

TECHNICAL MANUAL & USER GUIDE

VOCATION

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

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

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

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

Table of Contents

Introduction	4
1. General Description and Applicability	5
1.1. Main Areas of Application	6
1.2. Questionnaire Structure	6
1.3. Mode of Administration, Languages and Special Requirements	7
1.4. Social Desirability	7
1.5. Administration Instructions	8
1.6. Technical and Security Issues	9
1.6.1. Respecting the Ethical Charter	9
1.6.2. Confidentiality and Security of User Data	9
1.6.3. Hardware Architecture and Server Availability	10
1.6.4. Computer-generated Reports	10
1.7. Supply, Conditions and Availability	10
2. Conceptual Framework and Development	11
2.2. Development of the Test	13
2.2.1. Factors and Item Construction	13
2.2.2. Pretests	14
2.2.3. Factor Definitions	14
2.2.4. Combined Profiles	17
2.2.5. Ipsative Response Format	18
2.2.6. Thurstonian Item Response Theory (IRT) and Ipsative Responses	18
2.2.6.a. Item Facility	19
2.2.6.b. Item Difficulty and Discrimination	19
2.2.7. Pilot Study	19
2.3. The relationship between Raw Scores and Norms	20
2.4. Transforming Raw Scores to Standardised Scores	22
2.5. Standard Error of Measurement	22
2.6. Choosing an Appropriate Norm Group	23
2.7. Creating In-house Norms	23
2.8. Link between Scores and Report Comments	23
3. Sample Descriptives	25
4. Norms	26
5. Reliability	35
5.1. Internal Reliability	35

5.2. Test-Retest	36
6. Validity	37
6.1. Concurrent Validity	37
6.1.1. The ORVIS IPIP	37
6.2. Construct Validity	39
6.2.1. Inter-factor Correlations	40
6.2.2. Confirmatory Factor Analysis	40
6.2.3. Consequential Validity	42
7. VOCATION and Equal Opportunities	43
7.1. Measurement Equivalence	43
7.1.1. Differential Item Functioning (DIF)	43
7.1.1.a. DIF Analysis: Gender	45
7.1.1.b. DIF Analysis: Age	46
7.1.1.c. T-tests and ANOVAs	47
8. Additional Information and Considerations	53
8.1. Interpreting Scores	53
8.2. Organisational Need: Career selection, Recruitment, and Development	54
References	55

Introduction

VOCATION is a career interest assessment tool designed to analyse the occupational and vocational preferences of individuals. Based on John Holland's (1959) model of occupational interests (Holland Occupational Themes), **VOCATION** provides the user with an informative evaluation of the occupational areas that suit their interests.

VOCATION is an updated version of *Central Test's Occupational Interest Inventory-R*. This new version includes various benefits such as:

- More contemporary questions reflecting modern day vocations and fields
- Short and easy-to-understand questionnaire
- More situational questions
- Average completion time of 8 minutes
- Ipsative scaling to guard against social desirability and impression management
- Developed using state-of-the-art psychometric techniques such as Thurstonian Item Response Theory
- Practical and easy-to-understand reports
- Contemporary 3 factor combination profiles

In this manual and user guide you will find useful information about the questionnaire development, theoretical background, psychometric properties and practical applications.

Chapters 1, 2, 3 and 8 are specifically tailored for HR practitioners, while Chapters 4, 5, 6 and 7 comprise of detailed technical and psychometric information.

1. General Description and Applicability

VOCATION maps out the test-taker's career interests across six major factors. Given below are the definitions for each factor.

Realistic

Individuals with "realistic" interests tend to like physical and/or manual activities. Driven by concrete results, they are drawn to tasks that require great dexterity and physical endurance, and as a result are likely to enjoy outdoor activities or those that involve working manually with their hands, such as fixing and assembling objects.

Investigative

This factor corresponds to individuals that enjoy investigating theoretical knowledge. They like to solve problems, conduct surveys and research information in order to better understand a subject or the environment around them.

Artistic

Those with an interest in "artistic" activities are likely to want to express themselves through creative means such as music, drawing and art. They tend to be interested in activities that allow them to be intuitive and imaginative and may not enjoy being part of projects that are too conventional, with strict methods that allow for little improvisation.

Social

Individuals with an interest in "social" activities like to be around people, communicating with and/or helping them wherever possible. They may have a preference for activities that allow them to meet new people and to have a certain sense of community or of being part of a group.

Enterprising

Individuals with an interest in "enterprising" activities tend to enjoy tasks that involve an aspect of risk-taking and influencing others. They may enjoy having to persuade others of their ideas and/or leading or managing people. They are also likely to be dynamic and take initiatives.

Conventional

Individuals with an interest in "conventional" activities have a preference for conducting work in an orderly and meticulous fashion. They are likely to want to stick to a certain structure in the tasks that they undertake and to be organised. They are likely to enjoy tasks that require precision and attention to detail and that relate to dealing with numerical data.

1.1. Main Areas of Application

VOCATION was specifically intended for HR professionals and career counsellors, and designed to be used internationally. It can be used for career counselling, career development, mentoring and coaching. Table 1 below details the different uses of **VOCATION**:

Table 1. The Different Uses of VOCATION.

Career Counselling	Career Development	Mentoring/Coaching
To help students, trainees, recent university graduates and early career professionals choose a career path	To help professionals explore new occupational avenues and expand their skill-base (horizontal career development)	To understand the vocational interests of the mentee/coachee and guide them in their career

VOCATION can be combined with other personality questionnaires such as *Central Test's Professional Profile-2* and *Sales Profile-R* in order to acquire a more precise and comprehensive profile of the individual. Table 2 below gives two examples of how **VOCATION** can be used with *Professional Profile 2* and *Sales Profile-R*. However, **VOCATION** can also be applied with other assessments, as displayed below in Table 2.

Table 2. Possible Applications of VOCATION with Other Assessments.

Additional Test	Objectives	Advantages
Professional Profile 2	Career orientation	In-depth information about interest and personality fit tailored for a specific career
Sales Profile-R	Career development	Match interest in sales-related jobs with requisite skills

1.2. Questionnaire Structure

Format

Unlike traditional questionnaire-based career interest inventories, questions in **VOCATION** are presented in a forced choice format whereby respondents select the answer that best represents their preference from two listed options. This forces respondents to select the option that is most appealing to them. Forced-choice formats reduce response bias (Chan, 2003; ten Holt, van Duijn & Boomsma, 2010), susceptibility to impression management (Baron, 1996; Bartram, 2007) and are generally more difficult to fake (Brown & Maydeu-Olivares, 2011). Forced-choice formats make it impossible to provide the same response on all items thereby resulting in greater differentiation of scores within a profile (Cruise, 2011/2012). Recruiters, career counsellors and HR practitioners can therefore be confident that responses are more accurate representations of individual work preferences.

Questions

The questionnaire consists of 60 questions that measure 6 factors and 12 facets. All questions must be answered in order for respondents to progress to the next stage of the assessment. Shown below is a sample question.

You prefer activities for which you can...

- ☐ Follow procedures carefully.
- ☐ Take risks and implement change.

1.3. Mode of Administration, Languages and Special Requirements

VOCATION 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.

Like other modern questionnaires, the assessment 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 questionnaire is currently available in English and will soon be available in other languages. For more information on language availability please contact the *Central Test Consulting Service* (consulting@centraltest.com).

VOCATION 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 course of 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 8 minutes, **VOCATION** 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.

1.4. Social Desirability

Social desirability is the tendency to present oneself in a manner that will be viewed favourably 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 **VOCATION** 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 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.

VOCATION 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 responding compared to more commonly used Likert-scales (Bowen, Martin, & Hunt, 2002; Kluger, Reilly, & Russell, 1991). Not only does the ipsative format reduce the candidate's 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.

1.5. Administration Instructions

In most cases the assessment 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 **VOCATION** 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 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 their most truthful response.

Candidates receive a notification to take the questionnaire via e-mail 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 possible to configure the invitation in the way that the assessment must be completed at the same IP address once the questionnaire has been accessed from a particular IP address.

VOCATION has no time restriction. However, candidates generally take about 8 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; there are is right or wrong answer.

1.6. Technical and Security Issues

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

1.6.1. 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.

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, s/he must contact either the organisation or *Central Test* directly, who will delete his/her data.

1.6.2. 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 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 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.

1.6.3. Hardware Architecture and Server Availability

Central Test's 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 energysources
- 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 Test's 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

1.6.4. Computer-generated Reports

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

1.7. Supply, Conditions and Availability

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

- **VOCATION** is only available for Psychologists, HR practitioners and coaches.
- Training is compulsory prior to use and distribution.
- This manual is available to individuals who have been trained to use **VOCATION**.
- **VOCATION** can be ordered online, or by contacting any *Central Test* customer service representative or local partner worldwide (customer@centraltest.com).
- For information on **VOCATION** 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 Framework and Development

2.1. Theoretical Background

John Holland's Model

The Holland Typology, also known as the *RIASEC Model*, was developed by Holland (1959) and is based on his theory that a person's vocational interest or career choice is determined by underlying "personality types". The model is popular because of its practical applicability, simple structure, and availability in order to assess career interests.

Although every individual is complex, Holland (1977) believes that everyone can be easily categorised into 6 basic personality types or a combination of the 6 types – Doers, Thinkers, Creators, Helpers, Persuaders, and Organizers. Holland (1997) hypothesised that an individual's career choice is an expression of his personality, and that the RIASEC Model can also be used to identify appropriate work environments (Gottfredson, 1999).

Holland (1959) developed his theory throughout his career, starting in the 1950's when he enrolled in the army to serve as a psychological assistant. Nauta (2010) proposed that many of Holland's ideas for his theory are a result of his graduate training and the frustrations he experienced when providing career counselling services. Nauta (2010) also states that when Holland was serving in the army, he realised that people fall into categorical vocational personality types. Holland observed army recruits being classified based on the results of an interview, and this is when the idea to sort individuals into set psychological types struck him (Nauta, 2010).

Holland, being experienced in the field of career counselling, began to notice the setbacks of the career tests available at that time. Most of these tests consisted of long scoring items and there was no set list of occupations that could be matched with the results of the assessments. This led Holland (1985) to design the Vocational Preference Inventory (VPI).

Holland continued to work on his theory for the next few years. He also published the Self-Directed Search in 1970, which has since undergone four revisions. Holland also published the Occupations Finder (1977), which is an occupational classification system. Although Holland retired in 1980, he continued to work on his theory. In 1997, he revised it again and added an environmental identity aspect to it (Staff, 2008).

Based on six personality types, Holland developed the RIASEC Model. RIASEC is an acronym for Realistic (Doers), Investigative (Thinkers), Artistic (Creators), Social (Helpers), Enterprising (Persuaders) and Conventional (Organizers). Each type is characterised by a constellation of interests, preferred activities, beliefs, abilities, values, and characteristics (Nauta, 2010).

According to Nauta (2010), Holland stressed upon the fact that most people look for a work environment that allows them to use the skills and abilities that they possess, and take on roles that they are comfortable with.

Holland's theory is widely accepted and is instrumental in bringing about a radical paradigm shift in the measurement of career-related decisions, satisfaction and stability of work (Bullock, Andrews and Braud, 2010).

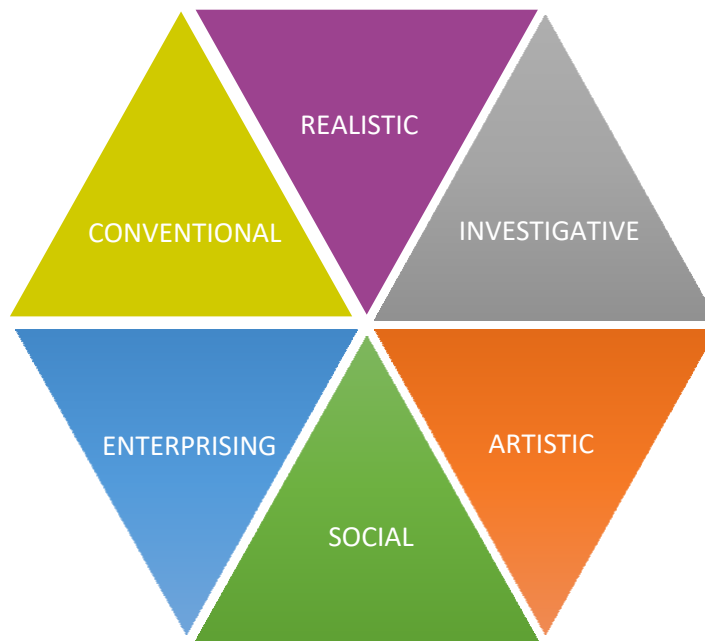


Figure 1. The Structural Framework of Holland's (1977) RIASEC Model of Occupational Themes Represented as a Hexagon.

Links between the RIASEC dimensions and personality traits

A significant amount of research has been conducted on the possible relationship between the RIASEC dimensions, and personality traits. Larson, Rottinghaus and Borgen (2002) conducted a meta- analysis on the relationship between Holland's RIASEC dimensions and the Big Five Personality factors (consisting of Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism), (Costa & McCrae, 1992). They found that there were significant correlations for men and women between the *Artistic* dimension and 'Openness' from the Big Five, *Enterprising* and 'Extraversion,' *Social* and 'Extraversion,' *Investigative* and 'Openness,' as well as *Social* and 'Agreeableness.' Barrick et al (2003) also found support for these relationships, as well as a further overlap between *Conventional* and 'Conscientiousness' and *Social* and 'Agreeableness.' However, research does not appear to show any correlation between the RIASEC Model and *Neuroticism* (Schinka, Dye, & Curtiss, 1997).

This research suggests that it is possible to infer occupational interests from the personality traits of an individual, and therefore that it may be possible to assume personality traits from one's occupational interests. However, research would suggest that occupational interests are more predictive of choice of employment than personality is (Fruyt & Mervielde, 1999), which can in turn be an accurate predictor of workplace success (Iddekinge, Putka, Roth, & Lanivich, 2011).

Sub-division of the RIASEC factors

Research has found that there are considerable differences between aspects within each dimension of Holland's RIASEC model (1958). The six factors; Realistic, Investigative, Artistic, Social, Enterprising and Conventional are broad, and considering the complexity of modern-day occupational fields, perhaps should be simplified or subdivided in order to represent a larger range of individual differences. Dingemanse, Van Amstel, De Fruyt, & Wille (2007) further explored the idea that there should be more separation *within* RIASEC types rather than just *between* them. They created the Career Insight Questionnaire (CIQ), in which they constructed 15 "building blocks of the Holland types"

(Dingemanse, Van Amstel, De Fruyt & Wille, 2015, p237) from the original six factor RIASEC model. For example, Realistic was sub-divided into 'Handcraft,' 'Outdoor,' and 'Mechanics/Construction,' and Investigative was sub-divided into 'Science' and 'Theory.' Then comparing their factors with Costa and McCrae's Big Five personality traits, Dingemanse et al (2015) found that correlations between personality traits from the Big Five were not consistent across aspects of each of the RIASEC factors, when sub-divided. For example, a negative correlation was found between Agreeableness and Holland's Investigative factor, but only for the scientific component of this factor, suggesting that an individual who scores highly on agreeableness, may have an interest in only part of Holland's 'I' interest.

Based on this research, it was decided that splitting each of the RIASEC factors into two parts for **VOCATION** would yield a more accurate representation of the range of existing occupational interests. Further details of these subdivisions can be found in Chapter 2.2.3. 'Factor Definitions'.

2.2. Development of the Test

VOCATION was developed from a thorough multi-stage process. The precursor to the questionnaire (*Occupational Interest Inventory-R*) was first developed in 2006. Questionnaire items and factors have been updated in order to ensure a) that items reflect contemporary ways of working, b) effective prevention of faking and social bias, c) that practitioners have a tool that improves the accuracy of assessment decisions and d) that individuals are clearly guided when assessing fit with desired roles and careers. As well as these aspects, throughout the revision of this assessment, it was made sure that the structure of factors and items met the highest standards of reliability and validity.

2.2.1. Factors and Item Construction

VOCATION was developed from the bank of items used from the original *Occupational Interest Inventory-R* questionnaire. The original questionnaire was reviewed in three major steps:

1. Feedback was collected from client interviews and training workshop feedback forms in order to be able to ensure that the questionnaire and the factors measured meet their needs.
2. A panel of experts consisting of psychometricians and researchers then reviewed the items for face validity purposes and to give their expert opinion.
3. The questionnaire was pretested six times with a large and representative sample of people. Results were analysed and items were removed and/or changed based on their psychometric qualities, in order to achieve the best reliability and validity.

All changes were made in order to ensure that all aspects of **VOCATION** accurately measure contemporary occupational domains and interests. Items were checked for cultural bias, clarity and accuracy. Two examples of changed items resulting from this process are depicted in Table 3 below.

Table 3. Examples of Changed Items Resulting from Review and Feedback.

Factor, Facet	Original Item	VOCATION
Artistic, A1	Assist with writing the script.	Express yourself through the writing of a script.
Enterprising, E2	Floor manager, ensuring the smooth running of the channel.	Ensuring that the teams are working efficiently.

2.2.2. Pretests

Pretests were conducted with a range of male and female professionals across the globe to ensure that items were accurate, relevant, clear, concise, culture-free, bias-free and reflected targeted work behaviours. After six English pretests, the final questionnaire was piloted with 309 English-speaking professionals from the United Kingdom. Pretest samples comprised of native English speakers from different continents in order to establish the questionnaire as an international measure of occupational interests.

Upon collection of pretest data, composite reliabilities for each dimension were calculated from generated item factor loadings. Factors were flagged if reliabilities were less than 0.55. Factors that did not converge were inspected for poorly-performing items which were then reviewed, modified or changed and then subjected to another round of pretest.

2.2.3. Factor Definitions

In light of the findings that the RIASEC factors are rather broad and should be further split up, as discussed in Chapter 2.1., **VOCATION** comprises of the original six factors, but each divided into two further facets.

On the test user's report, each factor is presented with a short indication of the personality traits, values, potential aptitudes, learning style and environment that may be endorsed for the individual, based on his/her significant interest in the factor. The *personality* aspects were developed using Holland's (1985) research, as well as further research conducted on the relationship between personality traits and the RIASEC model, as discussed in Chapter 2.1. The *values* associated with each factor signify the values which may be of importance to the individual. The *potential aptitudes* are the possible skills that the individual may hold. *Learning style* relates to the method of learning which may be most beneficial to the individual when taking in and comprehending new information. The *environment* then relates to the style of work environment that is likely to suit the individual, based on his/her interests.

The definitions for each factor, as well as their facet definitions and the additional information related to personality aspects, values, potential aptitudes, learning style and environment, are presented below.

Realistic

Individuals with "realistic" interests tend to like physical and/or manual activities. Driven by concrete results, they are drawn to tasks that require great dexterity and physical endurance, and as a result are likely to enjoy outdoor activities or those that involve working manually with the hands, such as fixing and assembling objects. This factor is divided into R1 and R2, as below:

(R1) Physical & Outdoor Activities: Refers to the desire to engage in physical activities and work outside of an office space and/or confined area.

(R2) Manual & Technical Interests: Refers to the desire to make concrete objects, and to engage in activities that involve manual work and/or the use of machinery.

Personality - Tough, genuine, natural and practical.

Values - Common sense, pragmatism and effort.

Potential aptitudes - Dexterity, comfortable with technical tasks, mechanical intelligence and physical stamina.

Learning style - A concrete and practical approach, where information is learnt through applying examples.

Environment - Most comfortable in environments which allow them to obtain concrete results, to use technical equipment or to work on-site. They tend to dislike being confined to indoor areas such as an office.

Investigative

This factor corresponds to individuals that enjoy investigating theoretical knowledge. They like to solve problems, conduct surveys and research information in order to better understand a subject or the environment around them. This factor encompasses both I1 and I2, as shown below:

(I1) Intellectual Curiosity & Learning: Refers to the preference for activities that require critical thinking and the desire to gain knowledge.

(I2) Science & Technology: Refers to an interest towards science and new technologies, as well as the desire to investigate and research such topics.

Personality - Curious about everything, and constantly seeking to learn about and understand their environment. They are often precise, intellectual and have a scientific mindset.

Values - Intellectual curiosity, critical thinking and logic.

Potential aptitudes - Research, logical reasoning, learning and analytical capabilities.

Learning style - Best suited to an approach that uses theoretical and systematic research.

Environment - Most comfortable in thought-oriented environments that favour expertise over productivity.

Artistic

Those with an interest in "artistic" activities are likely to want to express themselves through creative means such as music, drawing and art. They may feel the need to express their ideas and feelings in this way, and want the independence to do so. They tend to be interested in activities that permit them to be intuitive and imaginative, and may not have a desire to be a part of projects that are too conventional, with strict methods that allow for little improvisation. This factor is divided into A1 and A2, as below:

(A1) Aesthetic Sense & Expression: Refers to an interest in activities that require artistic sensitivity and expressing one's emotions in different forms.

(A2) Creativity & Design: Refers to a preference to conceive, imagine and innovate.

Personality - Intuitive, creative, idealistic and independent. Sensitive and attuned to art and aesthetics.

Values - An eye for beauty, originality, imagination and freedom.

Potential aptitudes - Originality, imagination and the ability to 'think outside of the box'/innovate.

Learning style - Best suited to a method that involves creativity, intuition, communicating through art, words, speech and visualising/conceptualising ideas.

Environment - Comfortable in environments that promote the expression of ideas and creativity, and where methods are not overly defined or structured.

Social

Individuals with an interest in "social" activities like to be around people, communicating with and/or helping them wherever possible. They may have a preference for meeting new people, as well as feeling a sense of co-operation or being part of a group. Furthermore, they are likely to want to be able to express themselves, as well as patiently listen to, and be attentive to, the needs of others. This factor encompasses both S1 and S2, as below:

(S1) Dedication to Others: Refers to a desire to help others, listen, inform, train, give advice and be helpful to the community.

(S2) Personal Relationships: Refers to a preference to be in contact with other people, explain, counsel and communicate with clients or co-workers.

Personality - Patient, understanding and attentive to others. They also tend to be kind, warm and welcoming.

Values - Altruism, cooperation and generosity.

Potential aptitudes - Empathy, listening and communication.

Learning style - Learn by sharing with others and by working as a team.

Environment - Comfortable in social environments that encourage personal contact, collaboration and communication.

Enterprising

Individuals with an interest in "enterprising" activities tend to enjoy tasks that involve an aspect of risk-taking and influencing others. They may enjoy selling and persuading others of their ideas and/or leading or managing others. They are also likely to enjoy communicating with others and taking initiatives. This factor is divided into E1 and E2, as below:

(E1) Enterprising: Refers to an inclination to develop new activities, materialise ideas or projects, take risks and manage difficult situations.

(E2) Leadership: Refers to the desire to encourage, motivate and manage or lead a team/other people.

Personality - Open, dynamic, assertive and entrepreneurial.

Values - Taking risks, status and competition.

Potential aptitudes - Take initiatives, lead and motivate others.

Learning style - Learn by doing and physically taking part in the type of situation that they are learning about.

Environment - Comfortable in competitive environments where they can take risks and push themselves out of their comfort zone in order to move up quickly.

Conventional

Individuals with an interest in "conventional" activities have a preference for conducting work in an orderly and meticulous fashion. They are likely to want to stick to a certain structure in the tasks that they undertake, to be organised and efficient with their time. They may as a result pay attention to small details, and be inclined towards projects involving numerical aspects, which would require such a precise approach. This factor is divided into C1 and C2, as below:

(C1) Methodical: Refers to the preference to organise, plan tasks, rigorously apply methods or rules, and control quality or structure.

(C2) Interest in Data and Numbers: Refers to the interest that an individual has to reason with numbers, analyse statistical data and informatics, as well as understand the laws of mathematics.

Personality - Meticulous, self-controlled, careful and stable.

Values - Accuracy, stability and efficiency.

Potential aptitudes - Organisation, data analysis, attention to detail and ability to work with numbers.

Learning style - Learn by following rules, instructions and established procedures.

Environment - Comfortable in structured and hierarchical environments where rules and procedures are clearly defined.

2.2.4. Combined Profiles

To date, most RIASEC tools derive individuals' occupational interest profiles by identifying their two or three main interests out of the 6 RIASEC factors, and combining them to explain which field or occupation is likely to interest the person. Holland himself noted the importance of using factor combinations, stating in his Self Directed Search tool (1979), which uses a three factor combination, that it is vital for the user to consider all the possible variations of the three factor combination. This suggests that an individual's occupational interests profile should not be inferred from their level of interest on one of the RIASEC factors alone.

In support of using the combination of several RIASEC factors to derive an individual's occupational interests profile, research has found that there is a large amount of overlap between the six RIASEC factors and personality. For example, Barrick et al (2003) and Larson, Rottinghaus and Borgen (2002) found a significant relationship between extraversion and both the Enterprising and Social factors of Holland's model. This suggests that, due to the complexity of personality, individuals are likely to be interested in multiple RIASEC factors.

In light of this, **VOCATION** reports derive an individual's occupational interests profile from the combination of their two highest interests (calculated at an equal level of significance) and their third highest interest (calculated at a slightly lower level of significance than the first two), and in turn present a list of possible appropriate trades. Using a combination of the three main factors allows for a more accurate and precise measurement of the occupations that individuals are likely to be interested in. Furthermore, keeping the third highest interest at a slightly lower level of significance allows for a more realistic view of the individual's occupational profiles, and leaves room for individuals to focus on their two main interests alone, in case they do not identify well with their third highest interest

2.2.5. Ipsative Response Format

Response options for the items in **VOCATION** are written in binary forced-choice format (FC). FC format items ask that respondents make comparative judgements by endorsing statements as 'most like me' or 'least like me', and have been shown to overcome many of the limitations associated with Likert-scale items (Brown & Maydeu-Olivares, 2011a; Chan, 2003; Saville & Willson, 1991; ten Holt, van Duijn, & Boomsma, 2010). The advantages of using FC items include:

- Removal of response biases, such as the tendency to agree with all items (acquiescence bias)
- Reduced impression management and faking
- More powerful intra-person comparisons

Despite the advantages of using forced choice items in questionnaires, they have been underutilised due to the nature of the data such questionnaires produce. These data are 'ipsative', meaning that the sum of individuals' scores remains constant. Historically, Classical Test Theory (CTT) has been considered inadequate for the scoring and validation of ipsative data (Cornwell & Dunlap, 1994; Loo, 1999; Meade, 2004), because:

- The relative nature of scores means that comparisons between individuals can be difficult.
- The linear dependency of scores distorts the construct validity of the scales.
- The lack of variance distorts correlations with external criteria.
- Assumptions required for traditional reliability estimates are violated.

2.2.6. Thurstonian Item Response Theory (IRT) and Ipsative Responses

Contemporary analysis techniques, developed by Brown and Maydeu-Olivares (2011, 2013), have allowed for the reliable use of FC items in questionnaire development. Using a combination of Item Response Theory (IRT) methodologies and Thurstonian comparison models (Thurstone, 1927; 1931), Brown & Maydeu-Olivares showed that the previous issues in the interpretation of ipsative data were able to be overcome, for example:

- 'Absolute' values could be obtained, allowing for meaningful comparisons between individuals.
- Distortions that lead to issues in validity can be corrected.

Despite their potential benefits, Thurstonian models were not previously used in the analysis of ipsative data for the following reasons (Brown, 2010):

- There was no efficient estimation method available.
- Computer technology was not powerful enough to facilitate analysis.
- Previous statistical programmes did not facilitate the calculation of scores because the variance of error for each question is 0.

For **VOCATION**, Thurstonian IRT was used to revise, rewrite and validate both items and factors. Figure 2 displays the model of Thurstonian IRT used to analyse items and factors on the questionnaire.

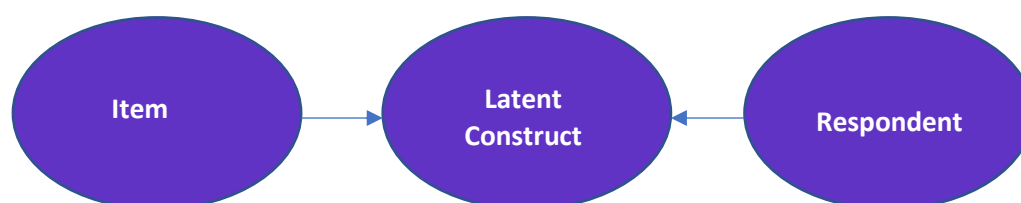


Figure 2. *Model of Thurstonian IRT Analyses Used in the Development of VOCATION.*

2.2.6.a. Item Facility

Item facility refers to an item's ability to differentiate between respondents and identify floor or ceiling effects. An item has floor or ceiling effects when the probability of endorsing a response option is much lower or greater than the probability of endorsing the alternate option. As items for **VOCATION** were binary, item facility was examined from the proportion of responses for each option. The more equivalent the proportion the more balanced the item. For **VOCATION** items with floor or ceiling effects were discarded or modified. The facility of changed or new items was then examined with subsequent pretests.

2.2.6.b. Item Difficulty and Discrimination

Here, an 'item' refers to the possible answer(s) for each question on **VOCATION**. Item difficulty is the first parameter of an IRT model. It is defined as the location on the ability continuum, where the probability of endorsing the item is equal to 50% of the participants. In an interest-based assessment, the difficulty parameter reflects the level of interest needed for the participant to have a 50% chance of choosing the item. An item is said to be 'easy' when the probability that 50% of individuals who endorsed it only have a low level on the interest measured by the item. (e.g. an "easy" item measuring Enterprising might be chosen by 50% of participants having only a low level of trait for Enterprising). On the contrary, an item is said to be "difficult", when the probability that 50% of individuals who endorsed it have a high level of trait (e.g. a "difficult" item measuring Realistic might reach the probability of being chosen by 50% of the participants only if they have a high level of trait for Realistic).

2.2.7. Pilot Study

This section is only about the final pilot study with the data used (sample) for the stats part of manual). The final questionnaire was piloted with 309 English-speaking professionals from the United Kingdom. Composite reliabilities for each factor were calculated from generated item factor loadings. Factors were flagged if reliabilities were less than 0.55. Factors that did not converge were inspected for

poorly-performing items which were then reviewed, modified or changed and then subjected to another round of pretest. The average scale composite reliability for the Pilot study was 0.83.

Scoring

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

2.3. 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, 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 3 shows the normal bell curve.

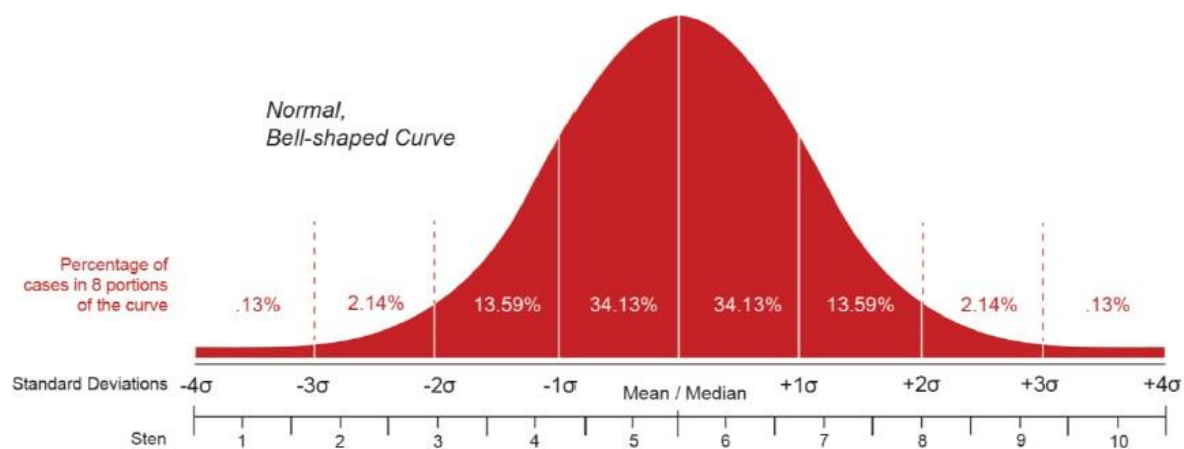


Figure 3. Normal Bell-shaped Curve Showing the Distribution of Percentages, Standard Deviations and Standard Ten (Sten).

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 facilitate the interpretation of test scores. For **VOCATION**, the forms of score standardization used were: 1) Mean; 2) Median; 3) Standard deviation; 4) Standard score; 5) Standard ten (sten); and 6) Standard error of measurement. Table 4 describes the various types of derived scores used in the validation of **VOCATION**.

Table 4. Types of Derived Scores Used to Standardise VOCATION.

Derived Score	Symbol	Definition and Characteristics
1	Mean	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: $\mu = (\sum X_i) / N,$ where $\sum X$ is the sum of all the population observations, N is the number of population observations, $\sum x$ is the sum of all the sample observations, and n is the number of sample observations.
2	Median	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	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	Estimates 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.
5	Standard Ten	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.

Sten scores are rounded to the nearest whole number with values less than 1 or greater than 10 rounded up or down respectively.

6	Standard Error of Measurement	SEM	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: $SEM = \sqrt{1 - r^2} \times \sqrt{\frac{SS_{\text{error}}}{n}}$
---	--------------------------------------	-----	--

2.4. Transforming Raw Scores to Standardised Scores

For the 60 items in **VOCATION**, 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 four-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:

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

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

Step 3 To enable the comparisons between individuals, scores for each scale are standardised using the mean and standard deviation across respondents. This process produces a z score using the formula:

$$z = (score - (\mu / \sqrt{n}))$$

Step 4 Standardised scores are then transformed into sten scores from 1 to 10 using the formula:

$$sten = z * 2 + 5.5$$

2.5. 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:

$$\text{Confidence band} = \text{score} \pm 1.96 \times \text{SEM}$$

Therefore, for **VOCATION**, 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.

2.6. 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 senior 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.

2.7. 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 **VOCATION** 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.

2.8. Link between Scores and Report Comments

Comments on **VOCATION** reports reflect the scoring process. An individual receives a score of 0 to 10 (as demonstrated in Chapter 3.2.). The comments provide an extensive interpretation of all facets within the six RIASEC factors of the test. They are chosen from six levels according to the score of the candidate, and detail individual preferences and orientation towards the attributes according to the

individual's score. Higher scores are accompanied by stronger descriptors, weaker 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 an interest' in the relevant factor. A high score on the other hand may describe an individual as having 'a strong interest' in the relevant factor. Scores should be interpreted in relative terms; the order of the targeted profiles is more important than their absolute score.

Combined profiles on **VOCATION** reports also reflect the scoring process. These profiles represent a combination of the three highest factor scores that the individual has endorsed. They are calculated on the basis that the scores on the first two interests will be of highest, equal significance, whereas the score on third interest will be of slightly lower significance.

3. Sample Descriptives

The sample comprised of 309 professionals from the UK, (142 men; 167 women) ranging in age from 18 to 72 years ($\mu=43.19$; $\sigma=13.92$). All the participants represented a range of occupational groups, representing 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 professional research and sampling firm named CINT, and each respondent received 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 8 minutes were excluded from analyses. Tables 5, 6 and 7 provide an overview of the sample descriptives.

Table 5. Country of Origin.

Candidate Country	Percentage (%)	N
United Kingdom	100	309

Table 6. Gender.

Gender	Percentage (%)	N
Females	54.05	167
Males	45.95	142

Table 7. Age Groups.

Age Group	Percentage (%)	N
18 to 25	14.24	44
26 to 33	15.53	48
34 to 41	14.89	46
42 to 49	12.30	38
50 to 56	22.33	69
Over 57	20.71	64

4. 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, or level of an attribute, interest or trait relative to his/her peers.

Tables 8 to 13 detail the mean, standard deviation, minimum, and maximum scores for the UK sample on factors of **VOCATION**, for the whole sample on gender, and is organised into categories based on age and gender.

Table 8. Factor Score Norms for the Whole Sample.

Factor	Mean	SD	Min	Max	Range	SE
R	4.62	2.13	0.00	10.00	10.00	0.12
I	5.65	1.74	1.00	9.50	8.50	0.10
A	5.22	2.31	0.00	10.00	10.00	0.13
S	4.82	2.18	0.00	10.00	10.00	0.12
E	4.42	1.59	0.00	8.50	8.50	0.09
C	5.27	2.33	0.00	10.00	10.00	0.13

Table 9. Facet Score Norms for the Whole Sample.

Facet	Mean	SD	Min	Max	Range	SE
R1	4.85	2.63	0.00	10.00	10.00	0.15
R2	4.39	2.43	0.00	10.00	10.00	0.14
I1	5.69	1.90	1.00	10.00	9.00	0.11
I2	5.61	2.15	0.00	10.00	10.00	0.12
A1	5.02	2.53	0.00	10.00	10.00	0.14
A2	5.43	2.48	0.00	10.00	10.00	0.14
S1	5.51	2.43	0.00	10.00	10.00	0.14
S2	4.13	2.50	0.00	10.00	10.00	0.14

E1	4.37	1.68	0.00	8.00	8.00	0.10
E2	4.47	2.25	0.00	10.00	10.00	0.13
C1	5.71	2.31	0.00	10.00	10.00	0.13
C2	4.83	2.97	0.00	10.00	10.00	0.17

Table 10. Factor Score Norms by Gender.

Factor	Mean	SD	Min	Max	Range	SE
Female						
R	4.04	1.83	0.00	8.50	8.50	0.14
I	5.40	1.71	1.00	9.00	8.00	0.13
A	5.59	2.26	0.50	10.00	9.50	0.17
S	5.18	2.24	0.00	10.00	10.00	0.17
E	4.31	1.59	1.00	8.50	7.50	0.12
C	5.48	2.36	0.50	10.00	9.50	0.18
Factor	Mean	SD	Min	Max	Range	SE
Male						
R	5.30	2.25	0.50	10.00	9.50	0.19
I	5.94	1.73	1.50	9.50	8.00	0.15
A	4.79	2.30	0.00	9.50	9.50	0.19
S	4.39	2.02	0.00	9.50	9.50	0.17
E	4.54	1.59	0.00	8.50	8.50	0.13
C	5.02	2.29	0.00	10.00	10.00	0.19

Table 11. Facet Score Norms by Gender.

Facet	Mean	SD	Min	Max	Range	SE
Female						
R1	4.45	2.45	0.00	10.00	10.00	0.19
R2	3.62	2.18	0.00	10.00	10.00	0.17

I1	5.58	1.92	1.00	10.00	9.00	0.15
I2	5.22	2.07	1.00	10.00	9.00	0.16
A1	5.43	2.45	0.00	10.00	10.00	0.19
A2	5.76	2.45	0.00	10.00	10.00	0.19
S1	6.09	2.39	0.00	10.00	10.00	0.18
S2	4.28	2.69	0.00	10.00	10.00	0.21
E1	4.29	1.68	0.00	8.00	8.00	0.13
E2	4.34	2.22	0.00	10.00	10.00	0.17
C1	6.09	2.21	1.00	10.00	9.00	0.17
C2	4.86	3.05	0.00	10.00	10.00	0.24

Facet	Mean	SD	Min	Max	Range	SE
<i>Male</i>						
R1	5.32	2.76	0.00	10.00	10.00	0.23
R2	5.29	2.39	0.00	10.00	10.00	0.20
I1	5.82	1.86	2.00	10.00	8.00	0.16
I2	6.07	2.17	0.00	10.00	10.00	0.18
A1	4.55	2.54	0.00	10.00	10.00	0.21
A2	5.04	2.47	0.00	10.00	10.00	0.21
S1	4.84	2.31	0.00	9.00	9.00	0.19
S2	3.95	2.25	0.00	10.00	10.00	0.19
E1	4.46	1.67	0.00	8.00	8.00	0.14
E2	4.62	2.28	0.00	10.00	10.00	0.19
C1	5.26	2.34	0.00	10.00	10.00	0.20
C2	4.79	2.88	0.00	10.00	10.00	0.24

Table 12. Factor Score Norms by Age Groups.

Factor	Mean	SD	Min	Max	Range	SE
18 to 25						
R	4.27	2.05	0.00	8.50	8.50	0.31
I	5.58	1.77	2.50	9.00	6.50	0.27
A	6.22	2.41	1.00	10.00	9.00	0.36
S	4.14	1.83	1.50	7.50	6.00	0.28
E	4.53	1.45	1.50	7.50	6.00	0.22
C	5.26	2.02	1.00	9.50	8.50	0.30
Factor	Mean	SD	Min	Max	Range	SE
26 to 33						
R	4.93	1.84	1.00	9.00	8.00	0.27
I	5.36	1.76	2.00	9.50	7.50	0.25
A	5.24	2.37	0.00	10.00	10.00	0.34
S	4.93	1.86	2.00	10.00	8.00	0.27
E	4.65	1.61	1.50	8.50	7.00	0.23
C	4.90	2.33	0.50	10.00	9.50	0.34
Factor	Mean	SD	Min	Max	Range	SE
34 to 41						
R	4.75	2.09	1.00	10.00	9.00	0.31
I	5.95	2.02	1.00	9.50	8.50	0.30
A	5.38	2.10	0.50	10.00	9.50	0.31
S	4.26	1.99	1.00	8.50	7.50	0.29
E	4.02	1.52	1.50	7.00	5.50	0.22
C	5.64	1.96	2.00	10.00	8.00	0.29
Factor	Mean	SD	Min	Max	Range	SE

42 to 49

R	5.14	1.80	0.50	9.50	9.00	0.29
I	5.55	1.59	2.50	8.00	5.50	0.26
A	5.37	2.50	0.00	10.00	10.00	0.41
S	4.95	2.80	0.00	10.00	10.00	0.45
E	3.82	1.53	0.00	8.00	8.00	0.25
C	5.17	2.61	1.00	10.00	9.00	0.42

Factor**Mean****SD****Min****Max****Range****SE**

50 to 56

R	4.83	2.29	1.00	9.00	8.00	0.28
I	5.83	1.51	2.00	9.00	7.00	0.18
A	4.70	2.03	0.50	9.00	8.50	0.24
S	4.99	2.14	0.50	9.00	8.50	0.26
E	4.57	1.61	1.50	8.50	7.00	0.19
C	5.09	2.43	1.00	10.00	9.00	0.29

Factor**Mean****SD****Min****Max****Range****SE**

Over 57

R	3.98	2.31	0.50	9.00	8.50	0.29
I	5.56	1.82	1.50	8.50	7.00	0.23
A	4.91	2.33	0.50	9.50	9.00	0.29
S	5.35	2.24	0.00	9.50	9.50	0.28
E	4.66	1.63	1.00	8.00	7.00	0.20
C	5.54	2.51	0.00	10.00	10.00	0.31

Table 13. Facet Score Norms by Age Groups.

Facet	Mean	SD	Min	Max	Range	SE
18 to 25						
R1	4.75	2.83	0.00	10.00	10.00	0.43
R2	3.80	1.96	0.00	8.00	8.00	0.30
I1	5.52	1.90	2.00	9.00	7.00	0.29
I2	5.64	2.11	1.00	10.00	9.00	0.32
A1	5.93	2.49	1.00	10.00	9.00	0.38
A2	6.50	2.64	1.00	10.00	9.00	0.40
S1	5.05	2.22	2.00	9.00	7.00	0.33
S2	3.23	2.13	0.00	8.00	8.00	0.32
E1	4.84	1.60	1.00	8.00	7.00	0.24
E2	4.23	1.92	0.00	7.00	7.00	0.29
C1	5.45	2.15	1.00	10.00	9.00	0.32
C2	5.07	2.43	0.00	10.00	10.00	0.37
26 to 33						
R1	5.33	2.27	1.00	10.00	9.00	0.33
R2	4.52	2.18	0.00	9.00	9.00	0.32
I1	5.15	1.95	2.00	10.00	8.00	0.28
I2	5.58	2.20	1.00	10.00	9.00	0.32
A1	5.04	2.61	0.00	10.00	10.00	0.38
A2	5.44	2.47	0.00	10.00	10.00	0.36
S1	5.48	2.01	2.00	10.00	8.00	0.29
S2	4.38	2.28	0.00	10.00	10.00	0.33
E1	4.81	1.96	0.00	8.00	8.00	0.28
E2	4.48	2.20	0.00	9.00	9.00	0.32
C1	5.25	2.36	0.00	10.00	10.00	0.34

C2	4.54	2.97	0.00	10.00	10.00	0.43
-----------	------	------	------	-------	-------	------

Facet	Mean	SD	Min	Max	Range	SE
--------------	-------------	-----------	------------	------------	--------------	-----------

34 to 41

R1	5.00	2.80	0.00	10.00	10.00	0.41
R2	4.50	2.29	0.00	10.00	10.00	0.34
I1	6.02	2.42	1.00	10.00	9.00	0.36
I2	5.87	2.11	1.00	10.00	9.00	0.31
A1	5.09	2.40	0.00	10.00	10.00	0.35
A2	5.67	2.11	0.00	10.00	10.00	0.31
S1	5.07	2.23	1.00	9.00	8.00	0.33
S2	3.46	2.53	0.00	9.00	9.00	0.37
E1	4.28	1.36	1.00	7.00	6.00	0.20
E2	3.76	2.30	0.00	9.00	9.00	0.34
C1	5.78	2.06	2.00	10.00	8.00	0.30
C2	5.50	2.68	0.00	10.00	10.00	0.40

Facet	Mean	SD	Min	Max	Range	SE
--------------	-------------	-----------	------------	------------	--------------	-----------

42 to 49

R1	5.45	2.36	0.00	10.00	10.00	0.38
R2	4.84	2.41	1.00	10.00	9.00	0.39
I1	5.74	1.73	2.00	8.00	6.00	0.28
I2	5.37	1.87	2.00	9.00	7.00	0.30
A1	5.26	2.53	0.00	10.00	10.00	0.41
A2	5.47	2.85	0.00	10.00	10.00	0.46
S1	5.71	2.89	0.00	10.00	10.00	0.47
S2	4.18	3.16	0.00	10.00	10.00	0.51
E1	3.71	1.54	0.00	7.00	7.00	0.25
E2	3.92	2.23	0.00	9.00	9.00	0.36

C1	5.66	2.40	2.00	10.00	8.00	0.39
C2	4.68	3.21	0.00	10.00	10.00	0.52

Facet	Mean	SD	Min	Max	Range	SE
--------------	-------------	-----------	------------	------------	--------------	-----------

50 to 56

R1	4.90	2.73	0.00	10.00	10.00	0.33
R2	4.77	2.58	0.00	10.00	10.00	0.31
I1	5.86	1.55	3.00	9.00	6.00	0.19
I2	5.80	2.12	1.00	10.00	9.00	0.26
A1	4.48	2.43	0.00	9.00	9.00	0.29
A2	4.91	2.17	1.00	10.00	9.00	0.26
S1	5.57	2.52	0.00	10.00	10.00	0.30
S2	4.42	2.35	0.00	9.00	9.00	0.28
E1	4.38	1.65	1.00	8.00	7.00	0.20
E2	4.75	2.26	0.00	10.00	10.00	0.27
C1	5.88	2.33	2.00	10.00	8.00	0.28
C2	4.29	3.06	0.00	10.00	10.00	0.37

Facet	Mean	SD	Min	Max	Range	SE
--------------	-------------	-----------	------------	------------	--------------	-----------

Over 57

R1	4.03	2.53	0.00	10.00	10.00	0.32
R2	3.94	2.74	0.00	10.00	10.00	0.34
I1	5.77	1.84	1.00	9.00	8.00	0.23
I2	5.36	2.39	0.00	10.00	10.00	0.30
A1	4.78	2.57	0.00	10.00	10.00	0.32
A2	5.03	2.55	0.00	10.00	10.00	0.32
S1	6.02	2.59	0.00	10.00	10.00	0.32
S2	4.69	2.40	0.00	10.00	10.00	0.30
E1	4.17	1.69	1.00	8.00	7.00	0.21

E2	5.14	2.27	1.00	10.00	9.00	0.28
C1	6.02	2.47	0.00	10.00	10.00	0.31
C2	5.06	3.21	0.00	10.00	10.00	0.40

5. Reliability

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

5.1. Internal Reliability

Internal reliability is a measure of the degree to which items on a scale are internally consistent and measure the same global construct or dimension of a construct. Composite reliability is a robust technique used to assess the internal consistency of items on a scale. It is mathematically defined as the amount of scale score variance accounted for by all underlying factors. Cronbach alpha, while a popular way of determining a measure's internal reliability, has a major weakness in that it assumes all items on a scale are equally important and therefore weighs all items similarly. As a consequence, if there is an increase in the number of items, reliability would artificially increase even though loadings may be low and not significant.

Therefore, composite reliability is a recommended alternative as it makes use of both item loadings and measurement error (Dunn, Baguley, & Brunsden, 2014; Raykov, 1997; Schmitt, 1996). Internal reliability values vary from 0 to 1, with higher values more desirable. As a rule of thumb, a reliability of .70 or higher is recommended (Nunnally, 1978), with values greater than .9 (excellent); .7 to .9 (good); .6 to .7 (acceptable); .5 to .6 (poor) and values less than .5 (unacceptable). Composite reliability should be $\geq .7$.

The formula for Composite Reliability (CR) is:

$$CR = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + (\sum \epsilon_i)}$$

The composite reliability for **VOCATION** was computed on the standardised British sample (N=309). Table 14 describes the internal reliability of **VOCATION**. All the factors showed a composite reliability of above .7 with R and S being .90 and above. On top of this, **VOCATION** has an overall composite reliability of .83, indicating that the scale has a very good internal consistency and is highly reliable.

Table 14. Composite Reliability for Factors.

Factor	CR
R	0.90
I	0.73

A	0.81
S	0.91
E	0.81
C	0.79

5.2. Test-Retest

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 Moment 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 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 **VOCATION**, test-retest reliability was calculated on the UK sample (n=88) at two points in time, between 3-6 weeks apart. Participants ranged in age from 21 to 72 years ($\mu = 48.27$, $\sigma = 12.22$). Test-retest reliability was computed from individual scores for all factors and was calculated using the Pearson Product Moment Correlation Coefficient. Table 15 shows the correlation coefficients between Time 1 and Time 2.

Table 15. Test-retest Reliability.

Factor	r
R	0.84
I	0.76
A	0.84
S	0.80
E	0.74
C	0.87

The test-retest correlations for VOCATION show excellent results, with the lowest correlation being at .74 and the average being .81. Therefore, it can be concluded that VOCATION is a highly reliable measure overtime as a person taking the test at two different points in time, will consistently get the same results.

6. 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 (Standards; 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. **VOCATION** was validated following a thorough multi-stage 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 **VOCATION** look, to potential candidates, like it measures occupational and vocational preferences?' 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 **VOCATION**.

6.1. 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 **VOCATION**, the components of the assessment must correlate highly with corresponding components on other established interest-based 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 16 to 18 show the results detailing the concurrent validity for **VOCATION**.

6.1.1. The ORVIS IPIP

The International Personality Item Pool (IPIP) is a collection of 2413 internationally recognised personality items. The project was initiated in 1996 and is available in over 25 languages.

The IPIP questionnaire for this study was extracted from the IPIP Website (Goldberg, 1999).

Out of the IPIP scales, The Oregon Vocational Interest Scales (ORVIS) was chosen to be the most adequate scale for concurrent validity as it measures the same constructs as Holland's RIASEC (Pozzebon, 2010). Table 16 below refers to which factors from **VOCATION** correspond to which factors

from ORVIS.

Table 16. VOCATION and Corresponding ORVIS Factors.

VOCATION	ORVIS
R	Production
I	Analysis
A	Creativity
S	Altruism
E	Leadership
C	Organization

Participants were instructed to “Without considering salary, please rate the degree to which you like to do the following activities, using the following scale”. Participants were asked to rate each statement on a 5-point Likert scale from Strongly dislike (1) to Strongly like (5).

Table 17. Factor Correlations Between OII-2 and the ORVIS IPIP.

	Production	Analysis	Creativity	Altruism	Leadership	Organization
R	0.50***	0.04	-0.03	-0.04	0.06	-0.11
I	0.12	0.65***	-0.09	-0.16	0.46***	0.31***
A	0.20**	-0.08	0.67***	0.17*	0.28***	-0.14
S	0.04	-0.23**	0.13	0.57***	0.43***	0.00
E	-0.04	0.04	-0.01	0.18*	0.78***	0.42***
C	-0.25**	0.20**	-0.39***	-0.29***	0.09	0.59***

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

Table 18. Facet Correlations Between the OII-2 and the ORVIS IPIP.

	Production	Analysis	Creativity	Altruism	Leadership	Organization
R1	0.47***	-0.18*	0.09	0.04	-0.28***	-0.40***
R2	0.51***	0.19*	0.00	-0.25***	-0.34***	-0.26***
I1	0.05	0.38***	-0.07	-0.27***	-0.37***	-0.12

I2	0.24**	0.43***	0.10	-0.18*	-0.28***	-0.26***
A1	0.25***	-0.16*	0.68***	0.18*	0.01	-0.31***
A2	0.16*	0.03	0.47***	-0.09	-0.18*	-0.28***
S1	0.11	-0.31***	0.18*	0.60***	-0.21**	-0.37***
S2	0.11	-0.26***	0.23**	0.50***	0.24***	-0.19**
E1	0.28*	0.06	0.54***	0.08	0.38**	0.10
E2	0.04	-0.01	0.13	0.28***	0.76***	0.29**
C1	-0.19*	-0.13	-0.30***	-0.34***	-0.52***	0.04
C2	-0.11	0.29***	-0.26***	-0.35***	-0.23***	0.37***

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

The correlations between the factors of the **VOCATION** and the ORVIS scale show excellent results. The correlations between corresponding factors ranged from .5 (between R and Production) to .78 (between E and Leadership), therefore showing an excellent level of concurrent validity, indicating that **VOCATION** is consistent with current existing models of occupational interest. For an in-depth look into the relationship between **VOCATION** and the ORVIS, the correlations between the facets were analysed. The results were mostly consistent with the factor scores, however there were a few differences. For example, E1 correlated highly with creativity, this is likely due to the fact that E1 measures the enterprising mindset, to be innovative and forward thinking, which requires a great deal of creative thought.

6.2. 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 is comprised of and pinpoints positive and negative attributes associated with each dimension.

For interest-based assessments, construct validity is important as it: a) adds legitimacy and acceptability to the standardised measurement of employee interests, 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 **VOCATION** define and classify occupational and vocational preferences in a way that represents theoretical and practical understanding?' For **VOCATION**, construct validity was established using two approaches:

inter-factor correlations and **confirmatory factor analysis**, as detailed in the sections below.

6.2.1. Inter-factor Correlations

The factors in **VOCATION** measure occupational and vocational preferences, however they all measure different facets of interests. In other words, if someone scores highly on a factor, it does not necessarily mean that they will score highly on all factors. Table 19 shows the inter-factor correlations.

Table 19. Inter-factor Correlations.

	R	I	A	S	E	C
R	1.00					
I	0.22*	1.00				
A	-0.15	-0.30**	1.00			
S	0.23*	-0.09	-0.01	1.00		
E	0.27**	0.32**	-0.04	0.66***	1.00	
C	-0.13	0.20	-0.45***	-0.33***	0.10	1.00

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

The correlations in Table 19 demonstrate the relationships between the various aspects measured by **VOCATION**. As can be seen in Table 19, the two factors which correlate most highly are S and E at .66, this is consistent with Holland's theory as these two factors are next to each other in the RIASEC hexagon, whilst A and C are opposite of each and are highly negatively correlated at - .45. However, we have observed differences, for example E and I are opposite each other in the hexagon, but are seen to be correlated in our model. Perhaps this reflects the investigative nature that is required for modern day entrepreneurship, in fact some of the greatest business leaders and entrepreneurs began as technological innovators, so it's certainly possible that the lines dividing the hexagon becoming less rigid.

6.2.2. 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 **VOCATION**, a strict confirmatory approach was adopted as there is only one formulated model for **VOCATION**.

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 **VOCATION**, 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).

Two CFA analyses were conducted on the 6 factor and 12 factor models. Tables 20 and 21 show the goodness of fit indices and model results.

Thurstonian IRT

“Item response theory (IRT) refers to a set of mathematical models that describe, in probabilistic terms, the relationship between a person’s response to a survey question and his or her level of the ‘latent variable’ being measured by the scale” (Hays & Fayers, 2005, p.55). In ipsative tests, a respondent must make a comparative judgement between two or more items to give an answer. Conversely, in normative tests, judgement is measured in absolute terms for each evaluated attribute and does not compare preferences between different attributes. Classical test scoring does not reflect this psychological process of comparison, but rather ignores the fact that respondents indicate a preference not an absolute judgement. Thurstone’s law of comparative judgements (Thurstone, 1927; 1931) generates results based on the comparative judgements of individuals. In order to analyse the

data using Thurstonian IRT modelling, raw data was converted into binary outcomes following the procedure described by Brown and Maydeu-Olivares (2011) for Item response modelling of paired comparisons and ranking data. For **VOCATION**, the 60 questions consisted of blocks of 2 items (response options). The observed responses to the paired comparisons were coded using the formula below, where l indicates the pair $\{a, b\}$;

$$y_l = \begin{cases} 1 & \text{if item } a \text{ is preferred over item } b \\ 0 & \text{if item } b \text{ is preferred over item } a \end{cases}$$

The Thurstonian IRT model was fitted to the binary outcomes of paired comparisons with a structure of both 6 and 12 latent factors. In total, 60 binary outcomes were generated from 60 blocks of two items. Mplus version 7.1 (Muthén & Muthén, 2014) was used to estimate item parameters from tetrachoric correlations using the unweighted least squares (ulsmv) algorithm.

Table 20. Confirmatory Factor Analysis using Thurstonian IRT for the 6 Factor Model.

N	X ²	df	X ² /df	p	RMSEA	P ≤ .05	CFI	TLI
309	2113.507	1635	1.293	0.00	0.031	1.000	0.850	0.837

Table 21. Confirmatory Factor Analysis using Thurstonian IRT for the 12 Factor Model.

N	X ²	df	X ² /df	p	RMSEA	P ≤ .05	CFI	TLI
309	1881.773	1584	1.188	0.00	0.025	1.000	0.907	0.896

Tables 20 and 21 show the model fit results for the two models. Due to the sample size, the more robust estimators of RMSEA and CFI/TLI were used to assess model fit. From these indicators, it can be seen that both models have a good fit and the 12 factor model fits better than the 6. This confirms that **VOCATION** is a valid model. The fact that the 12 factor fits better than the 6 also gives more credence to the idea that each factor in the RIASEC model can be split into two facets.

6.2.3. 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 **VOCATION**. Details of how consequential validity was measured are displayed below, in Chapter 8.

7. VOCATION 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. **VOCATION** is a fair and accurate measure of occupational and vocational preferences, 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)

7.1. 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.

7.1.1. 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 5). 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 4 and 5 show an example of item characteristic curves for uniform and non-uniform DIF.

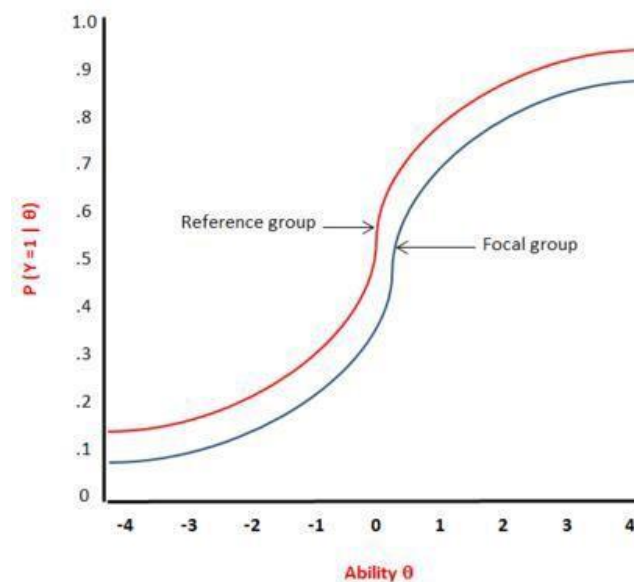


Figure 4. Item Characteristic Curve with Uniform DIF.

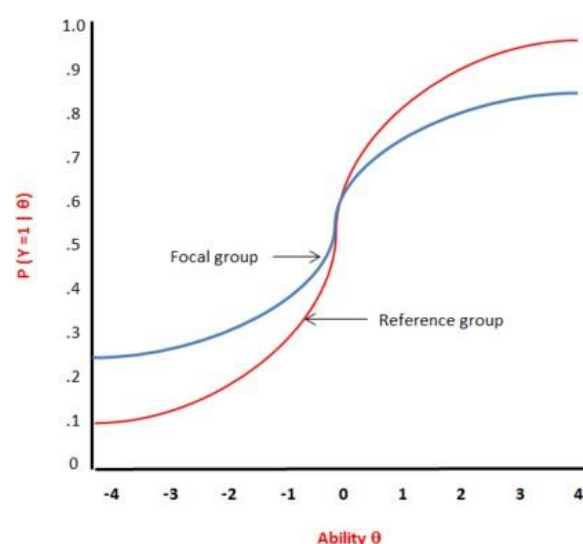


Figure 5. Item Characteristic Curve with Non-uniform DIF.

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 *Artistic* factor will have the same probability of endorsing the items measuring *Artistic*. However, true differences between groups may exist (e.g. some cultures are more interested in Artistic, others less). Such true differences are reflected in the norms (see Chapter 4).

7.1.1.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 **VOCATION** is a fair test for both genders. For this analysis, the UK norm sample was divided according to gender; 167 female and 142 male.

Table 22. Coefficients for Significant DIF and Effect Size for DIF Items.

Facet	Coefficient	R ²	Effect size
R1	7.417	0.023	Negligible
R1	6.010	0.02	Negligible
R2	7.593	0.022	Negligible
I1	8.013	0.034	Negligible
I1	8.225	0.028	Negligible
I1	8.063	0.026	Negligible
I2	7.299	0.023	Negligible
I2	7.115	0.025	Negligible
I2	9.992	0.037	Negligible
A1	6.619	0.019	Negligible
A1	6.345	0.018	Negligible
A2	6.194	0.019	Negligible
A2	8.093	0.028	Negligible
S1	7.823	0.025	Negligible
S2	6.005	0.016	Negligible
S2	9.925	0.037	Negligible
E1	6.710	0.025	Negligible
E1	6.977	0.028	Negligible
E2	6.669	0.025	Negligible

The analysis identified 19 out of the 120 items with a DIF effect for gender. Table 22 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 .13 is considered negligible. Out of the 19 items flagged for DIF 0 showed an effect size of .13 or over, therefore it can be considered that **VOCATION** does not contain any item that functions differently across genders. Two items (Figure 6 and Figure 7) have been chosen to demonstrate uniform and non-uniform DIF. Figure 6 shows uniform DIF for the item.

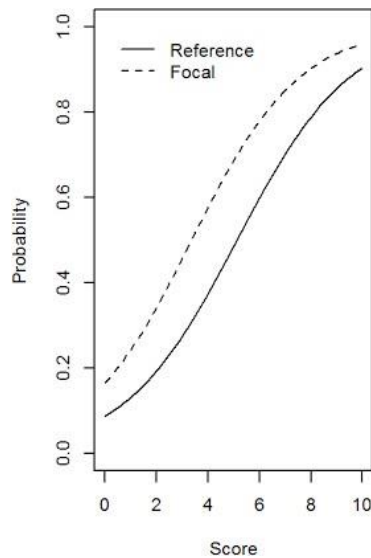


Figure 6. Uniform DIF for the Facet I2.

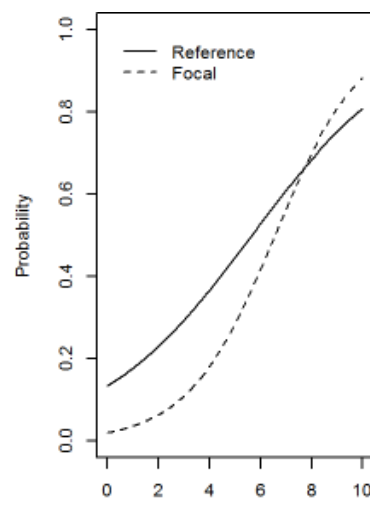


Figure 7. Non-uniform DIF for the Facet A2.

7.1.1.b. DIF Analysis: Age

DIF analysis was performed based on age in order to ensure that **VOCATION** was a fair test for people of all ages. For this analysis, two age groups were defined; 176 aged between 18 and 49 ($M=33.17$; $SD=9.47$), and 133 aged 50 or over ($M=56.45$; $SD=4.73$).

Table 23. Coefficient and Effect Size for DIF Items.

Question	Coefficient	R^2	Effect size
R2	6.240	0.019	Negligible
R2	7.993	0.03	Negligible
I1	8.025	0.026	Negligible
I1	6.059	0.02	Negligible
I2	7.358	0.022	Negligible
S1	6.121	0.019	Negligible
S1	9.268	0.032	Negligible
S1	9.387	0.036	Negligible
S1	9.381	0.024	Negligible

S1	9.211	0.036	Negligible
S2	8.607	0.031	Negligible
S2	6.016	0.018	Negligible
E2	7.478	0.029	Negligible
E2	9.659	0.032	Negligible

The analysis flagged 14 of the 120 items for DIF based on our age groups. However, as demonstrated in Table 23, the effect size for each of these items is negligible. Therefore, it can be concluded that these items are not discriminatory against people of an older age.

7.1.1.c. T-tests and ANOVAs

The DIF analysis demonstrates that **VOCATION** 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 24 to 29).

Age Groups by Factors

Participants were separated into six groups according to their age: 18 to 25 (n=44), 26 to 33 (n=48), 34 to 41 (n=46), 42 to 49 (n=38), 50 to 56 (n=69) and Over 57 (n=64).

Table 24. ANOVA for Factors by Age Group.

	18 to 25	26 to 33	34 to 41	42 to 49	50 to 56	Over 57	F	df	p-value
R	4.27	4.93	4.75	5.14	4.83	3.98	2.254	5.00	0.049*
I	5.58	5.36	5.95	5.55	5.83	5.56	0.733	5.00	0.599
A	6.22	5.24	5.38	5.37	4.70	4.91	2.742	5.00	0.019*
S	4.14	4.93	4.26	4.95	4.99	5.35	2.428	5.00	0.035*
E	4.53	4.65	4.02	3.82	4.57	4.66	2.371	5.00	0.039*
C	5.26	4.90	5.64	5.17	5.09	5.54	0.747	5.00	0.589

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

Table 25. Post-hoc T-tests for Factors by Age Groups.

Age Group	18 to 25	26 to 33	34 to 41	42 to 49	50 to 56
R					
26 to 33	1.00				
34 to 41	1.00	1.00			
42 to 49	0.94	1.00	1.00		
50 to 56	1.00	1.00	1.00	1.00	
Over 57	1.00	0.297	0.917	0.114	0.314
A					
26 to 33	0.609				
34 to 41	1.00	1.00			
42 to 49	1.00	1.00	1.00		
50 to 56	0.009**	1.00	1.00	1.00	
Over 57	0.053	1.00	1.00	1.00	1.00
S					
26 to 33	1.00				
34 to 41	1.00	1.00			
42 to 49	1.00	1.00	1.00		
50 to 56	0.601	1.00	1.00	1.00	
Over 57	0.063	1.00	0.138	1.00	1.00
E					
26 to 33	1.00				
34 to 41	1.00	0.824			
42 to 49	0.595	0.232	1.00		
50 to 56	1.00	1.00	1.00	0.281	
Over 57	1.00	1.00	0.56	0.141	1.00

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

The ANOVA for the RIASEC factors by age groups found significant differences for R, A, S and E. Post hoc analysis was conducted using pairwise t-tests (Bonferroni correction) in order to see where these differences occurred.

The R, S and E post hoc t-tests found there were no significant differences between the six age groups. However, it was found that for factor A, the age group 18 to 25 ($M=6.22$) was significantly higher than the age group 50 to 56 ($M=4.7$, $p=.009$), with the age group Over 57 ($M= 4.91$, $p=.053$) being close to significant. This indicates that younger people may be more interest in artistic endeavours. This may be because the artistic working field is becoming more and more popular, and accessible, in terms of work. Creative careers such as design and illustration are more desirable than ever at present, with companies realising the importance of having such elements in their work. As a result, more and more young people are surrounded by these topics and as the appeal of working in such fields increases, perhaps have more and more interest in the field.

Age Groups by Facets

Table 26. ANOVA for Age Groups by Facets.

Facet	18 to 25	26 to 33	34 to 41	42 to 49	50 to 56	Over 57	F	df	p-value
R1	4.75	5.33	5.00	5.45	4.90	4.03	2.044	5.00	0.072
R2	3.80	4.52	4.50	4.84	4.77	3.94	1.638	5.00	0.15
I1	5.52	5.15	6.02	5.74	5.86	5.77	1.277	5.00	0.273
I2	5.64	5.58	5.87	5.37	5.80	5.36	0.506	5.00	0.772
A1	5.93	5.04	5.09	5.26	4.48	4.78	2.006	5.00	0.078
A2	6.50	5.44	5.67	5.47	4.91	5.03	2.727	5.00	0.02*
S1	5.05	5.48	5.07	5.71	5.57	6.02	1.246	5.00	0.287
S2	3.23	4.38	3.46	4.18	4.42	4.69	2.815	5.00	0.017*
E1	4.84	4.81	4.28	3.71	4.38	4.17	2.822	5.00	0.017*
E2	4.23	4.48	3.76	3.92	4.75	5.14	2.925	5.00	0.014*
C1	5.45	5.25	5.78	5.66	5.88	6.02	0.804	5.00	0.548
C2	5.07	4.54	5.50	4.68	4.29	5.06	1.174	5.00	0.322

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

For an in-depth look at the differences between age groups for the RIASEC factors, an ANOVA was conducted on the facets. The results are shown in Table 26 above. The differences found in A were shown to exist in only A2 and not A1, the differences in S were also found to only be in S2 and not in S1. While the differences in E was found to be in E1 and E2, this further indicates that the RIASEC factors should be broken down into facets to be properly interpreted.

The post hoc t-tests for A2 showed that once again the differences existed between the 18 to 25 group ($M=6.50$) and the 50 to 56 group ($M=4.91$, $p=.013$) and Over 57 group ($M=5.03$, $p=.036$). This difference could be due to the fact that A2 is more centered around creativity and thinking outside the box, both of which are traits more often seen in the younger generations, who have a fresh outlook on the world. Perhaps the reason why this difference is not also seen in A1 is that A1 is aesthetic

appreciation, something which is far less likely to change with age.

The post hoc t-tests for E1 showed that the 18 to 25 group (M=4.84) and 26 to 33 group (M=4.81) were both significantly higher than the 42 to 49 group (M=3.71, $p=.033$ and $p=.035$) This may be because at a younger age taking risks is necessary in order to advance their careers as there is more to gain and less to lose. Whereas at 42 to 49 people are steady in their careers and need stability to settle down for family reasons.

The post hoc t-tests for E2 indicated that the Over 57 group (M=5.14) was significantly higher than the 42 to 49 group (M=3.92, $p=.014$). This may be due to the older generation feeling that they have greater wisdom to share and direct the younger generation, while the 42 to 49 group will be as above in stable careers, most likely with higher managers.

Table 27. Post-hoc T-tests for Facets by Age.

Age Group	18 to 25	26 to 33	34 to 41	42 to 49	50 to 56
A2					
26 to 33	0.578				
34 to 41	1.00	1.00			
42 to 49	0.892	1.00	1.00		
50 to 56	0.013*	1.00	1.00	1.00	
Over 57	0.036*	1.00	1.00	1.00	1.00
S2					
26 to 33	0.395				
34 to 41	1.00	1.00			
42 to 49	1.00	1.00	1.00		
50 to 56	0.189	1.00	0.61	1.00	
Over 57	0.040*	1.00	.153	1.00	1.00
E1					
26 to 33	1.00				
34 to 41	1.00	1.00			
42 to 49	0.033*	0.035*	1.00		
50 to 56	1.00	1.00	1.00	0.703	
Over 57	0.594	0.648	1.00	1.00	1.00
E2					

26 to 33	1.00				
34 to 41	1.00	1.00			
42 to 49	1.00	1.00	1.00		
50 to 56	1.00	1.00	0.285	0.949	
Over 57	0.536	1.00	0.021*	0.112	1.00

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

Gender

In order to analyse differences in scores between genders, an independent samples t-test was conducted to compare all **VOCATION** factors between females (n=167) and males (n=142).

Table 28. Gender T-tests for Factor Scores.

Factor	Female	Male	T.	df	p
R	4.04	5.30	-5.45	307	0.001***
I	5.40	5.94	-2.78	307	0.01**
A	5.59	4.79	3.08	307	0.001***
S	5.18	4.39	3.22	307	0.001***
E	4.31	4.54	-1.26	307	0.21
C	5.48	5.02	1.70	307	0.09

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

Table 29. Gender T-tests for Facets.

Facet	Female	Male	T.	df	p
R1	4.45	5.32	-2.93	307	0.001***
R2	3.62	5.29	-6.40	307	0.001***
I1	5.58	5.82	-1.09	307	0.28
I2	5.22	6.07	-3.54	307	0.001***
A1	5.43	4.55	3.08	307	0.001***
A2	5.76	5.04	2.58	307	0.01**
S1	6.09	4.84	4.66	307	0.001***

S2	4.28	3.95	1.14	307	0.26
E1	4.29	4.46	-0.89	307	0.37
E2	4.34	4.62	-1.11	307	0.27
C1	6.09	5.26	3.20	307	0.001***
C2	4.86	4.79	0.22	307	0.83

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

The t-tests found significant differences between males and females on the factors of R, I, A, and S. Males were found to be significantly more interested in R (male=5.3 vs female=4.04) ($t(t307)=-5.45$, $p=.001$), across both R1 (male=5.32 vs female=4.45) ($t(t307)=-5.45$, $p=.001$) and R2 (male=5.29 vs female=3.62) ($t(t307)=-6.4$, $p=.001$). This could be due to the idea that men are generally expected to be the ones taking manual labour jobs such as mechanics or building, etc.

Males were also found to be significantly more interested in I (male=5.94 vs female=5.40) ($t(t307)=-2.78$, $p=.01$), however this difference can only be seen in I2 (male=6.07 vs female=5.22) ($t(t307)=-3.54$, $p=.001$). This may be because I2 is focused on technology and research, a field of occupational interests which has been predominantly male-focused for many years, meaning that perhaps cognitive biases exist which cause females to be discouraged by such areas. This could be an interesting area of future research, as societal roles are becoming less entrenched, more schemes aimed at encouraging women into STEM fields are being implemented. We could expect these differences to reduce in younger generations and it may be possible to measure the effect of these schemes using occupational interest inventories.

Females were found to be significantly more A (male=4.79 vs female=5.59) ($t(t307)=3.08$, $p=.001$), for both A1 (male=4.55 vs female=5.43) ($t(t307)=3.08$, $p=.001$) and A2 (male=5.04 vs female=5.76) ($t(t307)=2.58$, $p=.01$). The A1 difference may be due to a tendency for women to be more in touch with their emotions and are therefore, perhaps more willing to express their emotions through artistic endeavours. The A2 difference could be down to an inclination for women to be more intuitive and therefore more interested in roles which require design and imagination.

Females were also found to be significantly more S (male=5.18 vs female=4.39) ($t(t307)=-3.22$, $p=.001$), however this difference was only found for S1 (male=4.84 vs female=6.09) ($t(t307)=4.66$, $p=.001$). This difference may exist as a result of women being expected by society to hold more social roles (nurses, teachers, customer service, etc.) which require more dedication to others.

8. Additional Information and Considerations

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

Report generation software for **VOCATION** provides graphs and personalised comments for all the 6 factors and 12 facets that relate to the candidate's score. The report should only be interpreted by professionals who have completed *Central Test's* accredited training course and are familiar with this manual. Training in and knowledge of the assessment process are vital as scores should be interpreted in light of information regarding reliability, validity, standard error of measurement, and any accommodations to the assessment or assessment session that were made. Scores should be presented within the context of the assessment and in a clear, meaningful and appropriate way for the target audience.

8.1. Interpreting Scores

It is important to note that all assessments by *Central Test* have been developed for HR purposes only. Therefore, no clinical conclusions can be drawn from the results – not even from extreme scores. There are six different levels of comments in total. Levels are determined by a candidate's score. Table 30 provides an example of descriptors used to describe different levels of a factor.

Extreme scores are defined as the lowest and highest possible values on a factor. For **VOCATION**, the minimum score is set at 0 and the maximum score is set at 10. Given the ipsative format of the assessment, it is impossible to obtain extreme scores on all factors.

Each factor consists of two facets. High scores on a factor indicate that there is a high probability that the person who is being evaluated can be characterised as having an interest in that factor. Middle scores (i.e. 4-7) indicate a high congruence with the norm group and may be less of a preference for the individual. However, all comments should be viewed as hypotheses that need to be confirmed with other sources of data, such as interviews or further assessments, CV content and references.

Table 30. Example Descriptors of Different Levels of a Trait in Report Comments for **VOCATION**.

Level 1 (score 0 to 2)	Level 2 (score 2.1 to 3.5)	Level 3 (score 3.6 to 5.0)	Level 4 (score 5.1 to 6.4)	Level 5 (score 6.5 to 7.8)	Level 6 (score 7.9 to 10)
Does not seem to indicate an interest	Indicates a very low interest	Indicates a low interest, may occasionally enjoy	Indicates an interest, but not a preference	Indicates an interest	Indicates a strong interest

8.2. Organisational Need: Career selection, Recruitment, and Development

Through the identification of 21st century work interests, **VOCATION** is an effective tool for both the employer, in the selection of job candidates on the basis of the occupational areas that suit a candidate's vocational interests, and the job candidate, in the successful selection of an appropriate career navigation.

Choosing an appropriate career path is a difficult task, and with the overwhelming number of working fields that exist in the modern-day occupational world, it can be confusing to know which to take. **VOCATION** helps to pinpoint a candidate's most appropriate job as a result of their recreational interests. By choosing a career that will make use of their interests, an individual can find greater job satisfaction.

This then leads to the importance of interest assessment in recruitment. In order to successfully hire a candidate that will feel satisfied with their job and as a result want to remain with the company, job satisfaction of the employee becomes significant to the recruiter. By using **VOCATION**, an employer will be able to infer how appropriate a job may be for a candidate based on an objective assessment of their interests. If the endorsed interests of said candidate do not align with those of the job role in question, there may not be a good match between the two.

Recruitment and selection is an important organisational function which can be very time consuming and expensive. Furthermore, bad recruitment decisions have additional effects on organisational productivity, capability and sustainability. Whilst someone in financial and legal administration for example, may feel more comfortable if they have a preference for tasks that involve following strict rules, someone working in marketing may be better suited to have an interest in tasks that are more imaginative and creative. It is recommended that prior to administering **VOCATION**, recruiters have a detailed job description and person specifications based on a job analysis. **VOCATION** helps recruiters to customise job profiles in their client account by specifying desired scores for each factor. The assessment allows the recruiter to easily compare the target profile with a candidate's profile.

Furthermore, **VOCATION** can be used in line with performance evaluations, which are a key component of employee development and success. Performance reviews should be fair and balanced assessments of an employee's achievements and capabilities, and should give the opportunity to discuss expectations and accomplishments. **VOCATION** supports HR practitioners with performance reviews that can help to discuss job performance in relation to occupational and vocational interests; if an individual is not performing as they could be at work, it may be down to a lack of interest in the particular area that they are working on. Identifying this can help the recruiter to identify whether or not the employee should be transferred to another area of the company, or to a different aspect of a project.

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

References

- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69(1), 49-56.
- Bartram, D. (2007). Increasing Validity with Forced-Choice Criterion Measurement Formats. *International Journal of Selection and Assessment*, 263-272.
- Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology*, 56, 45–74. <http://dx.doi.org/10.1111/j.1744-6570.2003.tb00143.x>
- Bullock, E. E., Andrews, L. M., Braud, J., & Reardon, R. C. (2009-2010, Winter). Holland's theory in an international context: Applicability of RIASEC structure and assessments. *Career Planning and Adult Development Journal*, 25(4), 29-58.
- Brown, A., & Maydeu-Olivares, A. (2011, 06). Item Response Modeling of Forced-Choice Questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502. doi:10.1177/0013164410375112
- Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, 18(1), 36-52. <http://doi.org/10.1037/a0030641>
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*. <http://doi.org/10.2333/bhmk.30.99>, 99-121.
- Christiansen, N., Sliter, M., & Frost, C. T. (2014b). What employees dislike about their jobs: Relationship between personality-based fit and work satisfaction. *Personality and Individual Differences*, 71, 25-29. <http://doi.org/10.1016/j.paid.2014.07.013>
- 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>
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor (NEO-FFI) inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- 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?* Germany: Lambert Academic Publishing.

- Cruise, P. A. (2012). The stability and organisational value of personality assessment. *Psyche Magazine* Vol. 64, Spring, pg 14-15. UK: Psychometrics Forum. Retrieved from <http://www.psychometricsforum.org/uploads/Psyche/Psyche%2064%20-%20Spring%202012.pdf>
- 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>
- Dingemanse, S. A., De Fruyt, F., Vergauwe, J., & Wille, B. (2015). A closer look at the psychological diversity within Holland interest types: Construct validation of the Career Insight. *Consulting Psychology Journal: Practice and Research*, 67(3), 234-257.
- Dingemanse, S. A., Van Amstel, B., De Fruyt, F., & Wille, B. (2007). *Loopbaan Inzicht Vragenlijst. Handleiding [Career Insight Questionnaire. Manual]*. Amsterdam: Harcourt Assessment.
- 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>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412. <http://doi.org/10.1111/bjop.12046>
- 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>
- Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *Academy of Management Review*, 18(4), 694-734. <http://doi.org/10.5465/AMR.1993.9402210155>
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

- Gottfredson, G. D. (1999). John L. Holland's contributions to vocational psychology: A review and evaluation. *Journal of Vocational Behaviour*, 55, 15-40.
- Harmon, L. W., & Zytowski, D. G. (1980). Reliability of Holland Codes Across Interest Measures for Adult Females. *Journal of Counselling Psychology*, 27 (5), 478-483.
- 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>
- 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>
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant Reactions to Selection Procedures: An Updated Model and Meta-Analysis. *Personnel Psychology*, 57(3), 639-683. <http://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Hays, R. D., & Fayers, P. (2005). Evaluating multi-item scales. *Assessing quality of life in clinical trials*, 41-54.
- Holland, J. L. (1958). A personality Inventory Employing Occupational Titles. *Journal of Applied Psychology*, 42(5), 336 – 342.
- Holland, J.L. and Holland, J. E. (1977). Vocational indecision: More evidence and speculation. *Journal of Counseling Psychology*, 24(5), 404-424.
- Holland, J. L. (1985). *Vocational preference inventory (VPI): Professional manual*. Odessa, Florida, US: Psychological Assessment Resources.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments (3rd ed.)*. Odessa, Florida, US: Psychological Assessment Resources.
- Holt, J. C., Duijn, M. A., & 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
- 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.
- 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>

- 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>
- Iddekinge, C. H., Putka, D. J., Roth, P. L., & Lanivich, S. E. (2011). A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, 1167-1194.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International.
- Kaplan, R. M., & Saccuzzo, D. P. (2008). *Psychological Testing: Principles, Applications, and Issues (7th ed.)*. Belmont, CA: Wadsworth Publishing Co Inc.
- Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of Big Six Interests and Big Five Personality Factors. *Journal of Vocational Behavior*, 61, 217-239.
- 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>
- 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.
- 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>
- Muthén, L. K., & Muthén, B. O. (2014). *Mplus User's Guide (7th ed.)*. Los Angeles, CA: Muthén & Muthén.
- 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.
- Nauta, M. M. (2010). The Development, Evolution, and Status of Holland's Theory of Vocational Personalities Reflections and Future Directions for Counseling Psychology. *Journal of Counseling Psychology*, 11-22.

- Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., & Goldberg, L. R. (2010). Psychometric characteristics of a public-domain self-report measure of vocational interests: The Oregon Vocational Interest Scales. *Journal of Personality Assessment*, 92, 168-174.
- 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>
- Schinka, J. A., Dye, D. A., & Curtiss, G. (1997). Correspondence Between Five-Factor and RIASEC Models of Personality. *Journal of Personality Assessment*, 355-368.
- 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>
- 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)*.
- Tett, R. P., & Simonet, D. V. (2011). Faking in Personality Assessment: A « Multisaturation » Perspective on Faking as Performance. *Human Performance*, 24(4), 302-321.
- Thurstone, L. L. (1927). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology*, 21(4), 384-400.
<http://doi.org/10.1037/h0065439>
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38(5), 406-427.
<http://doi.org/10.1037/h0069792>
- Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada: Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia.

CONTACT US

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

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



www.centraltest.com