to all human beings and is used to explain human nature and individual differences.

Considered the father of psychometrics, Sir Francis Galton (1884) was the first scientist to analyse the structure of personality using a lexical hypothesis. The lexical hypothesis is one of the most widely used underlying theories of personality and rests on two arguments: a) personality characteristics that are most important in people's lives will eventually become a part of their language, and b) more important personality characteristics are more likely to be encoded within language as a single word. This lexical hypothesis was used by Gordon Allport (1937) to generate a list of 17,953 unique terms to describe personality. Allport was the first to use the term "personality" in its modern application and produce an understandable classification of human traits that led to the clustering and later factor structure of personality traits by Raymond Cattell (1957) and Costa and McCrae (1985).

In the 21st century business world, psychometric measurement tools are increasingly used in the workplace as they help to predict behaviour and performance in a reliable way.

Across the globe, organisations use personality and ability assessments to make personnel decisions for recruitment, selection, promotion and development. Unlike judgements based on subjective perceptions, such as those derived from traditional job interviews, psychometric assessments offer a more objective, fair, reliable, valid and legally-defensible approach to identifying desired work behaviours. The measurement of personality represents a key component in identifying individuals who are best suited for particular situations, contexts and planned objectives as individual values, interests, preferences and abilities represent core composites of personality. Therefore, the use of

personality assessment is integral to any process where the individual is central to productivity (Cruise, 2012).

2.1.2. Different Approaches to Personality: Type vs. Trait

The two main theories for attempting to systematically categorize individuals in psychology are the trait and type theories. Type theories categorise personality as a set of interrelated elements that make an individual unique. People are assigned to discrete categories of personality that are different to each other (e.g., introvert vs. extrovert). Trait theories on the other hand, regard personality as degrees of expression along a gradient of behaviour. A trait could be defined as the tendency to act in a certain way and could be expressed by the behaviour of the individual in diverse situations (Nadeau, Acier, Miranda, Pervin & John, 2004).

For example, a type theorist would see introverts and extroverts as two different types of people, whereas a trait theorist would claim that extraversion and introversion are the two ends of a continuum and it is possible for people to fall somewhere in the middle. Having very different views into personality, both approaches can be useful for different areas of application. Traits express preferences that allow us to predict behaviour, understand changes in behaviour over time, and identify how to develop behaviours with intervention. As type theories regard opposing categories as two distinct classes, they are considered appropriate assessments for personal and team workplace development, career guidance, onboarding, and coaching. For recruitment purposes, trait theories are recommended over type theories as they give a more detailed description of an individual's potential instead of allocating unique individuals to a limited number of personality types.

Within **AVATAR**, users are provided with an assessment that incorporates the advantages of the two approaches by providing both a typological and trait-based interpretation of a candidate's test results. That way users can choose between two different reports according to their personal preferences and the field of application, or even see both reports to gain a deeper insight into the personality of a candidate.

2.1.3. Personality Theories used in AVATAR

2.1.3.a. Theory of Psychological Types – Carl G. Jung (1921)

In 1921, Carl Gustav Jung published his work "Psychological Types" in which he discusses the problem of types in literature, religion and philosophy. He took the view that much of the behaviour that appears to be random is, in fact, orderly and consistent. This, he argued, is due to a few basic and overseeable differences in mental functioning, and therefore his theory proposes that humans can be categorised according to two attitude types: Extraversion and Introversion, which are representative of an individual's attitude towards the world around them. While an introvert's main interests are in the inner world of ideas and thoughts, an extravert is more involved in the outer world of objects and people. Of course, no one is limited to just one attitude type, but instead is likely to have

a preference for one more than the other. As such, an introvert may feel more comfortable in a profession where they are able to work inside of their head, reflecting and generating ideas, whereas an extrovert may prefer a job that allows them to be in interaction with people and the environment around them.

In addition to these attitude types, Jung categorizes humans according to four “functioning types”, which describe how people perceive their environment - *Sensation* and *Intuition*, and how they judge what they perceive - *Thinking* and *Feeling*. People who process information through *Sensation* will pay close attention to details, whereas an individual who perceives through *Intuition* will rely more on broad patterns. Here, again, Jung argues that the two methods of perceiving compete for a person’s attention and that most people, although able to use both patterns, will enjoy one more than the other. The same is true for *Thinking* and *Feeling*. When an individual prefers *Thinking* over *Feeling*, he or she will be more likely to base their judgements on logic and factual information. The opposite is true for someone who prefers *Feeling* as they base their judgements on personal value (Arnau, Green, Rosen, Gleaves & Melancon, 2003).

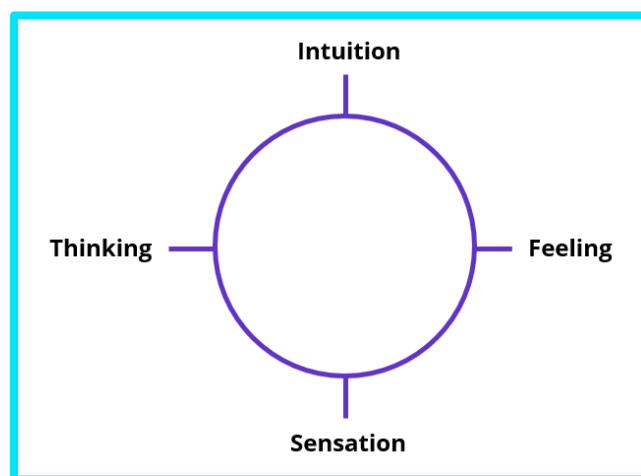


Figure 3. Jung's Model of Psychological Types.

2.1.3.b. Five Factor Model – Costa and McCrae (1987)

The Five Factor Model (FFM) of personality is considered the standard in trait personality measurement as it represents well-researched global personality traits. The model is comprised of: *Openness* - the extent to which an individual is intellectually curious, has insight and is imaginative; *Conscientiousness* - the extent to which an individual is reliable, persistent, and ambitious; *Extraversion* - the inclination to be sociable, desiring to be among others, thus requiring excitement and stimulation; *Agreeableness* - the tendency to be humanistic, cooperative, and trusting of others; and *Neuroticism* - the inclination to feel negative emotions including fear, guilt, sadness, and anger (John & Srivastava, 1999; Major, Turner & Fletcher, 2006; Mooradian, Renz, & Matzler, 2006). The five dimensions are sometimes named by their polar opposites (Soto & Jackson, 2013) namely

Antagonism, Lack of Direction, Introversion, Emotional Stability and Closedness to Experience, respectively.

2.1.3.c. HEXACO Model of Personality – Lee and Ashton (2004)

Lee and Ashton (2004) proposed a six-dimensional, trait-based model of personality founded on evidence of previous research on trait-based personality models. The model includes all dimensions measured by the Five Factor Model as well as the additional dimension of *Honesty-Humility (H)*. Within the model, dimensions are classified as: *Honesty-Humility (H)* – measures facets of sincerity, fairness, greed avoidance, and modesty; *Emotionality (E)* – assesses fearfulness, anxiety, dependence, and sentimentality; *Extraversion (X)* – identifies social self-esteem, social boldness, sociability, and liveliness; *Agreeableness (A)* – measures forgiveness, gentleness, flexibility, and patience; *Conscientiousness (C)* – assesses organisation, diligence, perfectionism, and prudence; and *Openness to Experience (O)* – examines an individual's tendency towards aesthetic appreciation, inquisitiveness, creativity and unconventionality. The HEXACO model is illustrated in Figure 4.

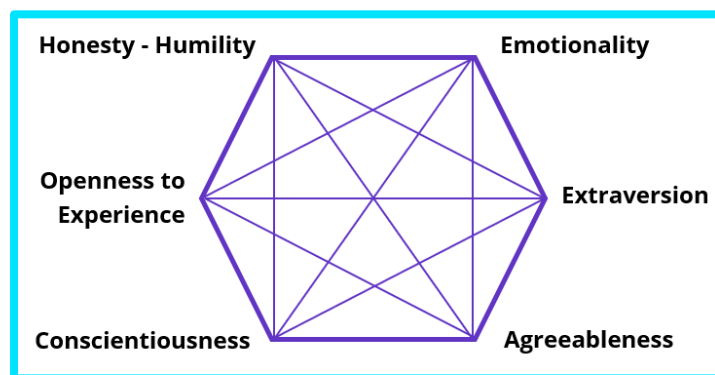


Figure 4. The HEXACO Model of Personality.

For the development of **AVATAR**, facets measured by the (H) dimension of HEXACO were important as an individual's values and expression of said values significantly influence role, team and organisational fit as well as his/her approach to making decisions.

2.1.3.d. DISC Model – William M. Marston (1928)

William M. Marston was an American psychologist who developed a personality inventory in order to investigate how human emotions lead to differences in behaviour, and how this behaviour changes over time. In 1928, Marston published his work "Emotions of normal people", which defined the DISC Theory. The Theory is named after its four central types, labelled by Marston as *Dominance, Inducement, Submission* and *Compliance*. It is possible to find behavioural styles that represent only one factor, e.g., a highly dominant individual, though even more often one will find that an individual has a primary factor as well as a secondary and tertiary one. Therefore, to gain a comprehensive insight into someone's personality, it is important to consider their primary, secondary, tertiary and even absent personality traits. It is important to note that none of the types are better or

worse than any other; that they simply reveal a person's preferences in various environments, their communication with those around them, and their abilities.

On the one hand, people with high *Dominant* styles enjoy dealing with problems, encountering challenges and are quick to make decisions. They are often described as forceful, strong-willed, ambitious, pioneering, discerning, and leading. On the other hand, people who score low on "D" are rather low-keyed; they take their time to commit a decision but are also fine to cooperate and follow instructions. They tend to be conservative, undemanding, mild, cautious, and peaceful. Although originally named *Inducement*, the next factor is often referred to nowadays as *Influence* and relates to an individual's behaviour in social situations and their style of communication. People with "I" as their primary factor are described as friendly, open-minded, sociable, convincing, persuasive, and magnetic. They possess very good social skills and can easily get into contact with others and win them over without big effort. In opposition to this, low "I" intensity scores represent those who are reflective, factual, logical, and critical. When convincing others, they do so not with their charisma but more with data and facts. The factor *Submission*, or better known as *Steadiness*, refers to an individual's approach on tasks. While high "S" styles are described as patient, persistent, thoughtful, stable, consistent, and rather unemotional, low "S" styles represent the opposite, being rather restless, impatient and impulsive. People with *Steadiness* as their dominant style prefer a stable environment while low "S" styles are likely to value variety and change. The final DISC factor is named *Compliance* and refers to the amount of structure and organisation a person values. The simplest way to describe high "C" styles is "rule-oriented". They are careful, accurate, tactful and unlikely to break the rules, while low "C" styles are rather self-willed, unsystematic, independent and arbitrary. Marston's DISC model is shown below, in Figure 5.

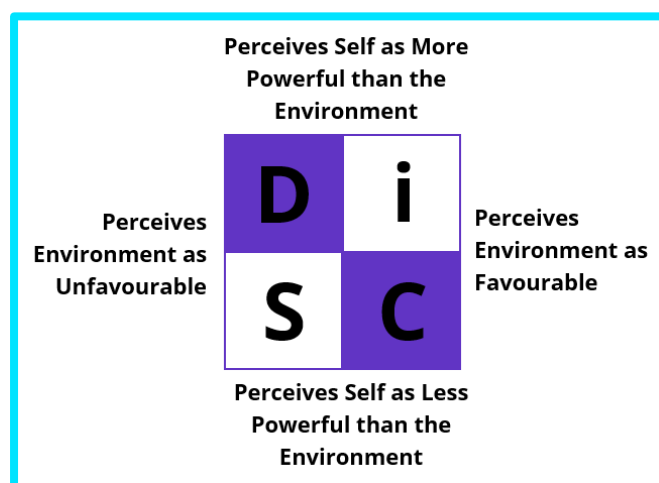


Figure 5. Marston's DISC Model.

Marston argued that the four behavioural types result from people's sense of self and their interaction with the environment. As a result, he proposed that the four central types are built upon two axes. The first axis relates to whether an individual sees their

environment as favourable or unfavourable. The second axis is whether they perceive themselves as having control or lack of control over their environment. Dominant and Influential individuals see themselves as rather powerful and will therefore try to change their environment while Steady and Compliant Individuals view themselves as less powerful and try to accommodate themselves to given circumstances. High “D” styles tend to see their environment as unfavourable and will try to change the status quo as much as they can. Compliant people see their environment as equally unfavourable, but unlike the “D” types, they are unlikely to be willing to break the rules. Influential and Steady people both see their environment as favourable.

2.1.3.e. Myers-Briggs Type Indicator – Briggs & Briggs-Myers (1980)

Following the development of Carl Jung’s Theory of Psychological Types (1921), Katherine Cook Briggs and Isabel Briggs-Myers constructed the MBTI to measure the difference in preferences among individuals in terms of how they perceive the world around them and make decisions. The MBTI sorts some of these psychological differences into four opposing pairs, or dichotomies, with a resulting 16 possible psychological types. People innately prefer one overall combination of type differences. The 16 types are typically referred to by an abbreviation of letters (S, N, T, F, E, I, P, J) as shown below.

Table 3. Four Pairs of Dichotomies in MBTI.

	Subjective	Objective
Deductive	Sensing/INtuition	Extraversion/Introversion
Inductive	Thinking/Feeling	Perception/Judgment

The interests, values, needs, and habits of the mind that naturally result from any set of preferences tend to produce a recognizable set of traits and potentials (Briggs-Myers & Myers, 1980). These preferences are considered as key components for workplace success as people will be happier and far more productive in a job that is in line with the perceptions and judgements that they tend to make. For example, an individual who is mostly introverted, as well as high on thinking and intuition, may be very well equipped for a job that requires them to analyse problems, discover solutions, and generate new ideas, such as theoretical science. Someone who is rather extraverted and high on feeling and sensing represents the complete opposite. Being very concerned with the well-being of others and relying a lot on factual information, this person may choose health professions, such as nursing, where they can talk to others and be involved in the action. It is possible for an individual to combine any attitude or function type, which is why this model classifies people into 16 personality types.

2.1.4. Emotional Intelligence

One of the first theories of emotional intelligence was an ability model by Salovey and Mayer (1990). Salovey and Mayer (1990) identified emotional intelligence as the “ability to monitor one’s own and other’s feelings and emotions, to discriminate among them, and to use this information to guide one’s thinking and action”.

Other theories, such as the Bar-On Emotional-Social Intelligence (ESI) model (Bar-On & Handley, 2003) and the Goleman model (Goleman, 1998), widened the scope of investigation to include components not closely related to cognitive abilities. The ESI model conceptualises emotional intelligence more openly than Mayer and Salovey (1997), as “an array of non-cognitive capabilities, competencies, and skills that influence one’s ability to succeed in coping with environmental pressures and demands” (Bar-On, 1997). Their model includes 15 interrelated social and emotional competencies, skills and facilitators categorised into 5 domains: 1) *Intrapersonal skills*, the ability to be aware of and be able to express oneself, including Self-regard, Emotional Self-awareness, Assertiveness, Independence and Self-actualisation; 2) *Interpersonal skills*, to be socially aware and successfully manage interpersonal relationships, includes Empathy, Social Responsibility, and Interpersonal Relationship; 3) *Stress management*, the regulation and management of emotions, including Stress Tolerance and Impulse Control; 4) *Adaptability*, the management of change, including Reality-Testing, Flexibility, and Problem-Solving; and finally, 5) *General mood*, concerned with self-motivation, includes Optimism and Happiness. The ESI is measured by the Emotional Quotient Inventory (Bar-On, 1997).

AVATAR adopts a trait approach similar to the ESI. This approach was chosen on the basis that it is important to have an Emotional Intelligence (EQ) subscale that is practical and informative in business environments. For example, the traits selected for the EQ subscale in **AVATAR** are all applicable to the workplace and will allow employers to gain a greater understanding of their employees’ emotional intelligence, how they will work, and how they will interact with others.

2.1.5. The Relationship between Personality and Emotional Intelligence (EQ)

According to Saville and Holdsworth (1999), personality refers to a person’s typical or preferred way of behaving, thinking and feeling. Therefore, personality is the particular combination of emotional, attitudinal, and behavioural response patterns of an individual, while the characteristics of personality include mood, attitude, dominance, sensitivity, aggression and other various traits that describe an individual’s actions or emotions.

In contrast, emotional intelligence as defined by Salovey and Mayer (1990) is “the ability of an individual to understand his own and others’ emotions and feelings, to distinguish between them and to use this knowledge in his decision-making process and actions”.

Goleman (1998), who popularised the term emotional intelligence, emphasised that emotional intelligence is an important factor in determining personal success.

A study by Ghiabi and Besharat (2011) examined the relationship between personality dimensions and emotional intelligence. Their findings showed that EQ could be positively and negatively predicted by extraversion and neuroticism, respectively. The finding by Ghiabi and Besharat (2011) supports earlier research, showing that overall emotional intelligence was significantly related to personality dimensions.

Another study done by Grehan, Flanagan and Malgady (2011) exploring the relationship between personality traits and emotional intelligence showed a relationship between the personality trait Conscientiousness and EQ. Additionally, an earlier study by Brackett and Mayer (2003) found a significant correlation between EQ and personality (extraversion, agreeableness, conscientiousness). Another study by Slaski and Cartwright (2002) also examined the relationship between EQ and job performance, with their findings indicating that those with higher EQ scores had significantly lower stress and distress.

To date, more and more companies are placing importance on the onboarding and professional development of their employees (Yusooff et al., 2013). Two of the most important tools used for this purpose are assessments of personality and emotional intelligence. Hence, with its combination of personality types and emotional stability subscale, **AVATAR** is a valuable tool for onboarding, personal development, and other HR purposes.

2.1.6. Constructual Mapping of AVATAR Factors

Table 4 below details the theories and frameworks to which each factor of **AVATAR** corresponds to.

Table 4. AVATAR Factors Mapped onto Other Theoretical Frameworks.

AVATAR Dimensions	Corresponding Dimension	Reference Model/Framework
Exploration/Security	Sensation/Intuition	Theory of Psychological Types (Jung, 1921)
	Conscientiousness, Openness	Five Factor Model (Costa & McCrae, 1985)
	Conscientiousness, Openness to Experience	HEXACO Model of Personality (Lee & Ashton, 2004)
	Steadiness, Compliance	DISC Model (Marston, 1928)
Spontaneous/Organised	Perception/Judgment	Myers-Briggs Type Indicator (Briggs-Myers & Myers, 1980)

Novelty-seeking/Traditional	Conscientiousness, Openness	Five Factor Model (Costa & McCrae, 1985)
	Conscientiousness, Openness to Experience Steadiness Sensing/Intuition	HEXACO Model of Personality (Lee & Ashton, 2004) DISC Model (Marston, 1928) Myers-Briggs Type Indicator (Briggs-Myers & Myers, 1980)
Experimental/Cautious	Conscientiousness, Openness	Five Factor Model (Costa & McCrae, 1985)
	Conscientiousness, Openness to Experience Influence, Steadiness Sensing/Intuition, Perception/Judgment	HEXACO Model of Personality (Lee & Ashton, 2004) DISC Model (Marston, 1928) Myers-Briggs Type Indicator (Briggs-Myers & Myers, 1980)

Control/Adaptation

Persuasive/Impressionable	Agreeableness	Five Factor Model (Costa & McCrae, 1985)
	Agreeableness, Honesty-Humility Influence	HEXACO Model of Personality (Lee & Ashton, 2004) DISC Model (Marston, 1928)
Directive/Participative	Agreeableness, Extraversion	Five Factor Model (Costa & McCrae, 1985), HEXACO Model of Personality (Lee & Ashton, 2004)
	Dominance Extraversion	DISC Model (Marston, 1928)
Driving/Easy-going	Agreeableness, Conscientiousness	Theory of Psychological Types (Jung, 1921) Five Factor Model (Costa & McCrae, 1985), HEXACO Model of Personality (Lee & Ashton, 2004)
	Dominance, Influence, Steadiness	DISC Model (Marston, 1928)

Goal-focus/People-focus

Autonomous/Affiliative	Extraversion	Theory of Psychological Types (Jung, 1921)
	Extraversion	Five Factor Model (Costa & McCrae, 1985), HEXACO Model of Personality (Lee & Ashton, 2004)
	Influence, Compliance Extraversion/Introversion	DISC Model (Marston, 1928) Myers-Briggs Type Indicator (Briggs-Myers & Myers, 1980)
Individualistic/Humanistic	Agreeableness	Five Factor Model (Costa & McCrae, 1985), HEXACO Model

		of Personality (Lee & Ashton, 2004)
	Extraversion/Introversion	Myers-Briggs Type Indicator (Briggs-Myers & Myers, 1980)
	Agreeableness, Neuroticism	Five Factor Model (Costa & McCrae, 1985), HEXACO Model of Personality (Lee & Ashton, 2004)
Vigilant/Trusting		
	Dominance	DISC Model (Marston, 1928)
<i>Subscale</i>		
Emotional Stability Measure	Emotional-Social Intelligence	ESI Model (Bar-On & Handley, 2003)

2.2. Development of AVATAR

AVATAR was developed via a thorough multistage process in a comprehensive and scientific manner. The development of the test started in 2015, as a result of accrued demand from companies using psychometric assessments in order to provide a “supplementary aid” in the recruitment process. Extensive empirical and theoretical research was conducted to create the dimensions, factors and items of **AVATAR**. The questionnaire was also pretested several times to ensure that it met the criteria for reliability and validity established by internationally recognised psychological institutions. Special attention was also given to guarding against social desirability as well as simplifying the ergonomics of the assessment.

Using robust theoretical concepts, practical and relevant work requirements, and advanced scientific analyses, all the items in the item bank were piloted and pretested with professionals from across the globe, ensuring that the assessed behaviours are relevant to current workplace needs, and are global and culture-free.

2.2.1. Dimensions, Factors and Item Construction

AVATAR assesses individual types across 3 broad dimensions, each of which consist of 3 main and opposing factors. Representing these factors are 63 questions in total, presented as paired statements (the equivalent of 126 items). Reviews ensured that the factors measured were clearly defined, and that items targeting each factor accurately reflected modern and business-relevant behaviours, thus resulting in the following:

- Greater differentiation between dimensions
- Inclusion of factors that target occupational and organisational fit
- Clearer item structure and wording
- Updated use of terms and relevance of items to match the current working world

In addition, **AVATAR** also comprises an Emotional Stability Subscale, with a total of 10 items that are presented in a 4-point Likert format.

Thus, items for **AVATAR** were written to promote ease of response as well as facilitate candidates' mental connection and understanding of statements. As the questionnaire aimed to capture work behaviours in multiple contexts, items were checked for cultural bias, clarity, offensiveness, and language accuracy.

2.2.2. Profile Calculation Method

Out of the eight total profiles, the overall profile which the candidate receives in the report is calculated using the dimensions from the three dichotomies with the highest average score that the candidate has obtained in the underlying factors. For example, a candidate will receive the 'Supporter' profile if their average scores are highest across the factors of Security, Adaptation and People-focus, than their average scores across the factors of Exploration, Control, and Goal-focus. Thus, a candidate could endorse two out of three factors from one dimension (e.g., Exploration), yet endorse the opposing dimension (e.g., Security) overall if their average score was higher across the factors of the Security dimension. Generally, however, the candidate will endorse the dimension with the most endorsed factors. This method was chosen as it takes into account all of the candidate's individual scores across all factors when calculating their dimensions and overall profile. For each profile type, the report also contains additional profile information providing a deeper understanding of the potential strengths, challenges, working style, communication style, and ideal learning environment of the candidate's personality type. This additional profile information complements the comments in the report with further profile-based insights.

2.2.3. Ipsative Response Format

Response options for the 65 items representing the core factors of **AVATAR** are written in a binary forced-choice format. In contrast to Likert scales, respondents choose between the two descriptors according to the extent of which one captures their preference, as ipsative formats have been found to reduce response bias (Chan, 2003; Saville & Willson, 1991; Ten Holt, van Duijn & Boomsma, 2010). The fact that respondents cannot endorse all the items eliminates acquiescence (the tendency to agree with all items on a questionnaire), reduces susceptibility to impression management and that it is generally more difficult to fake (Brown & Maydeu-Olivares, 2011). With an ipsative format, it is impossible for respondents to provide the same response on all items, thereby resulting in greater differentiation of scores within a profile. Ipsative scales are also beneficial because there is no need to include negatively worded items in order to balance dimensions. Rather, mirrored items facilitate measurement of behaviours for different traits along the same domains. In Likert scales, it is necessary to include negatively worded items in order to balance responses and monitor acquiescence (Brown & Maydeu-Olivares, 2013; Edwards & Kenney, 1946) but with ipsative formats, although all

items are worded positively, both positive and negative behavioural indicators are captured.

Despite such advantages, ipsative questionnaires are underutilised in work-based assessments due to the methodological challenges it poses in construct validation and interpretation, when compared to traditional normative scoring methodology (Cornwell & Dunlap, 1994; Loo, 1999; Meade, 2004).

In an ipsative test like **AVATAR**, the respondent must make a comparative judgement between two or more items to give an answer. Conversely, in a normative test, judgement is viewed in absolute terms for each evaluated personality trait and preferences are not compared between different traits. As Classical Test Theory (CTT) scoring does not reflect the psychological process of comparison, it ignores the fact that respondents indicate a preference and not an absolute judgement. As a consequence, ipsative test scores that are generated from CTT scoring approaches might be distorted. Therefore, the potential solution (configured in **AVATAR**) is to keep the ipsative format but change the way the scores are calculated, in order to reflect realistic comparative judgements.

2.2.4. Likert Response Format

AVATAR contains an Emotional Stability Subscale that follows the 4-point Likert format of responses. The agree-disagree approach of the Likert format allows to measure behaviours and attitudes that have for decades been ubiquitous in questionnaires of all kinds: market research, opinion polling, major government surveys and academic studies in fields ranging from political science to product design. Not only is it a pleasingly simple way of gauging specific opinions, but it also lends itself very easily to the construction of multiple-item measures (Johns, 2010).

The key to successful measurement via a Likert format is to convey the underlying construct to the respondents, so that they could then choose the response option that best reflects their position on that construct. This straightforward notion applied in **AVATAR** is illustrated below in Figure 6.

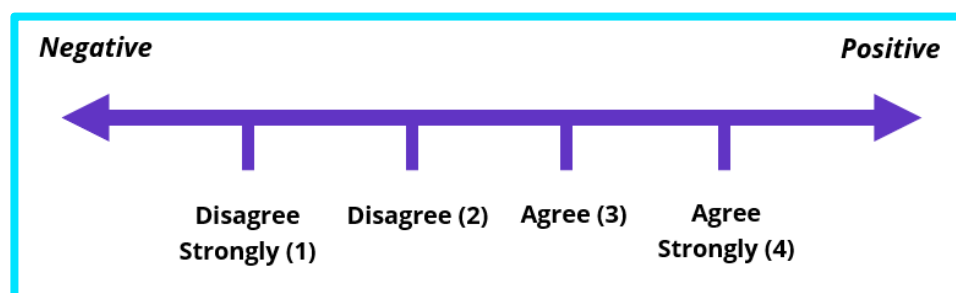


Figure 6. The Basis for Likert Measurement.

Attitudes or behaviours towards any object or on any issue vary along the same underlying negative-to-positive dimension. This has three significant implications. First, the method is universally applicable. Second, provided that the response options covered

the negative-to-positive dimension, their precise wording can vary. Third, because responses are comparable across different questions - in each case simply reporting how positively or negatively the respondent was disposed to the attitude or behaviour in question - they could be assigned the same numerical codes, as illustrated in Figure 5 above. Furthermore, with multiple items on the same broad object (such as those listed just above), these codes could be summed or averaged to give an indication of each respondent's overall positive or negative orientation towards that object. These advantages of the Likert format - above all, its simplicity and versatility - explain why this approach is ubiquitous.

3. Scoring

A raw score is the total number of responses endorsed by a candidate for each factor measured. In the context of a work-based assessment, raw scores mean little as they do not provide effective or appropriate comparisons between candidates across a specific trait. Norms enable Human Resource practitioners, managers and recruiters to evaluate an individual's score relative to the score of other delegates who complete the same assessment.

3.1. The Relationship between Raw Scores and Norms

The best norm group is one that is most representative of the characteristics of the individuals taking a test. Applied assessment assumes personality traits, abilities, interests, motivations, values, attitudes and perceptions are distributed along a normal probability or bell-shaped curve. Therefore, norm groups effectively identify individual scores across this distribution. In any given population, levels of a target attribute may be found. Therefore, for a given attribute or trait, the normal curve represents the norm or average performance of a population and the scores that are above and below the average for that population (Kaplan & Saccuzzo, 2008). For norms to be meaningful, the sample used must reflect the intended population of test takers. If a test is developed for engineers, the norm sample must consist of engineers in the population. Similarly, if an assessment was developed for nurses, it is not appropriate for use in the selection of chemists. Norm groups can range in size from a few hundred to a hundred thousand. However, the larger the norm group sample size the closer the approximation to a normal population distribution. Figure 7 below shows a normal bell curve.

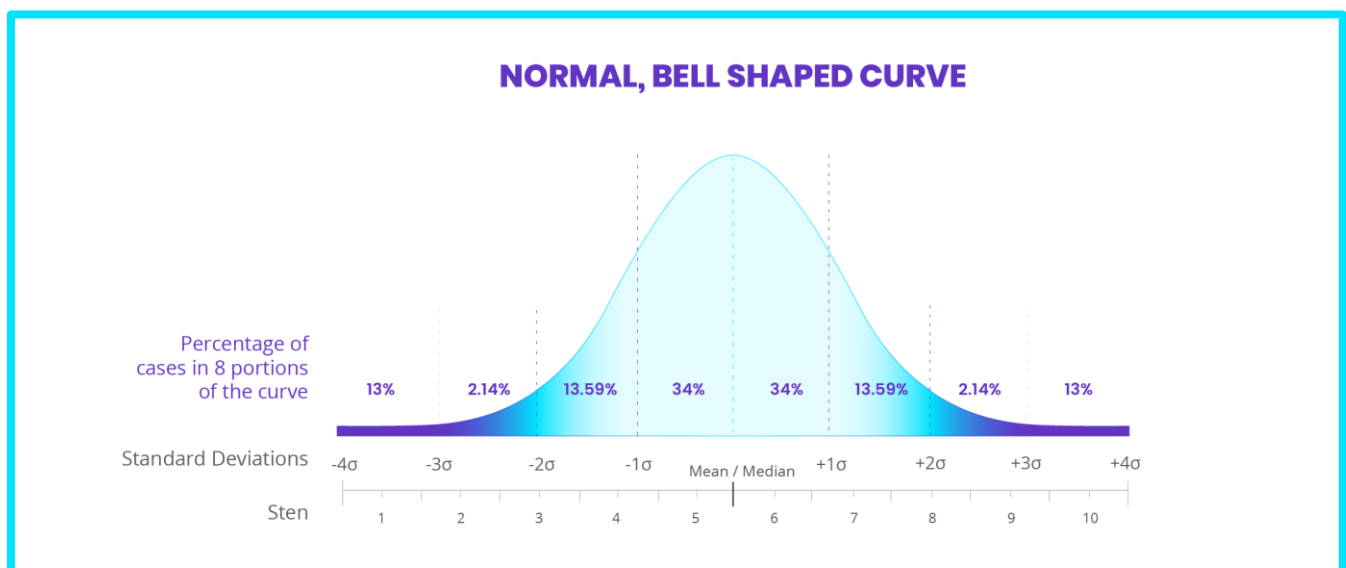


Figure 7. Normal Bell-shaped Curve showing the Distribution of Percentages, Standard Deviations and Standard Ten (Sten).

Key to the creation of norms is the conversion of raw scores into derived scores. This process provides a system of common metrics that readily facilitates the interpretation of test scores. For **AVATAR**, the forms of score standardization used are: 1) Mean; 2) Median; 3) Standard deviation; 4) Standard score; 5) Standard ten (sten); and 6) Standard error of measurement. Table 5 describes the various types of derived scores used in the validation of **AVATAR**.

Table 5. Types of Derived Scores used to Standardise AVATAR.

Sr. No.	Derived Score	Symbol	Definition and Characteristics
1	Mean	μ	The average score of all participants who complete an assessment. The mean is calculated by dividing the sum of all scores by the number of participants. The formula for calculating the Mean is: $\mu = (\sum X) / N,$ where $\sum X$ is the sum of all the population observations, and N is the number of population observations.
2	Median	$\tilde{x}$	The median is the score that separates the highest scores from the lowest. The median can be found by arranging all scores in order from lowest to highest and selecting the score in the middle. Like the mean, median is a measure of the general tendency of a group of scores.
3	Standard Deviation	σ	A measure of the amount of variation across scores for candidates who take an assessment. A small standard deviation indicates scores are close to the mean while a large value tells us there is much variability across individuals. The formula for calculating the Standard Deviation is: $\sigma = \sqrt{\sum (X_i - \mu)^2 / N},$ where μ is the population mean, X_i is the ith element from the population, and N is the number of elements in the population.
4	Standard Score	Z	An estimate of whether an individual's score is above average, average or below average compared to peers. Scored on a scale from -4 to 4 with zero in the middle representing average. Positive scores are above average while negative scores are below average. Z scores are calculated using the formula: $z = (x - \mu_x) / \sigma_x,$ where X is a normal random variable, μ is the mean of X, and σ is the standard deviation of X.

For a standard score with scores ranging from 1 – 10, here the mean is 5.5 and the standard deviation equals 2. The formula for calculating Sten scores is:		
$\text{Sten} = (X - \mu / \sigma) 2 + 5.5,$		
where X is the candidate's raw score, μ is the mean of raw scores for all candidates, σ is the standard deviation of raw scores on the standardisation sample, multiplied by 2 + 5.5.		
5	Standard Ten	Sten
Sten scores are rounded to the nearest whole number with values less than 1 or greater than 10 rounded up or down respectively.		
SEM is related to test reliability and provides an indication of the dispersion of measurement errors associated with test scores from a group of candidates. It helps us to identify level of errors when estimating candidate's true scores from their observed scores. The smaller the SEM, the more accurate are the assessments that are being made. The formula for calculating SEM is:		
6	Standard Error of Measurement	SEM
$\text{SEM} = \sigma \times \sqrt{1 - \text{reliability}}$		

3.2. Transforming Raw Scores to Standardised Scores

For all of the 73 items in **AVATAR**, the minimum and maximum overall score for each attribute is 0 and 10. The conversion of raw scores to standardised scores is achieved using a two-step process, as below:

Step 1. Ipsative scores are calculated for each scale by summing all forced-choice responses for each attribute. For each response, 1 is assigned to the response that is endorsed and 0 for the response that is not endorsed.

Step 2. Scores are repositioned on a scale from 0 to 10 by using the following formula:

$$\text{score} = \left(\frac{\text{raw score} - (\min)}{(\max) - (\min)} \right) * 10,$$

whereby max refers to the maximum possible score of each scale and min refers to the minimum possible score of each scale.

3.3. Standard Error of Measurement

No test score is perfect. Therefore, we need an estimate of the possible amount of error present within a score. The amount of error associated with a score is known as the standard error of measurement (SEM). SEM helps us to understand what is a candidate's true score and the extent to which it differs from the obtained score (Harvill, 1991). In an

ideal world, the SEM would be calculated by having an individual take an assessment on many occasions over a long period of time. The extent to which there was variation across each score obtained by the candidate would be classified as the standard error of measurement. However, as it is not feasible to assess candidates multiple times, the SEM must be calculated using a different approach.

The SEM is an estimate that is added to or subtracted from a candidate's score in order to establish a level of confidence about a score. From this confidence band, we are able to determine the extent to which a score may vary. Standard error of measurement is related to reliability. Therefore, as reliability increases, SEM decreases. For a series of scores, approximately 68% of the time, an observed score will lie within +1.0 and -1.0 SEM of the true score. On 95% of occasions, the observed score will be between +1.96 and -1.96 SEM of the true score. Whilst 99% of the time, the observed score will lie within +2.58 and -2.58 SEM of the true score. The 95% confidence band of a score is calculated using the formula:

Confidence band = score $\pm$ 1.96 x SEM

Therefore, for **AVATAR**, a factor with a score of 5 and SEM of 0.2, has a confidence band of 4.608 - 5.392. This means we can be 95% confident that an individual's true score lies in this range.

3.4. Choosing an Appropriate Norm Group

Norms give meaning to an individual's score. The interpretation of the same score may therefore differ depending on whether a score was obtained by a graduate, administrator or manager. For accurate interpretation, it is necessary to compare an individual's score with a group that is most similar to him/her. To choose the appropriate norm, first identify the country of the candidate. Then from that country, identify the age, gender, educational level, number of years' work experience, social professional status, industry or occupational group most appropriate for the candidate. Alternatively, it is possible to create your own internal norms, based on the candidates that have already been assessed.

3.5. Creating In-house Norms

It is possible to generate internal organisational norms if there is a relevant sample of 50 or more individuals. In-house norms ensure that comparisons between individuals and the reference group reflect the target role, tasks and work environment as closely as possible. In-house norms are particularly valuable in situations where standard norms are not available, roles are new, the organisation has a unique client-base or business function and in cases where the applicant pool for a role is significantly higher or lower than expected.

To create in-house norms, first calculate the mean and standard deviation for each factor on **AVATAR** and score the results based on the step-by-step guide outlined in Chapter 3.2. Organisations will then be able to compare individual applicants or employees to a local normal group.

3.6. Link between Scores and Report Comments

Comments on the **AVATAR** report reflect the scoring process. An individual receives a score of 0 to 10 (as demonstrated in Chapter 3.2.) which is converted into Low, Medium, and High scores for the personalised comments, providing an extensive interpretation of all the factors that are within the three dimensions of the test.

The candidate's scores indicate which factors within the three opposing pairs in each overarching dimension the candidate is inclined towards, the strength of their inclination, and individual attributes relating to the score. These are also presented on a separate table for each dimension. High scores are accompanied by stronger descriptors, low scores are accompanied by weaker descriptors, and so on. For example, a candidate with a low score on a factor may be described as 'not appearing to have a tendency' in the relevant factor. A high score on the other hand may describe an individual as having 'a strong tendency' in the relevant factor. Scores should be interpreted in relative terms as the order of the targeted profiles is more important than their absolute score.

4. Sample Descriptives

The sample consisted of 421 professionals from the UK, Ireland, USA, India and other countries (177 men; 244 women), ranging from 18 to 72 years. All the participants represented a variety of occupational groups, including technical, administrative, service, managerial and non-managerial roles, and a range of industries including banking and professional services, marketing and the service sector, manufacturing and product development. Data was collected from the internet crowdsourcing marketplace, *Amazon Mechanical Turk* (214 participants); an ad campaign on the social networking service company, *Facebook* (150 participants) as well as the Research and Innovation Testing Centre of *Central Test* (50 participants), with respondents receiving compensation for their participation. To ensure the quality of data obtained, participants whose answers were deemed unreliable or who completed the questionnaire in less than 5 minutes were excluded from analyses. Tables 6, 7 and 8 provide an overview of the sample descriptives.

Table 6. Sample Descriptives based on Country of Origin.

Candidate Country	Percentage	Frequency (N)
United Kingdom	5%	23
Ireland	23%	96
USA	49%	205
India	12%	49
Others	11%	48

Table 7. Sample Descriptives based on Gender.

Gender	Percentage	Frequency (N)
Female	58%	244
Male	42%	177

Table 8. Sample Descriptives based on Age Group.

Age (Years)	Percentage	Frequency (N)
18-25	14%	60
26-33	32%	133
34-41	23%	97
42-49	19%	78
50-56	9%	38
57+	4%	15

5. Norms

Norms help us to compare an individual's score on a test with the scores of others who completed the same assessment, usually under similar conditions. These individuals who took the test at the same time 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. In the case of work-based assessments, norm groups help Human Resource Managers, recruiters and coaches to:

- Determine a candidate's suitability for a role relative to the job specification.
- Evaluate a candidate's abilities across a range of desired competencies.
- Identify an employee's skill deficits and tailor training for development.
- Identify a candidate's potential team and organisational culture fit.
- Determine how a candidate is likely to respond to stress or changes in the work environment.
- Establish objective metrics of performance.

Tables 9 to 14 detail the Mean, standard deviation (SD), minimum (Min), maximum (Max), Range and standard estimate (SE) scores for the sample on the dimensions and factors of **AVATAR**, which is then organised into categories of gender and age.

Table 9. Dimension Score Norms.

Dimension	Mean	SD	Min	Max	Range	SE
Adaptation	5.02	2.66	0.00	10.00	10.00	0.12
Control	4.98	2.86	0.00	10.00	10.00	0.12
Security	5.56	2.83	0.00	10.00	10.00	0.13
Exploration	4.44	2.83	0.00	10.00	10.00	0.13
Goal-focus	5.29	2.82	0.00	10.00	10.00	0.13
People-focus	4.71	2.82	0.00	10.00	10.00	0.13
Emotional Stability Measure	7.03	1.29	3.00	10.00	7.00	0.06

Table 10. Factor Score Norms.

Factor	Mean	SD	Min	Max	Range	SE
Impressionable	4.31	3.38	0.00	10.00	10.00	0.16
Persuasive	5.69	3.38	0.00	10.00	10.00	0.16
Easy-going	5.79	2.98	0.00	10.00	10.00	0.15
Driving	4.21	2.98	0.00	10.00	10.00	0.15
Participative	5.11	3.50	0.00	10.00	10.00	0.17
Directive	4.89	3.50	0.00	10.00	10.00	0.17
Cautious	5.74	3.33	0.00	10.00	10.00	0.16
Experimental	4.26	3.33	0.00	10.00	10.00	0.16
Organised	5.38	3.25	0.00	10.00	10.00	0.16
Spontaneous	4.62	3.25	0.00	10.00	10.00	0.16
Traditional	5.68	3.62	0.00	10.00	10.00	0.18
Novelty-seeking	4.32	3.62	0.00	10.00	10.00	0.18
Autonomous	5.75	3.57	0.00	10.00	10.00	0.17
Affiliative	4.25	3.57	0.00	10.00	10.00	0.17
Individualistic	4.77	3.67	0.00	10.00	10.00	0.18
Humanistic	5.23	3.67	0.00	10.00	10.00	0.18
Vigilant	5.35	3.60	0.00	10.00	10.00	0.18
Trusting	4.65	3.60	0.00	10.00	10.00	0.18

Table 11. Dimension Score Norms by Gender.

Dimension	Mean	SD	Min	Max	Range	SE
<i>Female</i>						
Adaptation	5.19	2.86	0.00	10.00	10.00	0.18
Control	4.81	2.86	0.00	10.00	10.00	0.18
Security	6.11	2.82	0.00	10.00	10.00	0.18
Exploration	3.89	2.82	0.00	10.00	10.00	0.18
Goal-focus	5.26	2.94	0.00	10.00	10.00	0.19
People-focus	4.74	2.94	0.00	10.00	10.00	0.19
Emotional Stability Measure	6.86	1.26	3.00	10.00	7.00	0.08
Dimension	Mean	SD	Min	Max	Range	SE
<i>Male</i>						
Adaptation	4.91	2.62	0.00	10.00	10.00	0.20
Control	5.09	2.62	0.00	10.00	10.00	0.20
Security	4.90	2.93	0.00	10.00	10.00	0.22
Exploration	5.10	2.93	0.00	10.00	10.00	0.22
Goal-focus	5.33	2.97	0.00	10.00	10.00	0.22
People-focus	4.67	2.97	0.00	10.00	10.00	0.22
Emotional Stability Measure	7.26	1.29	3.30	10.00	6.70	0.10

Table 12. Factor Score Norms by Gender.

Factor	Mean	SD	Min	Max	Range	SE
<i>Female</i>						
Impressionable	4.50	3.49	0.00	10.00	10.00	0.22
Persuasive	5.50	3.49	0.00	10.00	10.00	0.22
Easy-going	5.82	3.07	0.00	10.00	10.00	0.20
Driving	4.18	3.07	0.00	10.00	10.00	0.20
Participative	5.26	3.52	0.00	10.00	10.00	0.23
Directive	4.74	3.52	0.00	10.00	10.00	0.23
Cautious	6.19	3.19	0.00	10.00	10.00	0.20
Experimental	3.81	3.19	0.00	10.00	10.00	0.20
Organised	5.80	3.25	0.00	10.00	10.00	0.21
Spontaneous	4.20	3.25	0.00	10.00	10.00	0.21
Traditional	6.32	3.41	0.00	10.00	10.00	0.22
Novelty-seeking	3.68	3.41	0.00	10.00	10.00	0.22
Autonomous	5.82	3.56	0.00	10.00	10.00	0.23
Affiliative	4.18	3.56	0.00	10.00	10.00	0.23
Individualistic	4.63	3.73	0.00	10.00	10.00	0.24
Humanistic	5.37	3.73	0.00	10.00	10.00	0.24
Vigilant	5.35	3.64	0.00	10.00	10.00	0.23
Trusting	4.65	3.64	0.00	10.00	10.00	0.23
<i>Male</i>						
Factor	Mean	SD	Min	Max	Range	SE

Impressionable	4.06	3.21	0.00	10.00	10.00	0.24
Persuasive	5.94	3.21	0.00	10.00	10.00	0.24
Easy-going	5.75	2.85	0.00	10.00	10.00	0.21
Driving	4.25	2.85	0.00	10.00	10.00	0.21
Participative	4.91	3.47	0.00	10.00	10.00	0.26
Directive	5.09	3.47	0.00	10.00	10.00	0.26
Cautious	5.12	3.42	0.00	10.00	10.00	0.26
Experimental	4.88	3.42	0.00	10.00	10.00	0.26
Organised	4.79	3.17	0.00	10.00	10.00	0.24
Spontaneous	5.21	3.17	0.00	10.00	10.00	0.24
Traditional	4.79	3.71	0.00	10.00	10.00	0.28
Novelty-seeking	5.21	3.71	0.00	10.00	10.00	0.28
Autonomous	5.65	3.6	0.00	10.00	10.00	0.27
Affiliative	4.35	3.6	0.00	10.00	10.00	0.27
Individualistic	4.97	3.59	0.00	10.00	10.00	0.27
Humanistic	5.03	3.59	0.00	10.00	10.00	0.27
Vigilant	5.36	3.56	0.00	10.00	10.00	0.27
Trusting	4.64	3.56	0.00	10.00	10.00	0.27

Table 13. Dimension Score Norms by Age Group.

Dimension	Mean	SD	Median	Min	Max	Range	SE
18-25							
Adaptation	5.63	2.48	5.71	0.00	10.00	10.00	0.32
Control	4.37	2.48	5.71	0.00	10.00	10.00	0.32

Security	5.95	2.80	6.19	0.00	10.00	10.00	0.36
Exploration	4.05	2.80	6.19	0.00	10.00	10.00	0.36
Goal-focus	5.11	2.65	4.77	0.00	10.00	10.00	0.34
People-focus	4.89	2.65	4.77	0.00	10.00	10.00	0.34
Emotional Stability Measure	6.87	1.45	7.00	3.30	10.00	6.70	0.19

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

26-33

Adaptation	4.77	2.68	4.77	0.00	10.00	10.00	0.23
Control	5.23	2.68	4.77	0.00	10.00	10.00	0.23
Security	5.29	2.68	5.70	0.00	10.00	10.00	0.23
Exploration	4.71	2.68	5.70	0.00	10.00	10.00	0.23
Goal-focus	5.21	2.83	5.23	0.00	10.00	10.00	0.25
People-focus	4.71	2.83	5.23	0.00	10.00	10.00	0.25
Emotional Stability Measure	7.09	1.13	7.00	3.30	10.00	6.70	0.10

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

34-41

Adaptation	5.43	2.83	5.70	0.00	10.00	10.00	0.29
Control	4.57	2.83	5.70	0.00	10.00	10.00	0.29
Security	5.98	2.91	6.19	0.00	10.00	10.00	0.29
Exploration	4.57	2.91	6.19	0.00	10.00	10.00	0.29
Goal-focus	6.06	2.98	6.19	0.00	10.00	10.00	0.30

People-focus	3.94	2.98	6.19	0.00	10.00	10.00	0.30
Emotional Stability Measure	6.87	1.29	7.00	3.00	10.00	7.00	0.13

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

42-49

Adaptation	4.85	2.79	4.29	0.00	10.00	10.00	0.32
Control	5.15	2.79	4.29	0.00	10.00	10.00	0.32
Security	5.76	3.16	5.71	0.00	10.00	10.00	0.36
Exploration	4.24	3.16	5.71	0.00	10.00	10.00	0.36
Goal-focus	5.01	3.02	5.00	0.00	10.00	10.00	0.34
People-focus	4.99	3.02	5.00	0.00	10.00	10.00	0.34
Emotional Stability Measure	7.05	1.35	7.00	3.00	10.00	7.00	0.15

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

50-56

Adaptation	5.26	2.95	4.53	0.00	10.00	10.00	0.48
Control	4.74	2.95	4.53	0.00	10.00	10.00	0.48
Security	5.39	3.04	5.49	0.48	9.53	9.05	0.49
Exploration	4.61	3.04	5.49	0.48	9.53	9.05	0.49
Goal-focus	4.65	3.23	4.76	0.00	10.00	10.00	0.52
People-focus	5.35	3.23	4.76	0.00	10.00	10.00	0.52
Emotional Stability Measure	6.98	1.10	6.90	4.80	10.00	5.20	0.18

Dimension	Mean	SD	Median	Min	Max	Range	SE
57+							
Adaptation	3.84	2.96	3.81	0.00	9.53	9.53	0.76
Control	6.16	2.96	3.81	0.00	9.53	9.53	0.76
Security	4.12	3.72	2.38	0.00	10.00	10.00	0.96
Exploration	5.88	3.72	2.38	0.00	10.00	10.00	0.96
Goal-focus	4.89	3.26	5.27	0.00	10.00	10.00	0.84
People-focus	5.11	3.26	5.27	0.00	10.00	10.00	0.84
Emotional Stability Measure	8.19	1.55	8.80	4.50	10.00	5.50	0.40

Table 14. Factor Score Norms by Age Group.

Factor	Mean	SD	Median	Min	Max	Range	SE
18-25							
Impressionable	4.86	3.24	5.00	0.00	10.00	10.00	0.42
Persuasive	5.14	3.24	5.00	0.00	10.00	10.00	0.42
Easy-going	6.31	2.70	7.10	0.00	10.00	10.00	0.35
Driving	3.69	2.70	7.10	0.00	10.00	10.00	0.35
Participative	5.74	3.42	7.10	0.00	10.00	10.00	0.44
Directive	4.26	3.42	7.10	0.00	10.00	10.00	0.44
Cautious	6.27	3.04	5.71	0.00	10.00	10.00	0.39
Experimental	3.73	3.04	5.71	0.00	10.00	10.00	0.39
Organised	5.98	2.99	7.10	0.00	10.00	10.00	0.39
Spontaneous	4.02	2.99	7.10	0.00	10.00	10.00	0.39
Traditional	5.60	3.59	5.70	0.00	10.00	10.00	0.46

Novelty-seeking	4.40	3.59	5.70	0.00	10.00	10.00	0.46
Autonomous	5.45	3.11	5.70	0.00	10.00	10.00	0.40
Affiliative	4.55	3.11	5.70	0.00	10.00	10.00	0.40
Individualistic	5.00	3.49	5.70	0.00	10.00	10.00	0.45
Humanistic	5.00	3.49	5.70	0.00	10.00	10.00	0.45
Vigilant	4.88	3.47	4.30	0.00	10.00	10.00	0.45
Trusting	5.12	3.47	4.30	0.00	10.00	10.00	0.45

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

26-33

Impressionable	3.93	3.03	4.29	0.00	10.00	10.00	0.26
Persuasive	6.07	3.03	4.29	0.00	10.00	10.00	0.26
Easy-going	5.23	3.07	5.70	0.00	10.00	10.00	0.27
Driving	4.77	3.07	5.70	0.00	10.00	10.00	0.27
Participative	5.15	3.37	5.70	0.00	10.00	10.00	0.29
Directive	4.85	3.37	5.70	0.00	10.00	10.00	0.29
Cautious	5.42	3.21	5.70	0.00	10.00	10.00	0.28
Experimental	4.58	3.21	5.70	0.00	10.00	10.00	0.28
Organised	5.06	3.15	5.70	0.00	10.00	10.00	0.27
Spontaneous	4.94	3.15	5.70	0.00	10.00	10.00	0.27
Traditional	5.39	3.32	5.70	0.00	10.00	10.00	0.29
Novelty-seeking	4.61	3.32	5.70	0.00	10.00	10.00	0.29
Autonomous	5.42	3.49	5.70	0.00	10.00	10.00	0.30
Affiliative	4.58	3.49	5.70	0.00	10.00	10.00	0.30

Individualistic	4.96	3.48	4.30	0.00	10.00	10.00	0.30
Humanistic	5.04	3.48	4.30	0.00	10.00	10.00	0.30
Vigilant	5.24	3.39	5.70	0.00	10.00	10.00	0.29
Trusting	4.76	3.39	5.70	0.00	10.00	10.00	0.29

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

34-41

Impressionable	4.74	3.55	4.30	0.00	10.00	10.00	0.36
Persuasive	5.26	3.55	4.30	0.00	10.00	10.00	0.36
Easy-going	6.12	2.75	5.71	0.00	10.00	10.00	0.28
Driving	3.88	2.75	5.71	0.00	10.00	10.00	0.28
Participative	5.43	3.67	5.70	0.00	10.00	10.00	0.37
Directive	4.57	3.67	5.70	0.00	10.00	10.00	0.37
Cautious	6.08	3.20	7.10	0.00	10.00	10.00	0.32
Experimental	3.92	3.20	7.10	0.00	10.00	10.00	0.32
Organised	5.76	3.30	5.71	0.00	10.00	10.00	0.34
Spontaneous	4.24	3.30	5.71	0.00	10.00	10.00	0.34
Traditional	6.11	3.64	7.10	0.00	10.00	10.00	0.37
Novelty-seeking	6.89	3.64	7.10	0.00	10.00	10.00	0.37
Autonomous	6.50	3.58	7.14	0.00	10.00	10.00	0.36
Affiliative	3.50	3.58	7.14	0.00	10.00	10.00	0.36
Individualistic	5.29	3.76	5.70	0.00	10.00	10.00	0.38
Humanistic	4.71	3.76	5.70	0.00	10.00	10.00	0.38
Vigilant	6.39	3.62	7.14	0.00	10.00	10.00	0.37

Trusting	2.61	3.62	7.14	0.00	10.00	10.00	0.37
----------	------	------	------	------	-------	-------	------

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

42-49

Impressionable	4.23	3.52	4.29	0.00	10.00	10.00	0.40
Persuasive	5.77	3.52	4.29	0.00	10.00	10.00	0.40
Easy-going	5.75	2.97	5.71	0.00	10.00	10.00	0.34
Driving	4.25	2.97	5.71	0.00	10.00	10.00	0.34
Participative	4.56	3.46	4.29	0.00	10.00	10.00	0.39
Directive	5.44	3.46	4.29	0.00	10.00	10.00	0.39
Cautious	5.88	3.48	7.10	0.00	10.00	10.00	0.39
Experimental	4.12	3.48	7.10	0.00	10.00	10.00	0.39
Organised	5.51	3.37	6.40	0.00	10.00	10.00	0.38
Spontaneous	4.49	3.37	6.40	0.00	10.00	10.00	0.38
Traditional	5.88	3.99	7.14	0.00	10.00	10.00	0.45
Novelty-seeking	4.12	3.99	7.14	0.00	10.00	10.00	0.45
Autonomous	6.03	3.70	7.14	0.00	10.00	10.00	0.42
Affiliative	3.97	3.70	7.14	0.00	10.00	10.00	0.42
Individualistic	3.96	3.74	2.86	0.00	10.00	10.00	0.42
Humanistic	6.04	3.74	2.86	0.00	10.00	10.00	0.42
Vigilant	5.04	3.75	4.29	0.00	10.00	10.00	0.42
Trusting	4.96	3.75	4.29	0.00	10.00	10.00	0.42

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

50-56

Impressionable	4.47	3.75	4.29	0.00	10.00	10.00	0.61
Persuasive	5.53	3.75	4.29	0.00	10.00	10.00	0.61
Easy-going	6.28	3.23	7.14	0.00	10.00	10.00	0.52
Driving	3.72	3.23	7.14	0.00	10.00	10.00	0.52
Participative	5.04	3.62	4.29	0.00	10.00	10.00	0.59
Directive	4.96	3.62	4.29	0.00	10.00	10.00	0.59
Cautious	5.75	3.6	5.71	0.00	10.00	10.00	0.58
Experimental	4.25	3.6	5.71	0.00	10.00	10.00	0.58
Organised	4.74	3.24	4.29	0.00	10.00	10.00	0.53
Spontaneous	5.26	3.24	4.29	0.00	10.00	10.00	0.53
Traditional	5.68	3.6	5.70	0.00	10.00	10.00	0.58
Novelty-seeking	4.32	3.6	5.70	0.00	10.00	10.00	0.58
Autonomous	4.77	3.83	5.00	0.00	10.00	10.00	0.62
Affiliative	5.23	3.83	5.00	0.00	10.00	10.00	0.62
Individualistic	4.25	3.88	3.58	0.00	10.00	10.00	0.63
Humanistic	5.75	3.88	3.58	0.00	10.00	10.00	0.63
Vigilant	4.93	3.84	5.00	0.00	10.00	10.00	0.62
Trusting	5.07	3.84	5.00	0.00	10.00	10.00	0.62

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

57+

Impressionable	2.76	3.56	1.40	0.00	8.60	8.60	0.92
Persuasive	7.24	3.56	1.40	0.00	8.60	8.60	0.92

Easy-going	5.51	3.57	7.10	0.00	10.00	10.00	0.92
Driving	4.49	3.57	7.10	0.00	10.00	10.00	0.92
Participative	3.24	3.22	2.86	0.00	10.00	10.00	0.83
Directive	6.76	3.22	2.86	0.00	10.00	10.00	0.83
Cautious	3.42	4.07	1.40	0.00	10.00	10.00	1.05
Experimental	6.58	4.07	1.40	0.00	10.00	10.00	1.05
Organised	4.19	3.84	4.29	0.00	10.00	10.00	0.99
Spontaneous	5.81	3.84	4.29	0.00	10.00	10.00	0.99
Traditional	4.76	4.28	4.29	0.00	10.00	10.00	1.10
Novelty-seeking	5.24	4.28	4.29	0.00	10.00	10.00	1.10
Autonomous	6.10	4.13	8.57	0.00	10.00	10.00	1.07
Affiliative	3.90	4.13	8.57	0.00	10.00	10.00	1.07
Individualistic	4.38	4.37	2.86	0.00	10.00	10.00	1.13
Humanistic	5.62	4.37	2.86	0.00	10.00	10.00	1.13
Vigilant	4.19	3.52	2.90	0.00	10.00	10.00	0.91
Trusting	5.81	3.52	2.90	0.00	10.00	10.00	0.91

Table 15. Dimension 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							
Adaptation	4.97	2.77	4.77	0.00	10.00	10.00	0.31
Security	5.09	2.93	5.00	0.00	10.00	10.00	0.33
Goal-focus	5.08	2.90	5.47	0.00	10.00	10.00	0.33
Emotional Stability Measure	7.11	1.21	7.00	4.00	9.80	5.80	0.14

Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 2: Bachelor's Degree, First-year Masters</i>							
Adaptation	4.90	2.84	4.77	0.00	10.00	10.00	0.23
Security	5.64	3.06	6.17	0.00	10.00	10.00	0.24
Goal-focus	5.36	2.93	4.77	0.00	10.00	10.00	0.23
Emotional Stability Measure	7.19	1.24	7.00	3.30	10.00	6.70	0.10
Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 3: 2 years of Vocational or University Studies</i>							
Adaptation	4.80	2.49	4.76	0.00	10.00	10.00	0.25
Security	5.63	2.72	5.71	0.00	10.00	10.00	0.27
Goal-focus	5.41	2.93	5.24	0.00	10.00	10.00	0.29
Emotional Stability Measure	6.93	1.19	7.00	3.00	9.80	6.80	0.12
Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 4: High School Diploma</i>							
Adaptation	5.87	2.89	5.27	0.00	10.00	10.00	0.37
Security	6.06	3.05	6.19	0.00	10.00	10.00	0.39
Goal-focus	5.38	3.12	5.24	0.00	10.00	10.00	0.40
Emotional Stability Measure	6.81	1.62	6.80	3.00	10.00	7.00	0.21

Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 5: Trades Training, No High School Diploma</i>							
Adaptation	5.20	2.78	4.77	0.48	9.05	8.57	0.84
Security	5.97	2.32	6.17	1.90	9.52	7.62	0.70
Goal-focus	4.46	3.32	3.33	0.00	9.52	9.52	1.00
Emotional Stability Measure	7.16	1.34	7.00	5.30	9.50	4.20	0.41

Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 6: Students</i>							
Adaptation	6.05	2.63	6.19	1.43	10.00	8.57	0.73
Security	5.50	2.68	5.71	0.00	9.07	9.07	0.74
Goal-focus	5.13	2.73	5.71	0.48	10.00	9.52	0.76
Emotional Stability Measure	6.37	1.06	6.50	4.50	7.50	3.00	0.29

Table 16. Factor Score Norms based on Study Level.

Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 1: Master's Degree, PhD, Engineering or Business School, MBA</i>							
Impressionable	3.95	3.35	2.90	0.00	10.00	10.00	0.38
Easy-going	5.64	2.98	5.71	0.00	10.00	10.00	0.34
Participative	5.31	3.51	5.70	0.00	10.00	10.00	0.40
Cautious	5.36	3.54	5.70	0.00	10.00	10.00	0.40
Organised	5.24	3.28	5.71	0.00	10.00	10.00	0.37
Traditional	4.65	3.52	4.30	0.00	10.00	10.00	0.40

Autonomous	5.55	3.80	5.70	0.00	10.00	10.00	0.43
Individualistic	4.60	3.63	4.29	0.00	10.00	10.00	0.41
Vigilant	5.09	3.53	5.00	0.00	10.00	10.00	0.40

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

Level 2: Bachelor's Degree, First-year Masters

Impressionable	4.11	3.35	4.29	0.00	10.00	10.00	0.27
Easy-going	5.61	3.02	5.70	0.00	10.00	10.00	0.24
Participative	4.97	3.69	5.00	0.00	10.00	10.00	0.29
Cautious	5.76	3.46	5.71	0.00	10.00	10.00	0.28
Organised	5.56	3.18	5.70	0.00	10.00	10.00	0.25
Traditional	5.59	3.76	5.71	0.00	10.00	10.00	0.30
Autonomous	5.97	3.51	7.10	0.00	10.00	10.00	0.28
Individualistic	4.94	3.74	4.30	0.00	10.00	10.00	0.30
Vigilant	5.17	3.59	4.30	0.00	10.00	10.00	0.29

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

Level 3: 2 years of Vocational or University Studies

Impressionable	4.17	3.13	4.29	0.00	10.00	10.00	0.31
Easy-going	5.51	2.96	5.71	0.00	10.00	10.00	0.30
Participative	4.71	3.16	4.30	0.00	10.00	10.00	0.32
Cautious	5.67	3.12	5.70	0.00	10.00	10.00	0.31
Organised	5.26	3.24	5.70	0.00	10.00	10.00	0.32
Traditional	5.96	3.41	5.71	0.00	10.00	10.00	0.34
Autonomous	5.64	3.62	5.70	0.00	10.00	10.00	0.36

Individualistic	4.83	3.63	4.30	0.00	10.00	10.00	0.36
Vigilant	5.77	3.51	5.71	0.00	10.00	10.00	0.35

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

Level 4: High School Diploma

Impressionable	5.36	3.69	5.71	0.00	10.00	10.00	0.47
Easy-going	6.41	2.94	7.10	0.00	10.00	10.00	0.38
Participative	5.83	3.45	5.71	0.00	10.00	10.00	0.44
Cautious	6.21	3.21	7.10	0.00	10.00	10.00	0.41
Organised	5.29	3.48	5.70	0.00	10.00	10.00	0.45
Traditional	6.68	3.61	8.57	0.00	10.00	10.00	0.46
Autonomous	5.83	3.56	5.71	0.00	10.00	10.00	0.46
Individualistic	4.73	3.80	4.29	0.00	10.00	10.00	0.49
Vigilant	5.57	3.65	7.10	0.00	10.00	10.00	0.47

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

Level 5: Trades Training, No High School Diploma

Impressionable	4.55	3.71	4.29	0.00	10.00	10.00	1.12
Easy-going	6.76	3.26	8.57	1.43	10.00	8.57	0.98
Participative	4.28	3.38	5.70	0.00	10.00	10.00	1.02
Cautious	5.84	3.16	7.14	0.00	10.00	10.00	0.95
Organised	5.58	3.41	5.71	0.00	10.00	10.00	1.03
Traditional	6.49	2.66	5.71	1.43	10.00	8.57	0.80
Autonomous	5.07	3.80	5.71	0.00	10.00	10.00	1.15
Individualistic	3.77	3.58	4.29	0.00	8.60	8.60	1.08

Vigilant	4.55	4.23	2.90	0.00	10.00	10.00	1.28
----------	------	------	------	------	-------	-------	------

Factor	Mean	SD	Median	Min	Max	Range	SE
<i>Level 6: Students</i>							
Impressionable	4.83	3.62	5.71	0.00	10.00	10.00	1.00
Easy-going	7.26	1.97	8.57	4.29	10.00	5.71	0.55
Participative	6.04	3.74	5.71	0.00	10.00	10.00	1.04
Cautious	5.94	2.85	5.71	0.00	10.00	10.00	0.79
Organised	5.06	3.38	5.71	0.00	10.00	10.00	0.94
Traditional	5.49	3.54	5.71	0.00	10.00	10.00	0.98
Autonomous	5.38	2.86	5.71	0.00	10.00	10.00	0.79
Individualistic	4.39	3.42	4.29	0.00	10.00	10.00	0.95
Vigilant	5.60	4.30	5.71	0.00	10.00	10.00	1.19

6. Reliability

Reliability measures the extent to which an assessment is free from error and the degree to which items on a scale measure the same construct. For example, all of the items that measure *Experimental* are coherent and reflect the definition of the factor. A primary concern, when developing and using assessments, 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 (reliability) with which the same results are produced. Reliability enables test users to be confident that the results accurately reflect the test taker's attitude, aptitude or behaviour in the workplace. In the context of a work-based assessment, reliability helps assure test users that the scores obtained reflect how a job candidate or employee is likely to behave in the workplace, interact with others, and work on specific tasks. If an assessment has high reliability, it also helps to guard against biases that may be caused from errors in human judgement during the recruitment and selection process.

6.1. Internal Reliability

Internal reliability measures the extent to which items on a scale are internally consistent and all measure the same construct. Internal reliability is measured by calculating Cronbach's Alpha (α) for each factor, which computes the average of all possible inter-item correlations without relying on a single indicator of reliability, which may contain large amounts of error. Cronbach's Alpha ranges from 0 to 1, with a minimum recommended value of .7 (1 being equal to perfect reliability (Nunnally, 1978). Values greater than .9 are considered excellent, between .7 to .9 are considered good, between .6 to .7 are acceptable, between .5 to .6 are poor, and values less than .5 are unacceptable. In order to ensure strong internal reliability for **AVATAR**, items for factors that obtained a Cronbach's Alpha (α) below .7 were altered during the pretest stage based on whether their item-factor correlation was below .5 or if the alpha for the factors and facets improved when deleting the item.

Table 17. Cronbach's Alpha for Factors.

Factor	Alpha
Impressionable - Persuasive	.81
Easy-going - Driving	.63
Participative - Directive	.81
Cautious - Experimental	.79

Organised - Spontaneous	.75
Novelty-seeking - Traditional	.81
Affiliative - Autonomous	.77
Humanistic - Individualistic	.76
Trusting - Vigilant	.85
Emotional Stability Measure	.84

Cronbach's Alpha was computed for the factors of **AVATAR** on the standardised sample (n = 421). Five factors show great internal consistency with alphas above .8 and four show very good consistency with alphas above .7. Only one factor, *Easy-going - Driving* obtained an alpha below .7, which is nonetheless considered acceptable (see Table 17).

Table 18. Cronbach's Alpha for Dimensions.

Dimension	Alpha
Adaptation - Control	.87
Security - Exploration	.89
People-focus - Goal-focus	.87
Emotional Stability Measure	.84

Cronbach's Alpha was computed for the dimensions of **AVATAR** on the standardised sample (n = 421). All the dimensions show great internal consistency with alphas above .80 (see Table 18).

AVATAR has an average Cronbach's Alpha of .86, indicating that the scale has very strong internal consistency and is therefore highly reliable in assessing personality types.

6.2. Test-retest Reliability

Test-retest reliability involves administering a test twice to the same respondents, under similar conditions, and comparing the results for each individual in order to determine the stability of the test scores over time. The more similar the scores on the first administration (Time 1) are to the scores on the second administration (Time 2), the greater the reliability.

The Pearson Product Correlation Coefficient gives us the correlation between the two administrations, and ranges between 0 and 1. The closer the correlation is to 1, the greater is the reliability.

The formula for the Pearson Product Moment Correlation is:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

where [x] and [y] are the two samples' means, and X and Y are the values of each of the two variables.

For **AVATAR**, test-retest reliability was calculated on an English-speaking sample (n = 101) twice, with an interval of 2 weeks. Participants ranged in age from 19 to 72 years (mean = 35.01, sd = 9.67). Test-retest reliability was computed from individual scores for all factors, and calculated using the Pearson Moment Correlation Coefficient. Table 17 shows the correlation coefficients between Time 1 and Time 2 for all the dimensions of **AVATAR**. Table 18 shows the correlation coefficient between Time 1 and Time 2 for all the factors of **AVATAR**.

Table 19. Test-retest Reliability for Dimensions.

Dimension	Correlation
Adaptation - Control	0.81
Security - Exploration	0.74
People-focus - Goal-focus	0.78
Emotional Stability Measure	0.77

Table 20. Test-retest Reliability for Factors.

Factor	Correlation
Impressionable - Persuasive	0.72

Easy-going - Driving	0.70
Participative - Directive	0.60
Cautious - Experimental	0.76
Organised - Spontaneous	0.64
Novelty-seeking - Traditional	0.76
Affiliative - Autonomous	0.72
Humanistic - Individualistic	0.71
Trusting - Vigilant	0.65....

The test-retest correlations for **AVATAR** show good results, with the lowest correlation being at .74 for the dimensions, and .60 for the factors. It can therefore be concluded that **AVATAR** is a reliable measure over time, as a person taking the test at two different points in time will consistently get the same results.

7. 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.

Validity is derived from the Latin word 'validus', meaning strong. According to the Standards for Educational and Psychological Testing (APA, AERA, & NCME, 2014), validation is a process of providing evidence for the scientific basis of interpreting assessment scores. The process of validation therefore depends on evidence supporting: a) test content, b) response processes, c) internal structure, d) relations to other variables, and e) consequences of testing. **AVATAR** was validated by following a thorough multistage process that ensured items and factors were accurate, representative, consistent, clear, fair and culturally-free.

Face validity is a measure of whether an assessment looks like it is measuring what it claims to be measuring, by those who take it (Nevo, 1985). This is an important aspect to take into account in order to obtain the candidate's co-operation and commitment. Face validity and attitudes towards tests influence candidates' motivation, concentration and test-taking anxiety (McCarthy & Goffin, 2003). As a result, the question of 'Does AVATAR look, to potential candidates, like it measures personality types?' was addressed during the creation of the assessment, by way of ensuring that aspects such as the situational questions were not misleading or confusing to the user. The following sections will present, in more detail, further aspects of validity that were taken into account for **AVATAR**.

7.1. Concurrent Validity

Concurrent validity measures whether a method is concurrent with the results of another similar tried and tested method. In this case, for concurrent validity to be established for **AVATAR**, the components of the assessment must correlate highly with corresponding components on other established questionnaires. Concurrent validity also requires that the two tests are taken at the same time or with little time between them, so that the correlation is pure and unaltered by changes over time. Tables 21 and 22 show the results detailing the concurrent validity for **AVATAR** and the NEO-PI-R Scale.

7.1.1. The NEO-PI-R Scale

The NEO-PI-R Scale (Costa & McCrae, 1992), covers 30 facets of the Five Factor Model via 120 items (Maples et al., 2014).

Tables 21 and 22 show the correlations coefficient between the dimensions and factors of AVATAR and the NEO-PI-R Scale.

The correlation coefficient is a statistical measure of the relationship strength between the relative movements of two variables and ranges between -1 and 1. A negative correlation describes the extent to which two variables move in opposite directions, while a positive correlation signified that both variables move in tandem. The relationship is the strongest when it is closest to 1.

Table 21. Correlation between Dimensions of AVATAR and the NEO-PI-R Scale.

AVATAR→ NEO↓	ES	EXPL	GOAL	CONT
O	0.21	0.00	-0.14	-0.02
C	0.52	-0.41	0.06	0.16
E	0.34	0.40	-0.50	0.50
A	0.41	-0.25	-0.13	-0.16
N	-0.75	0.11	0.17	-0.21

ES - Emotional Stability
EXPL - Exploration
GOAL - Goal-focus
CONT - Control
O - Openness to Experience
C - Conscientiousness
E - Extraversion
A - Agreeableness
N - Neuroticism

The strongest negative correlation for the dimension ES - Emotional Stability is associated with N - Neuroticism (-0.75), while the strongest negative correlation for the dimension EXPL - Exploration is associated with C - Conscientiousness (-0.41). For the dimension GOAL - Goal-focus, the strongest negative correlation is associated with E - Extraversion (-0.50) and for the dimension CONT - Control, the strongest positive correlation is associated with E - Extraversion (0.50).

Table 22. Correlation between Factors of *AVATAR* and the NEO-PI-R Scale.

AVATAR→ NEO↓	Per	Dir	Dri	Exp	Spo	Nov	Aut	Ind	Vig
O1	0.10	0.02	0.05	0.19	0.25	0.13	-0.08	0.00	0.04
O2	0.10	0.01	-0.14	-0.12	-0.13	-0.12	-0.05	-0.21	-0.04
O3	0.10	-0.05	-0.05	-0.16	-0.07	-0.17	0.00	-0.01	0.09
O4	0.10	0.11	-0.05	0.35	0.36	0.35	-0.29	-0.20	-0.41
O5	0.20	0.04	-0.11	-0.17	-0.14	-0.17	0.06	-0.19	0.03
O6	-0.09	-0.23	-0.40	-0.05	0.01	-0.07	0.01	-0.12	0.06
C1	0.19	0.29	0.14	0.07	0.03	0.01	-0.18	-0.21	-0.16
C2	0.14	0.03	-0.01	-0.29	-0.43	-0.37	0.15	0.01	0.26
C3	0.18	0.04	-0.08	-0.27	-0.37	-0.32	0.06	-0.11	0.05
C4	0.24	0.20	0.11	-0.14	-0.28	-0.21	-0.01	-0.06	0.09
C5	0.16	0.19	0.04	-0.30	-0.41	-0.38	0.09	-0.01	0.03
C6	0.15	0.01	-0.11	-0.32	-0.40	-0.35	0.22	0.01	0.17
E1	0.16	0.33	0.29	0.26	0.21	0.33	-0.45	-0.14	-0.48
E2	0.04	0.31	0.27	0.37	0.28	0.40	-0.45	-0.14	-0.54
E3	0.37	0.65	0.61	0.15	0.09	0.12	-0.27	-0.04	-0.24
E4	0.25	0.36	0.32	0.08	-0.04	0.02	-0.29	-0.17	-0.25
E5	0.09	0.28	0.39	0.56	0.48	0.58	-0.40	-0.11	-0.37
E6	0.21	0.27	0.29	0.14	0.14	0.19	-0.42	-0.21	-0.28
A1	-0.06	0.24	0.23	0.21	0.22	0.30	-0.48	-0.17	-0.68

A2	0.01	-0.20	-0.33	-0.31	-0.33	-0.36	0.21	-0.14	0.16
A3	0.18	0.07	0.05	-0.05	0.05	0.02	-0.26	-0.39	-0.25
A4	0.06	-0.16	-0.34	-0.37	-0.32	-0.34	0.28	-0.05	0.17
A5	-0.07	-0.43	-0.40	-0.29	-0.17	-0.25	0.19	-0.14	0.29
A6	0.13	0.01	0.01	-0.02	0.14	0.04	-0.20	-0.37	-0.10
N1	-0.14	-0.14	0.09	-0.02	0.08	0.02	0.10	0.23	0.15
N2	-0.10	-0.01	0.13	0.08	0.18	0.11	0.04	0.11	0.17
N3	-0.17	-0.14	-0.02	0.02	0.15	0.06	0.05	0.15	0.15
N4	-0.37	-0.33	-0.18	-0.07	0.14	0.04	0.17	0.10	0.13
N5	-0.23	-0.19	-0.05	0.16	0.27	0.21	-0.12	0.08	0.00
N6	-0.26	-0.33	-0.14	-0.10	0.09	-0.02	0.14	0.20	0.15

Per - Persuasive
Dir - Directive
Dri - Driving
Exp - Experimental
Spo - Spontaneous
Nov - Novelty-seeking
Aut - Autonomous
Ind - Individualistic
Vig - Vigilant

O1 - Imagination
O2 - Artistic Interests
O3 - Emotionality
O4 - Adventurousness
O5 - Intellect
O6 - Liberalism
C1 - Self-efficacy
C2 - Orderliness
C3 - Dutifulness
C4 - Achievement-striving
C5 - Discipline
C6 - Cautiousness
E1 - Friendliness
E2 - Gregariousness
E3 - Assertiveness

E4 - Activity Level
E5 - Excitement-seeking
E6 - Cheerfulness
A1 - Trust
A2 - Morality
A3 - Altruism
A4 - Cooperation
A5 - Modesty
A6 - Sympathy
N1 - Anxiety
N2 - Anger
N3 - Depression
N4 - Self-consciousness
N5 - Immoderation
N6 - Vulnerability

The strongest positive correlation for the factor *Per - Persuasive* lies with E3 - Assertiveness (0.37) while the strongest negative correlation is with N4 - Self-consciousness (-0.37). The strongest positive correlation for the factor *Dir - Directive* lies with E3 - Assertiveness (0.65) while the strongest negative correlation is with A5 - Modesty (-0.43). The strongest positive correlation for the factor *Dri - Driving* lies with E3 - Assertiveness (0.61) while the strongest negative correlation is with A5 - Modesty (-0.40).

The strongest positive correlation for the factor *Exp - Experimental* lies with E5 - Excitement-seeking (0.56) while the strongest negative correlation is with A4 - Cooperation (-0.37). The strongest positive correlation for the factor *Spo - Spontaneous* lies with E5 - Excitement-seeking (0.48) while the strongest negative correlation is with C5 - Discipline (-0.41). The strongest positive correlation for the factor *Nov - Novelty-seeking* lies with E5 - Excitement-seeking (0.58) while the strongest negative correlation is with C5 - Discipline (-0.38). The strongest positive correlation for the factor *Aut - Autonomous* lies with A4 - Cooperation (0.28) while the strongest negative correlation is with A1 - Trust (-0.48). The strongest positive correlation for the factor *Ind - Individualistic* lies with N1 - Anxiety (0.23) while the strongest negative correlation is with A3 - Altruism (-0.39). The strongest positive correlation for the factor *Vig - Vigilant* lies with A5 - Modesty (0.29) while the strongest negative correlation is with A1 - Trust (-0.68).

7.2. Construct Validity

Construct validity measures the degree to which a test measures what it claims to measure. It refers to the appropriateness of inferences made on the basis of observation or measurement. Construct validity is essential to the overall validity (perceived and actual) of a scale and is particularly valuable when there is much subjectivity to concepts. When a construct is not clearly or thoroughly defined, we are unable to make direct inferences between the construct and other behaviours (Cronbach & Meehl, 1955). Often there is no accepted unit of measurement for constructs; even fairly well known concepts such as intelligence are open to debate. Therefore, the construct validity of a scale may be threatened if the construct is: a) defined as too broad or too narrow, b) mislabelled as a specific dimension of or component within the construct, c) confounded by other related constructs that mask the effects of the main construct, or d) the construct was developed with insufficient research. Therefore, in order to develop a scale that measures occupational and vocational preferences, the construct must reflect known theories that define interests, identify the components that it consists of and pinpoints positive and negative attributes associated with each dimension.

For personality-based assessments, construct validity is important as it: a) adds legitimacy and acceptability to the standardised measurement of an employee's behaviour, b) facilitates understanding of organisational implications of findings, and c) enables organisations and individuals to tailor interventions to meet desired targets. Construct validity addresses the question 'Does **AVATAR** define and classify personality preferences in a way that represents theoretical and practical understanding?' For **AVATAR**, construct validity was established using two approaches: inter-factor correlations and confirmatory factor analysis, as detailed in the sections below.

7.2.1. Inter-factor Correlations

Inter-factor correlations of scale dimensions help us to determine the degree to which dimensions on a construct are related and consistent. Dimensions that are theoretically and empirically related must generate higher correlation scores. Similarly, dimensions on

a scale not proven to be linked should generate correlation values that reflect this lack of or negative relationship. Factor correlations provide an effective guide to identify the degree to which dimensions on a scale are logical, appropriate and typical of the target behaviour. Related dimensions tell us the strength of relationships between dimension and criteria, and therefore help us to make inferences about the relationship between an attribute and desired outcomes.

The correlation coefficient is a numerical way to quantify the relationship between two variables and ranges between -1 and 1. The closer the value to 1, the stronger the relationship. Subscale inter-correlations provide evidence for convergent and discriminant validity. Whilst the correlation coefficient is easy to calculate and interpret it does not imply causality, merely the association between two variables. Traditionally, inter-factor correlations are measured using Pearson (1985) correlation.

The factors in **AVATAR** measure personality preferences, however they all measure different facets of personality. In other words, if someone scores highly on a factor, it does not necessarily mean that they will score highly on all factors. Table 23 shows the inter-factor correlations.

Table 23. Inter-factor correlations.

	Factor	1	2	3	4	5	6	7	8	9
1	Impressionable - Persuasive									
2	Easy-going - Driving	.48***								
3	Participative - Directive	.61***	.57***							
4	Cautious - Experimental	.33***	.18***	.38***						
5	Organised - Spontaneous	.19***	-.02	.17***	.56***					
6	Novelty-seeking - Traditional	.20***	.14**	.26***	.75***	.52***				
7	Affiliative - Autonomous	-.08	.07	-.02	.17***	.18***	.24***			
8	Humanistic - Individualistic	-.15**	-.06	-.10*	.08	.11*	.11*	.57***		

9	Trusting - Vigilant	-.22***	-.05	-.15**	.06	.09	.13**	.49***	.43***	
10	Emotional Stability Measure	-.20***	.01	-.17***	-.18***	-.13**	-.17***	-.09	-.05	-.08

Note. *p < .05, ** p < .01, ***p < .001.

Inter-factor correlations also provide an indication of the discriminant validity of a model, i.e., the extent to which factors or concepts that are supposed to be unrelated, truly are unrelated. It is generally accepted that correlations of below .85 indicate that discriminant validity is likely, with lower scores indicating greater discriminant validity (Kline, 2011). As a result, it can be concluded that all factors of **AVATAR** are likely to have discriminant validity, with no inter-factor correlations of .85 or above.

7.2.2. 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, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are usually performed. EFA and CFA are both forms of factor analysis. Factor analysis examines the factors (hypothetical constructs) of an assessment which are developed to interpret clusters of strong inter-correlations between items in a correlation matrix. Factor analysis aims to simplify a range of items into a few common variables; these can be identified by extracting and grouping items which have high correlations on a factor to reveal the common variables.

The goal of EFA is to identify and maximise the importance of each dimension (percentage of variance), while CFA is a hypothesis driven approach that determines the degree to which observed indicators reflect the underlying construct by specifying the number of factors expected and then classifying the value of expected relationships (Schreiber, Nora, Stage, Barlow & King, 2006). In doing so, CFA provides a numerical estimate of the degree to which the proposed factor structure is an accurate fit of the phenomenon in the real world. CFA therefore investigates the degree to which measures of a construct are consistent with established understanding and theoretical arguments for that construct. A confirmatory factor analysis (CFA) with continuous factor indicators was conducted in order to check the convergence of the results with the model of the assessment. CFA is a form of factor analysis used to measure that a construct is consistent with its intended meaning. There are two key approaches to confirming the factor structure of a scale. A strict confirmatory approach is one in which a model is tested using structural equation modelling (SEM) goodness of fit tests to determine if the model specified by the research is consistent with the data. On the other hand, an alternative model approach allows two or more models to be tested to identify which has the best fit. For the development of **AVATAR**, a strict confirmatory approach was adopted as there is only one formulated model for **AVATAR**.

CFA is used to determine how well the model fits the data, and therefore evaluating the goodness of fit indices is crucial. 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 (sensitivity to sample sizes and inappropriateness when distributional assumptions are violated), measures of relative fit are strongly recommended. 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 **AVATAR**, 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 the least likely to be affected by sample size (Fan, Thompson & Wang, 1999). Furthermore, using a variety of indices is necessary as different indices reflect different aspects of model fit (Hooper, Coughlan & Mullen, 2008). Like the X^2 , the RMSEA is a measure of absolute fit that assesses the amount of 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). This is because RMSEA is related to the residuals in the model and assesses the degree to which the proposed model is inaccurate. Both the CFI and TLI are incremental or relative fit indices that assess overall improvement of a proposed model compared with an independence model 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). CFA analysis conducted for the **AVATAR** model is shown in Tables 24 and 25.

Table 24. CFA of Model Fit by Dimension.

Chi sq	df	p value	CFI	TLI	RMSEA	p value
2182.418	1793	.00	.92	.916	.026	1

The results of the CFA by dimension, displayed in Table X, show an RMSEA of .026 and CFI of .92, indicating that the dimensions of **AVATAR** have an excellent model fit and exist in the defined structure.

Table 25. CFA of Model Fit by Factor.

Chi sq	df	p value	CFI	TLI	RMSEA	p value
2735.237	1826	.00	.872	.867	.039	1

The results of the CFA by factor, displayed in Table X, show a RMSEA of .039 and CFI of .867, indicating that the factors of **AVATAR** have an excellent model fit and exist in the defined structure.

7.2.3. Consequential Validity

Consequential validity emerged from ethical concerns related to societal ramifications of assessments. It can be described as a type of validity evidence that concerns the intended and unintended, positive as well as negative, consequences related to the interpretation and use of assessments (Messick, 1995). Messick (1980, p. 1012) argued that “not only should tests be evaluated in terms of their measurement properties but that testing applications should be evaluated in terms of their potential social consequences”. Consequential validity is considered to be an important component of the evaluation of validity, rather than an isolated conceptualisation of it (Hubley & Zumbo, 2011). Given the financial costs associated with hiring a candidate, as well as the employment legislation governing diversity and inclusion of specific demographic groups, consequential validity is important for **AVATAR**. Details of how consequential validity was measured are displayed below, in Chapter 8.

8. AVATAR 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. **AVATAR** is a fair and accurate measure of personality type, and HR practitioners and psychologists can be confident of its diverse application.

Within the context of assessment, measures used must not result in adverse impact when applied to different groups. Adverse impact occurs when there is a significant difference in the rate of selection and employment decisions for one protected group over another. This is particularly important for organisations as there is a risk of selecting the wrong candidate for a job, as well as legal and financial costs associated with discrimination claims. Adverse impact is classified as discrimination when a candidate believes (and validity evidence fails to show otherwise) that not being offered a role was based on membership to certain demographics. For most countries across Europe, North America and Australia, fairness in the selection process is subject to legislation. Specifically for the USA, the Employee Testing and other Procedures Act 1971 (US Department of Labor) protects candidates against unfair assessment procedures, whilst fair employment assessment in the UK is protected by the following:

- Race Relations Act (1976)
- Sex Discrimination Act (1975;1986)
- Disability Act (1998)
- Employment Equality Regulation (2003)
 - Sexual Orientation
 - Religion and Belief
- Age Discrimination (2006)

8.1. 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). Understanding ME is important in assessments in order to establish validity, confidence and equivalence across groups. Equivalence of measurement is obtained when candidates from different cultural backgrounds, age groups, ethnicities, genders and occupational groups with the same level on a trait, have an equal chance of being selected for a role that requires the desired level on this trait. 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, 1984). One approach to assessing measurement equivalence is differential item functioning analysis.

8.1.1. 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 (see Figure 8). Item Response Theory (IRT) facilitates DIF analysis by examining the probability of endorsing items through the item characteristic curve. Here, the logistic regression approach is used, whereby separate analyses are run for each item and maximum likelihood estimation used to calculate the probability of endorsing an item. Figures 8 and 9 show an example of item characteristic curves for uniform and non-uniform DIF.

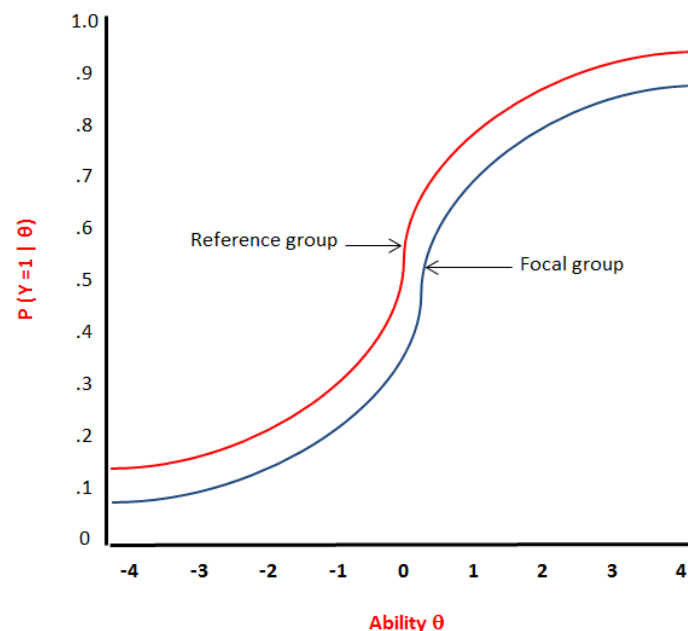


Figure 8. Item Characteristic Curve with Uniform DIF.

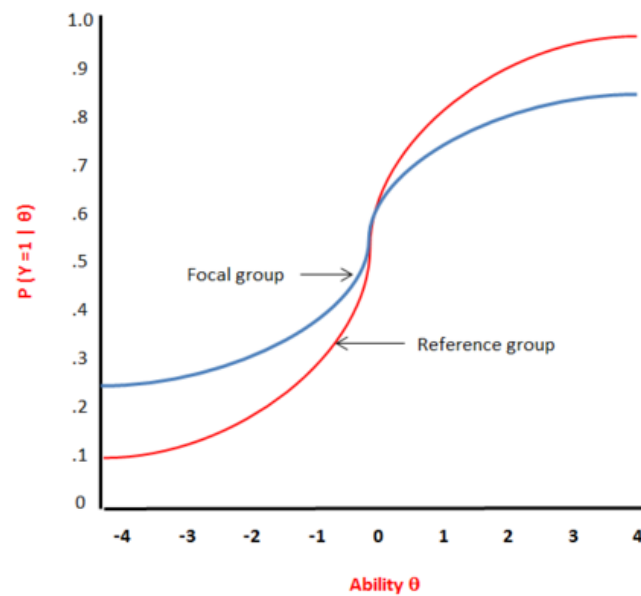


Figure 9. Item Characteristic Curve with Non-uniform DIF.

The DIF analysis provides information on whether an item is perceived in a similar way across gender, age and culture. If there is no DIF, it means that there is no bias that would favour one group over another. For example, two individuals from different countries with a similar absolute interest in the Laissez faire factor will have the same probability of endorsing the items measuring Laissez faire. However, true differences between groups may exist (e.g., some cultures are more interested in Laissez faire, others less). Such true differences are reflected in the norms (see Chapter 4).

8.1.1.a. DIF Analysis: Gender

DIF analysis was performed across gender in order to ensure that there was not a bias towards either males or females, and therefore that **AVATAR** is a fair test for both genders. For this analysis, the global norm sample was divided according to gender; 244 female and 177 male.

Table 26. Effect Sizes for DIF Items by Gender.

Item	Effect Size (Nagelkerke's R^2)	Magnitude
Spontaneous Item 10	0.012	Negligible
Organised Item 10	0.012	Negligible
Driving Item 8	0.012	Negligible

Driving Item 10	0.022	Negligible
Easy-going Item 10	0.017	Negligible
Individualistic Item 1	0.012	Negligible
Individualistic Item 9	0.010	Negligible
Humanistic Item 1	0.012	Negligible
Humanistic Item 9	0.010	Negligible
Affiliative Item 7	0.011	Negligible
Affiliative Item 10	0.012	Negligible

The analysis flagged 7 items for DIF, based on gender. Table 26 shows the DIF coefficient as well as the R^2 effect size. According to Zumbo and Thomas' (1997) classification of effect size, any R^2 under 0.13 is considered negligible. Out of the 7 items flagged for DIF, none showed an effect size of 0.13 or over. Therefore, it can be considered that **AVATAR** does not contain any item that functions differently across genders.

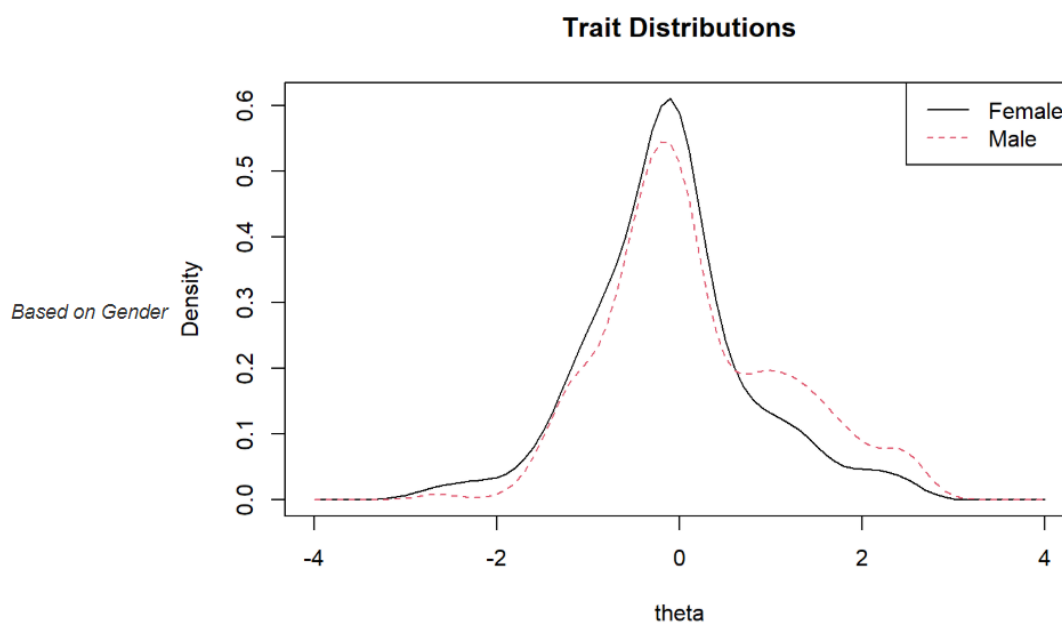


Figure 10. DIF Analysis based on Gender.

Figure 10 shows the distribution of the trait items between male (dashed line) and female (solid line). The graph shows a smooth histogram for both genders as measured by the **AVATAR** traits.

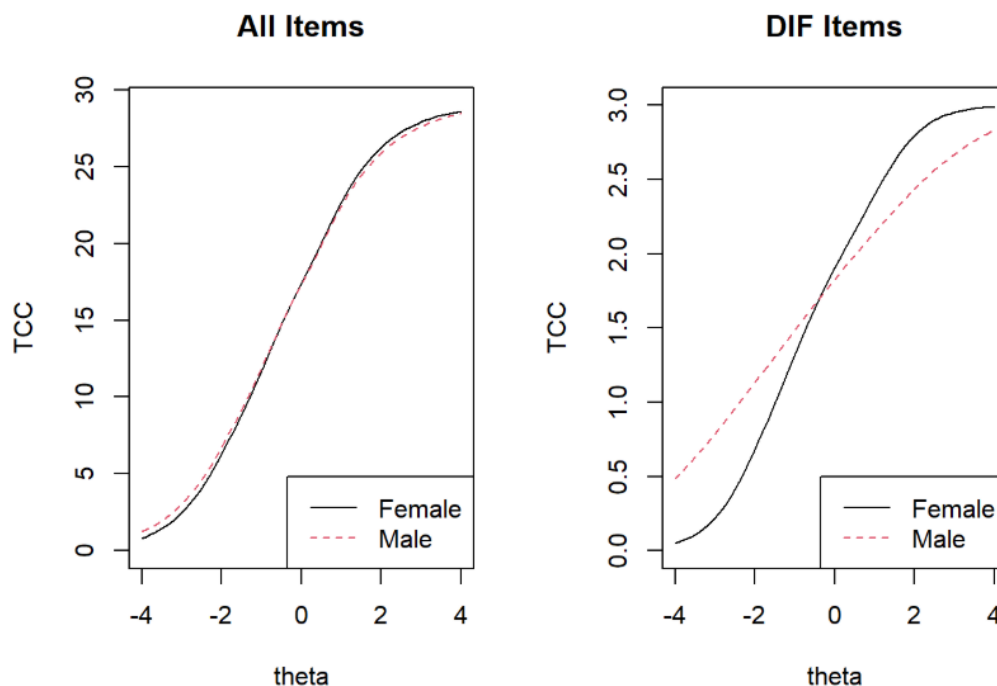


Figure 11. Item Characteristic Curves for All Items and DIF Items based on Gender.

8.1.1.a. DIF Analysis: Age

DIF analysis was performed across age groups in order to ensure that there was not a bias towards either those under 25 or those above 25, and therefore ensure that **AVATAR** is a fair test for both age groups. For this analysis, the global norm sample was divided according to age; 60 aged under 25, and 361 aged over 25.

Table 27. Effect Sizes for DIF Items by Age Groups.

Item	Effect Size (Nagelkerke's R^2)	Magnitude
Spontaneous Item 3	0.013	Negligible
Spontaneous Item 10	0.010	Negligible
Organised Item 3	0.013	Negligible
Organised Item 10	0.010	Negligible
Driving Item 1	0.011	Negligible

Driving Item 5	0.016	Negligible
Trusting Item 5	0.017	Negligible
Vigilant Item 5	0.017	Negligible
Autonomous Item 3	0.022	Negligible
Autonomous Item 9	0.011	Negligible
Affiliative Item 3	0.021	Negligible
Affiliative Item 9	0.015	Negligible

The analysis flagged 7 items for DIF, based on gender. Table 27 shows the DIF coefficient as well as the R^2 effect size. According to Zumbo and Thomas' (1997) classification of effect size, any R^2 under 0.13 is considered negligible. Out of the 7 items flagged for DIF, none showed an effect size of 0.13 or over. Therefore, it can be considered that **AVATAR** does not contain any item that functions differently across age groups.

8.1.1.c. T-tests and ANOVAs

The DIF analysis demonstrates that **AVATAR** has high consequential validity, and although several items were flagged with DIF, there was a negligible impact on the overall assessment, meaning that the assessment does not bias towards gender or age. Therefore, any significant differences found between genders or age groups can be attributed to real differences between the groups and not biases of the assessment. To further explore the differences between the groups, t-tests and ANOVAS were performed between genders and age groups (see Tables 28 to 33).

Age Groups by Dimensions

Participants were separated into six groups according to their age: 18 to 25 (n=60), 26 to 33 (n=133), 34 to 41 (n=97), 42 to 49 (n=78), 50 to 56 (n=38) and Over 57 (n=15).

Table 28. ANOVA for Dimensions by Age Groups.

Dimension	18-25	26-33	34-41	42-49	50-56	57+	F	df	p-value
Adaptation	5.63	4.77	5.43	4.85	5.26	3.84	1.90	5	0.09

Security	5.95	5.29	5.98	5.76	5.39	4.12	1.66	5	0.14
Goal-focus	5.11	5.21	6.06	5.01	4.65	4.89	1.96	5	0.08
Emotional Stability Measure	6.87	7.09	6.87	7.05	6.98	8.19	3.06	5	0.01

Note. *p < .05, ** p < .01, ***p < .001.

The ANOVA for the dimensions of **AVATAR** found no significant differences for the dimensions by age groups.

Age Groups by Factors

Table 29. ANOVA for Factors by Age Group.

Factor	18-25	26-33	34-41	42-49	50-56	57+	F	df	p-value
Impressionable	4.86	3.93	4.74	4.23	4.47	2.76	1.64	5	0.15
Easy-going	6.31	5.23	6.12	5.75	6.28	5.51	1.79	5	0.11
Participative	5.74	5.15	5.43	4.56	5.04	3.24	1.82	5	0.11
Cautious	6.27	5.42	6.08	5.88	5.75	3.42	2.26	5	0.05
Organised	5.98	5.06	5.76	5.51	4.74	4.19	1.65	5	0.14
Novelty-seeking	5.6	5.39	6.11	5.88	5.68	4.76	0.69	5	0.63
Affiliative	5.45	5.42	6.50	6.03	4.77	6.10	1.86	5	0.10
Humanistic	5.00	4.96	5.29	3.96	4.25	4.38	1.46	5	0.20
Trusting	4.88	5.24	6.39	5.04	4.93	4.19	2.42	5	0.03

Note. *p < .05, ** p < .01, ***p < .001.

The ANOVA for the dimensions of **AVATAR** found significant differences for *Cautious* and *Vigilant* factors. Post-hoc analysis was conducted using pairwise t-tests (Bonferroni Correction), in order to see where these differences occurred.

For the factor *Cautious*, the post-hoc tests found that those in the age group 57+ (M = 3.42) scored significantly lower than those across the age groups 26-33, and 50-56.

For the factor *Vigilant*, the post-hoc tests found that those in the age groups 50-56 and 57+ scored significantly lower than those in the age group 34-41 (M = 6.39).

Table 30. ANOVA for Dimensions by Study Level.

Dimension	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	F	df	p-value
Adaptation	4.97	4.90	4.80	5.87	5.20	6.05	1.70	5	0.13
Security	5.09	5.64	5.63	6.06	5.97	5.50	0.83	5	0.53
Goal-focus	5.08	5.36	5.41	5.38	4.46	5.13	0.32	5	0.90
Emotional Stability Measure	7.11	7.19	6.93	6.81	7.16	6.37	1.74	5	0.13

Note. *p < .05, ** p < .01, ***p < .001.

The ANOVA for the dimensions of **AVATAR** found significant differences for all dimensions and the emotional stability measure. Post-hoc analysis was conducted using pairwise t-tests (Bonferroni Correction), in order to see where these differences occurred.

For the dimension *Adaptation*, the post-hoc tests found that those in Level 2 (M = 4.90) scored significantly lower than those in Level 1 (M = 4.97). Those in Level 3 (M = 4.80) scored significantly lower than those in Level 1 and Level 2. Those in Level 5 (M = 5.20) scored significantly higher than those in Level 1, Level 2, Level 3, and Level 4. Those in Level 6 (M = 6.05) scored significantly higher than those in Level 4 and Level 5.

For the dimension *Security*, the post-hoc tests found that those in Level 3 (M = 5.63) scored significantly lower than those in Level 2. Those in Level 4 (M = 6.06) scored significantly higher than those in Level 2 and Level 3. Those in Level 5 (M = 5.97) scored significantly higher than those in Level 1, Level 2, and Level 3, and significantly lower than those in Level 4. Those in level 6 (M = 5.50) scored significantly higher than those in Level 1, and significantly lower than those in Level 2, Level 3, Level 4, and Level 5.

For the dimension *Goal-focus*, those in Level 2 (M = 5.36) scored significantly higher than those in Level 1. Those in Level 3 (M = 5.41) scored significantly higher than those in Level 1 and Level 2. Those in Level 4 (M = 5.38) scored significantly higher than those in Level 1 and Level 2, and significantly lower than those in Level 3. Those in Level 5 (M = 4.46) scored significantly lower than those in Level 1, Level 2, Level 3, and Level 4. Those in Level 6 (M = 5.13) scored significantly lower than those in Level 2, Level 3, and Level 4, and significantly higher than those in Level 1 and Level 5.

For the emotional stability measure, those in Level 2 (M = 7.19) scored significantly higher than those in Level 1. Those in Level 4 (M = 6.81) scored significantly lower than those in Level 3. Those in Level 5 (M = 7.16) scored significantly lower than those in Level 2, and significantly higher than those in Level 1, Level 3, and Level 4.

Table 31. ANOVA for Factors by Study Level.

Factor	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	F	df	p-value
Impressionable	3.95	4.11	4.17	5.36	4.55	4.83	1.58	5	0.16
Easy-going	5.64	5.61	5.51	6.41	6.76	7.26	1.73	5	0.13
Participative	5.31	4.97	4.71	5.83	4.28	6.04	1.19	5	0.31
Cautious	5.36	5.76	5.67	6.21	5.84	5.94	0.46	5	0.81
Organised	5.24	5.56	5.26	5.29	5.58	5.06	0.20	5	0.96
Novelty-seeking	4.65	5.59	5.96	6.68	6.49	5.49	2.49	5	0.03
Affiliative	5.55	5.97	5.64	5.83	5.07	5.38	0.30	5	0.92
Humanistic	4.60	4.94	4.83	4.73	3.77	4.39	0.29	5	0.92
Trusting	5.09	5.17	5.77	5.57	4.55	5.60	0.60	5	0.70

Note. *p < .05, ** p < .01, ***p < .001.

The ANOVA for the dimensions of **AVATAR** found significant differences for the *Traditional* factor. Post-hoc analysis was conducted using pairwise t-tests (Bonferroni Correction), in order to see where these differences occurred. The post-hoc tests found that those in Level 2 (M = 5.59) scored significantly higher than those in Level 1. Those in Level 4 (M = 6.68) scored significantly higher than those in Level 2.

Gender

In order to analyse differences in scores between genders, an independent samples t-test was conducted to compare all **AVATAR** factors between females (n= 244) and males (n=177).

Table 32. Gender T-tests for Dimension scores.

Dimensions	Female	Male	T.	df	p
Adaptation	5.19	4.91	1.03	419	0.30
Security	6.11	4.90	4.26	419	0.00
Goal-focus	5.26	5.33	-0.20	419	0.80
Emotional Stability Measure	6.86	7.26	-3.20	419	0.00

Note. *p < .05, ** p < .01, ***p < .001.

Table 33. Gender T-tests for Factor Scores.

Factor	Females	Males	T.	df	p
Impressionable	4.50	4.06	1.31	419	0.19
Easy-going	5.82	5.75	0.22	419	0.83
Participative	5.26	4.91	0.99	419	0.32
Cautious	6.19	5.12	3.31	419	0.00
Organised	5.80	4.79	3.20	419	0.00
Novelty-seeking	6.32	4.79	4.38	419	0.00
Affiliative	5.82	5.65	0.49	419	0.62
Humanistic	4.63	4.97	-0.96	419	0.34
Trusting	5.35	5.36	-0.04	419	0.97

Note. *p < .05, ** p < .01, ***p < .001.

The analyses found that there were no statistically significant differences between genders for both dimension and factor scores.

References

- Allport, G. W. (1937). *Personality: a psychological interpretation*. H. Holt and Company.
- Alonso, P., & Lewis, G. B. (2001). Public Service Motivation and Job Performance Evidence from the Federal Sector. *The American Review of Public Administration*, 31(4), 363-380.
<http://doi.org/10.1177/02750740122064992>
- American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). Standards for educational and psychological testing.
- Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. (2004). Customer Satisfaction and Shareholder Value. *Journal of Marketing*, 68(4), 172-185.
<http://doi.org/10.1509/jmkg.68.4.172.42723>
- Arnau, R. C., Green, B. A., Rosen, D. H., Gleaves, D. H., & Melancon, J. G. (2003). Are Jungian preferences really categorical?: An empirical investigation using taxometric analysis. *Personality and Individual Differences*, 34(2), 233-251.
[https://doi.org/10.1016/S0191-8869\(02\)00040-5](https://doi.org/10.1016/S0191-8869(02)00040-5)
- Barnard, C. I. (1968). *The functions of the executive*. Cambridge: Harvard University Press.
- Bar-On, R. (1997). *The Emotional Quotient Inventory (EQ-I): A test of emotional intelligence*, 14. Toronto, Canada: Multi-Health Systems, Inc.
- Bar-On, R., & Handley, R. (2003). *The Bar-On EQ-360*. Toronto, Canada: Multi-Health Systems.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, 44(1), 1-26.
<http://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Bowen, C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *The International Journal of Organizational Analysis*, 10(3), 240-259.
<http://doi.org/10.1108/eb028952>
- Boyatzis, R. E., & Skelly, F. R. (1991). The impact of changing values on organizational life. In A. D. Kolb, I. M. Rubin, and J. S. Osland (Eds.). *The Organizational Behavior Reader*. Englewood Cliffs, NJ: Prentice Hall Press.

- Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, 29, 1147-1158.
- Briggs-Myers, I. & Myers, P. (1980). *Gifts differing: Understanding personality type*. CA: Davies-Black.
- Bright, L. (2007). Does Person-Organization Fit Mediate the Relationship Between Public Service Motivation and the Job Performance of Public Employees? *Review of Public Personnel Administration*, 27(4), 361-379.
<http://doi.org/10.1177/0734371X07307149>
- Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*. New York: World Book.
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, 30(1), 99-121.
<http://doi.org/10.2333/bhmk.30.99>
- Chatman, J. A. (1989). Improving interactional organizational research: A model of person-organization fit. *The Academy of Management Review*, 14(3), 333-349.
<http://doi.org/10.2307/258171>
- Chiu, C. H. (2002). Relationship between *Employee Attitudes and Alignment of Employees-Organizational Values (Doctoral dissertation)*, Graduate Institute of Healthcare Organization Administration, National Taiwan University.
- Chung, Y. I., Chi, N. W., Chen, C. Y. (2008). Development and Valuation of Measurement on Work Values of Sales. *NTU Management Review*, 19(1): 51-82.
- 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>
- Cornwell, J. M., & Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville & Willson (1991). *Journal of Occupational and Organizational Psychology*, 67(2), 89-100.
<http://doi.org/10.1111/j.2044-8325.1994.tb00553.x>
- Costa, P. T., Jr., & McCrae, R. R. (1992). Professional manual for the revised NEO personality inventory. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

- Costa., P. T., Jr., & McCrae. R. R. (1987). Personality assessment in psychosomatic medicine: Value of a trait taxonomy. In G. A. Fava & T. N. Wise (Eds.). *Advances in psychosomatic medicine: Vol. 17. Research paradigms in psychosomatic medicine.* (pp. 71-82). Basel, Switzerland: Karger.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
<http://doi.org/10.1037/h0040957>
- Cruise, P. A. (2011). *The Role of Culture in Personnel Selection: Does culture affect hiring the right person for the job?* Lambert Academic Publishing. Germany.
- Cruise, P. A. (2012). The stability and organisational value of personality assessment. *Psychometrics Forum UK.*
- De Fruyt, F., & Mervielde, I. (1999). RIASEC types and big five traits as predictors of employment status and nature of employment. *Personnel Psychology*, 52(3), 701-727.
<http://doi.org/10.1111/j.1744-6570.1999.tb00177.x>
- Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*, 95(1), 134-135.
<http://doi.org/10.1037/0033-2909.95.1.134>
- 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>
- Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, 36, 487-495.
<http://doi.org/10.1016/j.chb.2014.04.005>
- Edwards, A. L., & Kenney, K. C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology*, 30(1), 72-83.
<http://doi.org/10.1037/h0062418>
- 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.
<http://doi.org/10.1080/10705519909540119>
- Finegan, J. E. (2000). The impact of person and organizational values on organizational commitment. *Journal of Occupational and Organizational Psychology*, 73(2), 149-169. <http://doi.org/10.1348/096317900166958>

- Galton, S. F. (1884). The Measurement of Character. *The Fortnightly Review*.
- Ghiabi, B., & Besharat, M. A. (2011). An investigation of the relationship between personality dimensions and emotional intelligence. *Procedia - Social and Behavioral Sciences, Volume 30*, 416-420.
<https://doi.org/10.1016/j.sbspro.2011.10.082>
- Goleman, D. (1998). *Working with emotional intelligence*. New York, NY: Bantam Books.
- Grehan, P. M., Flanagan, R., & Malgady, R. G. (2011). Successful graduate students: The roles of personality traits and emotional intelligence. *Psychology in the Schools, 48* (4), 317-331.
<https://doi.org/10.1002/pits.20556>
- Griffin, Em. (1991). *A First Look at Communication Theory*. McGraw-Hill, Inc.
www.afirstlook.com
- 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>
- Harvill, L. M. (1991). Standard Error of Measurement. *Educational Measurement: Issues and Practice, 10*(2), 33-41.
<http://doi.org/10.1111/j.1745-3992.1991.tb00195.x>
- Holland, J. L. (1959). A theory of vocational choice. *Journal of counseling psychology, 6*(1), 35.
- Holtzman, S., & Vezzu, S. (2011). Confirmatory Factor Analysis and Structural Equation Modeling of Noncognitive Assessments using PROC CALIS. Retrieved from <http://www.lexjansen.com/nesug/nesug11/sa/sa07.pdf>
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Articles, 2*.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.
<http://doi.org/10.1080/10705519909540118>
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research, 103*(2), 219-230.
<http://doi.org/10.1007/s11205-011-9843-4>
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *The Journal of Applied Psychology, 85*(6), 869-879.

- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research (L. Pervin and O.P. John)*. Elsevier.
- Jung, C.G. (1921) Psychological Types. The Collected Works of C.G. Jung, Vol.6, Bollingen Series XX.
- Kaplan, R. M., & Saccuzzo, D. P. (2008). *Psychological Testing: Principles, Applications, and Issues (7th ed.)*. Belmont, CA: Wadsworth Publishing Co Inc.
- Kline, R. B. (2011). Principles and practice of structural equation modeling. New York: Guilford Press.
- Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, 76(6), 889-896. <http://doi.org/10.1037/00219010.76.6.889>
- Kovach, K. A. (1995). Employee motivation: Addressing a crucial factor in your organization's performance. *Employment Relations Today*, 22(2), 93-107. <http://doi.org/10.1002/ert.3910220209>
- Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, 39(2), 329-358. http://doi.org/10.1207/s15327906mbr3902_8
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705-717. <http://doi.org/10.1037/0003066X.57.9.705>
- Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, 90(1), 125-152. <http://doi.org/10.1037/0033-2909.90.1.125>
- Loo, R. (1999). Issues in factor-analyzing ipsative measures: The Learning Style Inventory (LSI-1985) example. *Journal of Business and Psychology*, 14(1), 149-154. <http://doi.org/10.1023/A:1022918803653>
- Major, D. A., Turner, J. E., & Fletcher, T. D. (2006). Linking proactive personality and the Big Five to motivation to learn and development activity. *The Journal of Applied Psychology*, 91(4), 927-935. <http://doi.org/10.1037/0021-9010.91.4.927>

- Maples, J. L., Guan, L., Carter, N. T., & Miller, J. D. (2014). A test of the International Personality Item Pool representation of the revised NEO Personality Inventory and development of a 120-item IPIP-based measure of the Five-Factor Model. *Psychological Assessment*, 26, 1070-1084.
- Marston, W. M. (1928). *Emotion of normal people*. Harcourt, Brace.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370-396.
<http://doi.org/10.1037/h0054346>
- Massey, M. E. (1979). *The people puzzle: understanding yourself and others*. Reston Pub. Co.
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? *Emotional Development and Emotional Intelligence: Educational Implications*, 3-27. New York: Basic Books.
- McCarthy, J. M., & Goffin, R. D. (2003). Is the Test Attitude Survey Psychometrically Sound? *Educational and Psychological Measurement*, 63(3), 446-464.
<http://doi.org/10.1177/0013164403063003007>
- McCrae, R. R., & John, O. P. (1992). An introduction to the five factor model and its applications. *Journal of Personality*, 176-207.
- 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.
- Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77(4), 531-552.
<http://doi.org/10.1348/0963179042596504>
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027.
<http://doi.org/10.1037/0003-066X.35.11.1012>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
<http://doi.org/10.1037/0003-066X.50.9.741>
- Mooradian, T., Renzl, B., & Matzler, K. (2006). Who Trusts? Personality, Trust and Knowledge Sharing. *Management Learning*, 37(4), 523-540.
<http://doi.org/10.1177/1350507606073424>

- Mueller-Hanson, R., Heggstad, E. D., & Thornton III, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88(2), 348-355.
- Nadeau, L., Acier, D., Miranda, D., Pervin, L.-A., & John, O.-P. (2004). *La personnalité : De la théorie à la recherche*. Bruxelles : Saint-Laurent, Québec: De Boeck.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287-293.
<http://doi.org/10.1111/j.1745-3984.1985.tb01065.x>
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Pardee, Ronald L. (1990). *Motivation Theories of Maslow, Herzberg, McGregor & McClelland: A Literature Review of Selected Theories Dealing With Job Satisfaction and Motivation*. Washington, D.C.: Distributed by ERIC Clearinghouse.
- Pearson, K. (1895). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186, 343-414.
<http://doi.org/10.1098/rsta.1895.0010>
- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302), 157-175.
<http://doi.org/10.1080/14786440009463897>
- Porter, L. W., Bigley, G. A., & Steers, R. M. (2003). *Motivation and work behavior*. McGraw-Hill.
- Rainlall, S. (2004). A review of employee motivation theories and their implications for employee retention within organizations. *The Journal of American Academy of Business*, 9, 21-26.
- Rothmann, S., & Coetzer, E. (2003). The Big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, North America, 29.
- Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, 9(3), 185-211.
- Saville, P., & Holdsworth, R. (1999). *Occupational Personality Questionnaires*, Course Notes for OPQ. UK: SHL.

- Saville, P., & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64(3), 219-238.
<http://doi.org/10.1111/j.2044-8325.1991.tb00556.x>
- Schneider, B. (1987). The People Make the Place. *Personnel Psychology*, 40(3), 437-453.
<http://doi.org/10.1111/j.1744-6570.1987.tb00609.x>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, 99(6), 323-338.
<http://doi.org/10.3200/JOER.99.6.323-338>
- Schutz, W. C. (1958). FIRO: A three-dimensional theory of interpersonal behavior (Vol. ix). Oxford, England: Rinehart.
- Schwartz, S. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, 2(1).
<http://doi.org/10.9707/2307-0919.1116>
- Slaski, M., & Cartwright, S. (2002). Health performance and emotional intelligence. *An exploratory Study of Retail Managers Stress and Health*, 18, 63- 68.
- 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 (Vol. 758).
- Studenska, Anna. (2012). Predictors of Learning Autonomy and its Components. *Literacy Information and Computer Education Journal (LICEJ)*, Volume 3, Issue 1, 599
- Ten Holt, J. C., van Duijn, M. A. J., & Boomsma, A. (2010). Scale construction and evaluation in practice: A review of factor analysis versus item response theory applications. *Psychological Test and Assessment Modeling*, 52(3), 272-297.
- Tett, R. P., & Simonet, D. V. (2011). Faking in Personality Assessment: A "Multisaturation" Perspective on Faking as Performance. *Human Performance*, 24(4), 302-321.
<http://doi.org/10.1080/08959285.2011.597472>
- Thompson, Sheila R. (2001). The Five Most Important Entry-Level Employability Skills That Employers of the Chippewa Valley Seek in Entry-Level Job Applicants, 3-6. *American Psychological Association (APA) Publication Manual*.
- Yusooff, F., Desa, A., Ibrahim, N., Kadir, N. B. A., & Rahman, R. M. A. (2014). A study of the relationship between EQ and personality among lecturers at a research university. *Procedia - Social and Behavioral Sciences* (114), 351 – 354.

CONTACT US

Client Service: customer@centraltest.com

Training and Consulting: consulting@centraltest.com



www.centraltest.com

*Central Test – Empowering Talent © 2021
All rights reserved*