

TECHNICAL MANUAL & USER GUIDE

ORIENTATION

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

*The assessment **ORIENTATION** and the Central Test logo are registered trademarks.*

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

Table of Contents

Introduction	4
1. General Description and Applicability	5
1.1. Factor Definitions	5
1.2. Main Areas of Application	6
1.2.1. Use with other assessments	6
1.3. Questionnaire Structure	7
1.4. Languages and Special Requirements	7
1.5. Administration	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	11
1.8. Report Content	11
2. Conceptual Framework and Development	13
2.1. Theoretical Background	13
2.1.1. John Holland's Model	13
2.1.2. Links between the RIASEC dimensions and personality traits	14
2.1.3. Subdivision of the RIASEC factors	14
2.2. Development	15
2.2.1. Development Phases for the English Version	15
2.2.2. Combined profiles	15
2.2.3. Ipsative Response Format	16
3. Scoring	18
3.1. Raw Scores	18
3.2. Conversion of Raw Scores to Standard Scores	18
3.3. Scoring in the Reports	18
4. Psychometric Properties	20
4.1. Sample Descriptives	20
4.2. Norms	22
4.3. Reliability	26
4.3.1. Internal reliability	27
4.3.2. Marginal reliability	28

4.4. Validity	29
4.4.1. Construct validity	30
4.4.2. Inter-factor Correlations	31
4.4.3. Confirmatory factor analysis	32
4.4.3.a. Thurstonian IRT	33
5. ORIENTATION and Equal Opportunities	35
5.1. Measurement Equivalence	35
5.1.1. Differential Item Functioning (DIF)	36
5.1.1.a. DIF Analysis: Gender	38
5.1.1.b. DIF Analysis: Age	38
5.2. T-tests and ANOVAs	39
5.2.1. Gender	39
5.2.2. Age Groups by Factors	40
5.2.3. Study Level Groups by Factors	42

Introduction

ORIENTATION assesses career interests based on personality. Its report is particularly useful in career guidance processes for the younger audience as both the item design and the report have been specifically developed for this age group. In addition, the report includes information on the education or training required for certain professional fields.

The purpose of **ORIENTATION** is to provide a ranking of career interests to support career guidance processes and help with study or career choice.

The following sections of this *Technical Manual & User Guide* describe the factors, questionnaire structure, report sections, and provide information on score administration and interpretation, the theoretical model, details on the assessment's development, norms, reliability and validity studies conducted to date.

1. General Description and Applicability

ORIENTATION is an assessment designed to assess career interests based on personality, providing a ranking of those interests. It is built on the RIASEC model and serves as a support tool in career guidance processes for young people and students. The following section explains the dimensions and factors assessed by the assessment.

1.1. Factor Definitions

Professional success is mostly dependent on the alignment between an individual's personality and interests and the tasks they perform. Working in a profession that matches one's interests makes it easier to stay motivated and gives greater meaning to the tasks carried out.

REALISTIC: People with "Realistic" interests prefer physical or manual activities. Goal-oriented, they are drawn to tasks that require skill and physical endurance. They enjoy activities carried out outdoors or manual activities, such as assembling or repairing objects.

INVESTIGATIVE: The "Investigative" interest includes people who enjoy exploring and developing theories. They like researching, conducting studies or problem-solving to gain a better understanding of a field of knowledge or the world around them.

ARTISTIC: People with an "Artistic" interest tend to express themselves creatively through music, drawing or other forms of artistic expression. They often feel the need to convey their ideas and emotions in an artistic way. They prefer activities that allow them to be intuitive and imaginative and may feel resistant toward conventional projects that do not allow for improvisation.

SOCIAL: People with a "Social" interest prefer to be around others, communicating or helping whenever possible. They enjoy meeting new people, as well as cooperating and feeling like they are part of a group. They also like connecting with others, listening attentively and are sensitive to other people's needs.

ENTERPRISING: People with an "Enterprising" interest often enjoy activities that involve some level of risk and influencing others. They may like guiding, coordinating, persuading others about their ideas or engaging in negotiations. They enjoy taking the initiative and interacting with others.

CONVENTIONAL: The term "Conventional" refers to work that is organised and

meticulous, involving the application of rules and planning. People with a “Conventional” profile like to follow clear rules, be organised, and work efficiently. They are not discouraged by routine tasks.

1.2. Main Areas of Application

ORIENTATION is especially useful for educational institutions in their career development and guidance programmes. It is also valuable for companies and organisations in the professional development of new, young, or trainee staff, in apprenticeship programmes. It is also applicable for organisations or social centers that support professional development, such as public employment centers or social development organisations. Its scope of application includes:

- Career development and career guidance coaching.
- Initial assessment in personalised intervention programmes aimed at staff in companies and organisations, students in higher education institutions and users in community organisations or centers dedicated to career assistance and guidance.

1.2.1. Use with other assessments

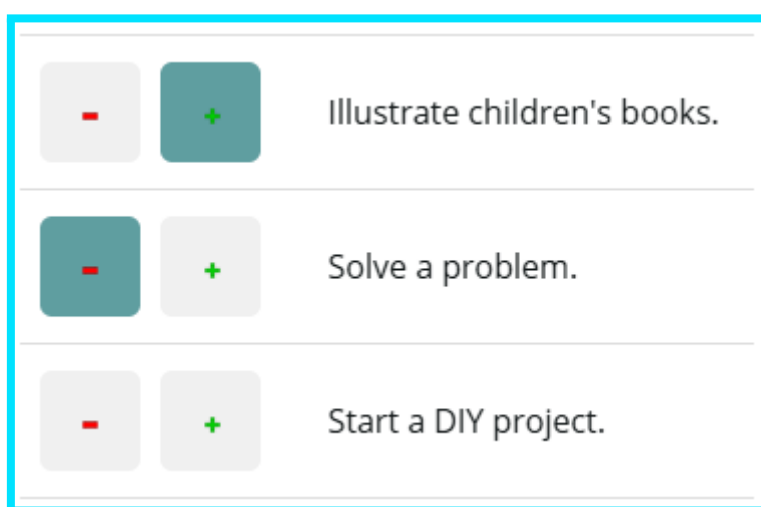
It is possible to combine **ORIENTATION** with other assessments to obtain a more complete profile of the individual. For example, in the educational field, it can be combined with:

- **PROFESSIONAL PROFILE 2** - to create a profile of the competencies needed for most job roles.
- **ENTREPRENEUR TEST** - to gain a more tailored profile of development needs in cases where there is an interest in entrepreneurship.
- **MOTIVATION+** - to complement the entrepreneurial profile with a motivational profile, which is useful when deciding which projects, roles, or professions might be the best fit.
- **EMOTION 2.1** - to provide additional information on emotional intelligence competencies, which are key to professional success.

In addition, for more precise information on international readiness, language assessments can be used: **BUSINESS ENGLISH TEST**, **FRENCH LANGUAGE TEST**, and/or **GERMAN LANGUAGE TEST**.

1.3. Questionnaire Structure

ORIENTATION consists of 40 questions in total, with an average completion time of 8 minutes. In each question, the participant must choose the interest they identify with the most and the interest they identify with the least.



The image shows a screenshot of a questionnaire interface. It features three rows of questions, each with two buttons (a red minus sign and a green plus sign) and a text description. The first row has the minus button selected and the text 'Illustrate children's books.' The second row has the plus button selected and the text 'Solve a problem.' The third row has the minus button selected and the text 'Start a DIY project.' The entire interface is enclosed in a light blue border.

☒	☐	Illustrate children's books.
☐	☒	Solve a problem.
☒	☐	Start a DIY project.

Figure 1. Example of a question in **ORIENTATION**.

To ensure a pleasant answering experience, each question is presented individually, and once it is answered, the assessment automatically moves on to the next question. In addition, the upper left corner of the screen displays the time spent on the assessment, the number of questions viewed and the number of questions answered in relation to the total number of questions.

1.4. Languages and Special Requirements

ORIENTATION is administered online and is compatible with any current operating system, including Windows, MacOS and Linux. Like most assessments today, it can be taken remotely without the need for supervision. However, group assessments can also be conducted under controlled conditions if needed.

It is important to note that reports are generated online only. Paper-and-pencil

administration is not recommended, as it requires manually entering responses into the system to generate the report, which increases the likelihood of errors.

ORIENTATION is currently available in English, French and Spanish. For more information on language availability, you can contact the consulting service at consulting@centraltest.com.

ORIENTATION was designed for use by individuals from all social groups and/or with perceptual difficulties. The assessment's on-screen presentation can be adjusted according to individual needs. For example, if there is a visual impairment or a larger font size is required, it can be increased using "CTRL" and "+".

The assessment instructions are available at all times. The assessment publisher, *Central Test*, can be contacted for any guidance regarding administration.

Although response time is recorded, the assessment does not have a time limit, so candidates who require more time to complete it are not penalised.

1.5. Administration

ORIENTATION is designed to be administered online through *Central Test's* virtual platform, so supervision is usually not required. In cases where supervised administration is conducted, the following recommendations can be taken into account.

Although the instructions for completing **ORIENTATION** can be read on the screen, it can be helpful to address any questions before starting the assessment.

- If the candidate has questions after the assessment has begun, respond in a way that does not disturb others.
- If the candidate asks for clarification on a question, answer in a way that avoids influencing their response.
- If there are language-related difficulties, consider administering the assessment in the language most suitable for the candidate.

Once a candidate has been invited to take the assessment, they will receive an email invitation with a link to the website and their access credentials. The questionnaire link is protected with a unique password assigned to each individual account. It is possible

to set a deadline for completing the assessment. The invitation can also be configured so that the assessment must be completed from a specific IP address.

Completion time is recorded for supervision and follow-up purposes, allowing the identification of assessments completed too quickly or those requiring unusually long completion times, to investigate potential scoring errors. It is preferable for participants to answer spontaneously, without rushing, and without overthinking their responses.

1.6. Technical and Security Issues

Guidelines governing the security of administering **ORIENTATION** are respected to ensure that every administration of **ORIENTATION** 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. As of 25th May 2018, *Central Test* complies with the GDPR (General Data Protection Regulation) regulations made by the European Parliament & Council.

Assessment 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 their personal data be permanently deleted. To do this, they must contact either the organisation or Central Test directly, who will delete their 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, 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 uses Google Cloud Platform, offering hosting on the same infrastructure that Google uses internally for products such as its search engine. All servers used are located within the European Union and are run on a Unix/Linux type operating system, featuring:

- 6 Compute Engines
- Load Balancing
- High performance of the SSD disk (680,000 IOPS)
- A 2-hour Incident Response 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:

- Firefox 35 or above
- Google Chrome 30 or above
- Safari 11 or above
- Microsoft Edge 12 or higher

1.6.4. Computer-generated Reports

Reports can be generated within minutes of candidates completing the questionnaire. 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 vary from country to country. However, the following points generally apply:

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

1.8. Report Content

The **ORIENTATION** report is designed for career development and presents the ranking of a person's career interests, along with information on education and professional suitability. Below are its sections, along with useful details for interpretation.

Header: The header at the top of the report provides the assessment name, date it was taken and completion time. It allows for information verification and tracking.

Profile summary and survey: This section highlights the highest and lowest areas of interest.

Graph: The graph provides a quick and easy way to view the scores. It is a radar graph, with a highlighted dashed line indicating the midpoint. The score for each factor is shown in brackets. The purpose of the graph is to make interpretation easier by offering a visual summary of the profile.

Personalised comments by factor: This section provides the complete profile information.

- **Score:** Includes the score displayed on a scale from 0 to 10. In general, a score above 6.5 indicates a high level of interest, between 3.5 and 6 indicates a medium level, and below 3 indicates a low level.
- **Definition:** Includes the definition of the assessed factor, serving as a reference for interpreting the score. It is accompanied by information on potential skills, learning style, activity sector and examples of social and leisure activities related to that interest.

Combined profile and professional fit: Career interests often manifest in combinations, and analysing them together provides a better understanding of the profile. In the report, 15 combinations of career interests are ranked based on the scores obtained. For each combination, the percentage of match is shown, with details on the relevant factors and ideal scores. A definition is also provided, along with examples of professions that best match the combination.

Matchings with studies and professions: This section lists professions based on the percentage of match and the results obtained in the assessment, with details on the relevant factors and ideal scores. A match by professional sector is also presented, including guidance on which studies or training programmes may be relevant.

2. Conceptual Framework and Development

In the development of a psychometric test, it is necessary to carry out preliminary research to understand the theoretical models that can serve as a framework, their constructs, and their relationship with behavior. In the development of **ORIENTATION**, the relationship between professional interests and behavior, as well as job fit, was investigated.

2.1. Theoretical Background

2.1.1. John Holland's Model

The Holland Typology, also known as the RIASEC Model, was developed by Holland (1958) and is based on his theory that a person's vocational interest or career choice is determined by underlying "personality types".¹ The model is developed based on 6 basic personality types or a combination of the 6 types. RIASEC is an acronym for Realistic (Doers), Investigative (Thinkers), Artistic (Creators), Social (Helpers), Enterprising (Persuaders) and Conventional (Organisers). Each type is characterised by a constellation of interests, preferred activities, beliefs, abilities, values, and characteristics.²



Figure 2. Structural framework of Holland's (1977) RIASEC model of occupational themes.

¹ Holland, J. L. (1958). A personality inventory employing occupational titles. *Journal of Applied Psychology*, 42(5), 336–342. <https://doi.org/10.1037/h0047330>

² 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*, 57(1), 11–22. <https://doi.org/10.1037/a0018213>

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.³ Holland also stressed that most people look for a work environment that allows them to use the skills and abilities they possess, and take on roles that they are comfortable with, which brings about a paradigm shift in measuring career-related decisions, satisfaction, and work stability.⁴

2.1.2. Links 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, and found that there were significant correlations 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'.⁵ Further overlap was also found between Conventional and 'Conscientiousness' and Social and 'Agreeableness'.⁶

2.1.3. Subdivision of the RIASEC factors

Dingemanse, Van Amstel, De Fruyt, and Wille (2007) explored the separation within RIASEC types and created the Career Insight Questionnaire (CIQ), making subdivisions for each RIASEC factor.⁷ For example, Investigative was subdivided into 'Science' and 'Theory'. Dingemanse et al. (2015) found that correlations between personality traits from the Big Five were not consistent across aspects of each subdivision.⁸ 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 individuals may have an interest in only some parts of the Investigative

³ Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.).

Odessa, FL: Psychological Assessment Resources; Gottfredson, G. D. (1999). John Holland's contributions to vocational psychology: A review and evaluation. *Journal of Vocational Behavior*, 55(1), 15–40. <https://doi.org/10.1006/jvbe.1999.1695>

⁴ 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*, 57(1), 11–22.

<https://doi.org/10.1037/a0018213>; Bullock, E. E., Andrews, L. M., Braud, J., & Reardon, R. C. (2010). Holland's theory in an international context: Applicability of RIASEC structure and assessments. *Career Planning and Adult Development Journal*, 25(4), 29–58.

⁵ 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(2), 217–239. <https://doi.org/10.1006/jvbe.2001.1854>

⁶ 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(1), 45–74. <https://doi.org/10.1111/j.1744-6570.2003.tb00143.x>

⁷ Dingemanse, L., Van Amstel, F., De Fruyt, F., & Wille, B. (2007). The Career Insight Questionnaire (CIQ): Development and psychometric evaluation. *International Journal for Educational and Vocational Guidance*, 7(3), 141–157. <https://doi.org/10.1007/s10775-007-9123-6>

⁸ Dingemanse, L., Wille, B., De Fruyt, F., & Van Amstel, F. (2015). Disentangling RIASEC types into aspects: Relations with personality traits and vocational interests. *Journal of Vocational Behavior*, 88, 28–37. <https://doi.org/10.1016/j.jvb.2015.02.002>

interest. As such, for **ORIENTATION**, it was decided that splitting each of the RIASEC factors into two parts would yield a more accurate representation of the range of existing occupational interests.

2.2. Development

ORIENTATION in English is the adaptation of the **Atout Orientation** test, originally developed and validated for the French language and audience. It was created in response to the need in the English market for a career interest assessment designed to support career guidance for students, graduates, and young professionals.

2.2.1. Development Phases for the English Version

- **Item translation:** Items from the latest French version were translated into English and reviewed by the Central Test R&D team. The review ensured that the language was clear, culturally appropriate, and suitable for the target age group.
- **Pretesting:** The translated questionnaire was pretested with English-speaking populations from different countries. Analyses were conducted to evaluate item quality, factor structure, and reliability. Items with unsatisfactory results were revised or removed.
- **Final version:** Following these reviews and refinements, the final version of **ORIENTATION** was implemented on the Central Test platform and made available to clients.

2.2.2. Combined profiles

Most RIASEC-based tools determine occupational interest profiles by identifying two or three primary interests from the six RIASEC factors and combining them to indicate which professional field or activity might appeal to the individual. Holland emphasised the importance of using combinations of factors, as explained in the manual for his self-administered test, the SDS.⁹ Holland used a combination of three factors, requiring the user to consider all possible variations of these three-factor combinations. Consequently, an occupational interest profile should not be inferred solely from the degree of interest in a single RIASEC factor.

The combination of RIASEC factors to create an individual's occupational interest profile

⁹ Holland, J. L. (1970). *Self-Directed Search (SDS): A career planning tool for adults and students*. Odessa, FL: Psychological Assessment Resources.

is supported by various studies demonstrating the overlap between the six RIASEC factors and personality traits. For example, both Barrick et al. (2003) and Larson, Rottinghaus, and Borgen (2002) found a strong relationship between extraversion and the Enterprising and Social factors in Holland's model.¹⁰ These studies show that personality is complex and that it is normal for individuals to have different professional interests. Therefore, it is important for profiles to include more than one RIASEC factor.

In the **ORIENTATION** report, the percentage match for 15 two-factor combinations is presented, as shown in the image. For each combination, an interpretation commentary and examples of professions are provided.

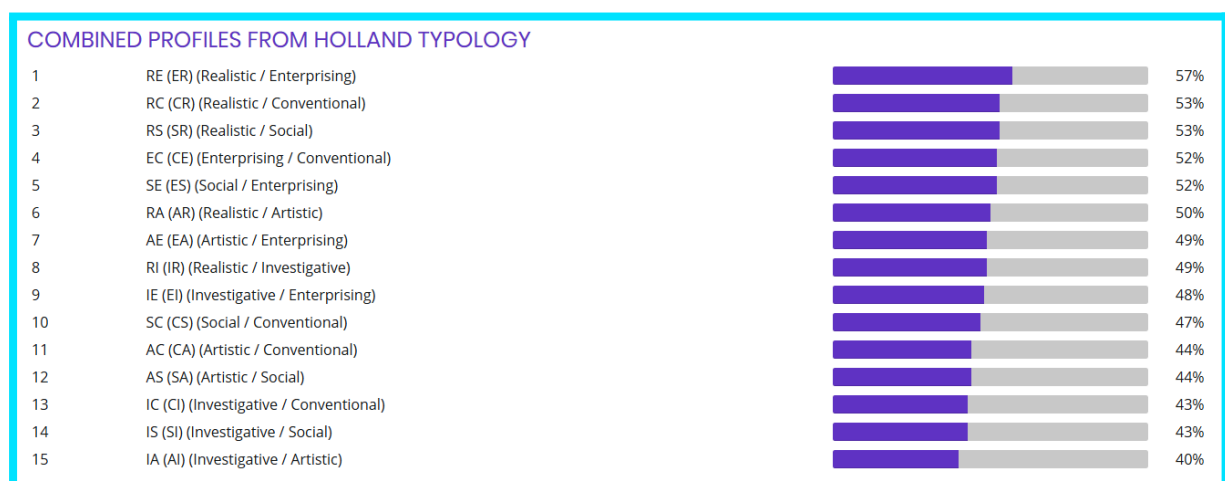


Figure 3. Matching of the 15 combined RIASEC profiles in the report.

2.2.3. Ipsative Response Format

In **ORIENTATION**, the items are grouped into three options per question, using a forced-choice format in which respondents must select the option they identify with the most and the one they identify with the least. Ipsative or forced-choice items require respondents to make comparative judgments such as “most like me” or “least like me.” This format helps overcome many of the limitations associated with Likert scale items.¹¹

¹⁰ 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(1), 45–74. <https://doi.org/10.1111/j.1744-6570.2003.tb00143.x>; 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(2), 217–239. <https://doi.org/10.1006/jvbe.2001.1854>

¹¹ Brown, A., & Maydeu-Olivares, A. (2011a). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460–502. <https://doi.org/10.1177/0013164410375112>; Chan, K. Y. (2003). An exploratory study of the reliability and validity of forced-choice personality measures. *Educational and Psychological Measurement*, 63(5), 835–850. <https://doi.org/10.1177/0013164403251310>; 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. <https://doi.org/10.1111/j.2044-8325.1991.tb00556.x>; Holt, J. M., van Duijn, M. A. J., & Boomsma, D. I. (2010). Scale construction and evaluation in forced-choice questionnaires. *Multivariate Behavioral Research*, 45(3), 299–328. <https://doi.org/10.1080/00273171.2010.483394>

The ipsative format is particularly relevant for career interest tests because it provides a ranking of an individual's interests from most important to least important.

Below are some of the advantages of using forced-choice items:

- Avoiding response biases such as acquiescence bias (agreeing with everything).
- Limiting the possibility of response distortion and social desirability bias.
- Allowing a more precise comparison of an individual's scores across interests.

Using a combination of Item Response Theory (IRT) methodologies and Thurstonian comparison models, Brown & Maydeu-Olivares showed that 'Absolute' values could be obtained, allowing for meaningful comparisons between individuals, and distortions that lead to issues in validity can be corrected.¹² For **ORIENTATION**, Thurstonian IRT was used to conduct the analyses.

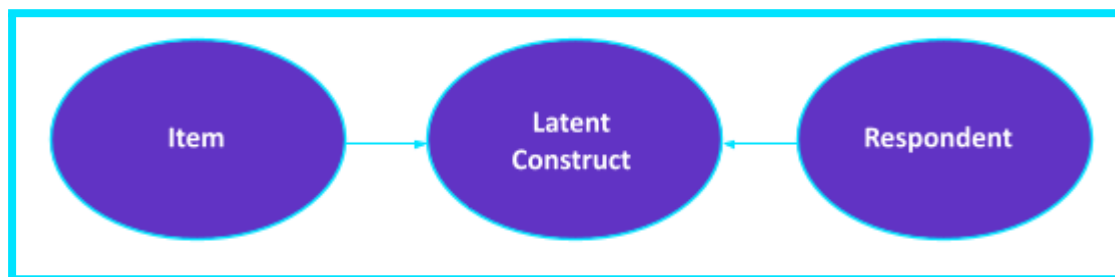


Figure 4. Model of Thurstonian IRT analyses used in the development of **ORIENTATION**.

¹² Thurstone, L. L. (1927). *A law of comparative judgment*. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>; Thurstone, L. L. (1931). *The indifference function*. *Journal of Social Psychology*, 2(2), 139–167. <https://doi.org/10.1080/00224545.1931.9919096>; Brown, A., & Maydeu-Olivares, A. (2011a). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460–502. <https://doi.org/10.1177/0013164410375112>

3. Scoring

3.1. Raw Scores

A raw score is the total number of responses endorsed by a candidate for each factor measured, evaluating an individual's score relative to the score of other delegates who completed the same assessment.

3.2. Conversion of Raw Scores to Standard Scores

For each **ORIENTATION** factor, the maximum possible raw score depends on the number of items included under each factor, while the minimum possible raw score is always 0.

The conversion of raw scores is carried out in a two-step process:

Step 1: Ipsative scores are calculated for each scale by summing the values assigned to all forced-choice responses within that factor. In each response, the option selected as “most” is scored 1, the option selected as “least” is scored -1, and the unchosen option is scored 0.

Step 2: The scores are then transformed to a 0-to-10 scale using the following formula:

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

Where:

- max is the maximum possible score for each factor
- min is the minimum possible score for each factor

3.3. Scoring in the Reports

ORIENTATION reports reflect the scoring process, with each individual receiving a score from 0 to 10. The report provides a detailed description of each of the six RIASEC factors assessed in the assessment. The level corresponding to the candidate's score is selected from six possible levels, detailing their preferences and the activities or occupations they are most inclined toward.

Higher scores reflect stronger intensity, while lower scores reflect weaker intensity. For

example, a candidate with a low score in a particular factor might be described as “does not seem to show much interest,” whereas a high score might indicate “a strong interest” in the relevant factor. Scores should be interpreted in relative terms, with the ranking of profiles being more important than the absolute score.

The combined profiles in the report also reflect the scoring process. These profiles represent the combination of the three factors with the highest scores. They are calculated by considering that the first two interest areas have the highest and equal importance, while the third interest area carries slightly less weight.

4. Psychometric Properties

4.1. Sample Descriptives

The sample consisted of $N = 300$ candidates from English-speaking countries ($n = 150$ males, $n = 146$ females, $n = 4$ undisclosed) ranging in age from 18 to 77 years ($M = 40.00$, $SD = 15.68$). Participants represented a wide range of occupational groups, including technical, administrative, service, managerial and non-managerial roles, and industries such as banking and professional services, marketing and the service sector, manufacturing, and product development.

Education level was coded into six categories: Level 1 (*Master's degree, PhD, Engineering or Business School, MBA*), Level 2 (*Bachelor's degree, first-year Master's*), Level 3 (*two years of vocational or university studies*), Level 4 (*high school diploma*), Level 5 (*trades training or no high school diploma*), and *Students*.

Data were collected from the professional research and sampling firm CINT and our in-house testing platform, Testing Center (EN). Participants from the TC received a report, and those from CINT received monetary compensation for their participation. To ensure data quality, participants whose answers were deemed unreliable or who completed the questionnaire in less than four minutes were excluded from analyses.

Tables 1–4 provide an overview of the sample descriptives.

Table 1. *Gender*

Gender	%	N
Female	48.67	146
Male	50.00	150
N/A	1.33	4

Table 2. *Age groups*

Age group	%	N
18-25	20.67	62
26-33	18.67	56

34-41	24.00	72
42-49	11.33	34
50+	24.00	72
N/A	1.33	4

Table 3. *Study level*

Study level	%	N
Level 1	27.00	81
Level 2	25.67	77
Level 3	12.00	36
Level 4	21.67	65
Level 5	4.33	13
Student	2.67	8
N/A	6.67	20

Table 4. *Country of origin*

Country	%	N
Congo, Democratic Republic	0.33	1
Costa Rica	0.33	1
Egypt	0.33	1
France	0.33	1
Ghana	0.33	1
Hungary	0.33	1
India	29.33	88
Italy	0.33	1
Kenya	0.67	2
Mali	0.33	1
Mauritius	0.33	1

Norway	0.33	1
Saudi Arabia	5.67	17
Spain	0.33	1
Sweden	0.33	1
United Arab Emirates	0.33	1
United Kingdom	19.00	57
Uruguay	0.33	1
USA	39.00	117
Virgin Islands (UK)	0.33	1
N/A	1.33	4

4.2. Norms

The normative population refers to the larger group that the assessment is intended to represent. It is defined by collecting data from a sufficiently large number of candidates who complete the assessment. From these data, we calculate descriptive statistics that are generalized to reflect what would be expected in the broader population. In this way, the normative population serves as the statistical reference point for interpreting individual scores.

For each RIASEC factor, the results are summarized through descriptive norms. These include the mean (average score), standard deviation (degree of variability), minimum (lowest observed score), maximum (highest observed score), range (spread of scores from minimum to maximum), and the standard error (precision of the mean estimate). Together, these indices provide an empirical context for interpretation. For example, if the mean score for a factor is 4.5, it shows that people in the normative population typically score around 4.5 on that factor. In this way, descriptive norms establish the statistical framework in which individual results can be located and evaluated relative to the broader population distribution.

Tables 5–8 detail the mean (*M*), standard deviation (*SD*), minimum (*Min*), maximum (*Max*), range, and standard error (*SE*) scores on factors of **ORIENTATION**, for the whole sample, for the different genders, age groups, and study levels.

Table 5. Factor score norms for the whole sample

Factor	Mean	SD	Min	Max	Range	SE
Whole sample norms (N = 300)						
R	4.85	1.19	1.00	8.00	7.00	0.07
I	5.29	1.19	1.80	8.80	7.00	0.07
A	4.30	1.26	1.30	8.30	7.00	0.07
S	5.33	1.34	1.00	9.00	8.00	0.08
E	4.79	1.11	2.30	8.80	6.50	0.06
C	5.59	1.11	2.50	8.80	6.30	0.06

Table 6. Factor score norms by gender

Factor	Mean	SD	Min	Max	Range	SE
Female (n = 146)						
R	4.81	1.20	1.00	8.00	7.00	0.10
I	5.36	1.15	3.00	8.80	5.80	0.10
A	4.41	1.27	1.30	8.30	7.00	0.11
S	5.38	1.35	2.30	9.00	6.70	0.11
E	4.45	1.00	2.30	7.30	5.00	0.08
C	5.74	1.15	2.80	8.50	5.70	0.10
Male (n = 150)						
R	4.91	1.18	2.00	7.50	5.50	0.10
I	5.22	1.24	1.80	8.50	6.70	0.10
A	4.22	1.22	1.30	8.00	6.70	0.10
S	5.26	1.32	1.00	8.80	7.80	0.11
E	5.10	1.12	2.80	8.80	6.00	0.09
C	5.45	1.06	2.50	8.80	6.30	0.09

Table 7. Factor score norms by age group

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

18-25 (n = 62)

R	4.82	1.10	1.50	7.50	6.00	0.14
I	5.15	1.10	1.80	7.30	5.50	0.14
A	4.45	1.44	1.30	8.30	7.00	0.18
S	5.44	1.40	3.00	9.00	6.00	0.18
E	4.87	1.27	2.50	8.50	6.00	0.16
C	5.40	0.92	3.50	8.00	4.50	0.12

26-33 (n = 56)

R	4.68	1.11	1.50	6.80	5.30	0.15
I	5.22	1.13	3.00	8.50	5.50	0.15
A	4.51	1.09	1.30	6.50	5.20	0.15
S	5.51	1.30	2.80	8.50	5.70	0.17
E	4.90	1.02	2.30	7.30	5.00	0.14
C	5.33	0.95	3.30	8.00	4.70	0.13

34-41 (n = 72)

R	5.00	1.26	2.00	8.00	6.00	0.15
I	5.24	1.19	2.00	7.80	5.80	0.14
A	4.42	1.26	1.50	8.00	6.50	0.15
S	5.21	1.41	2.30	8.50	6.20	0.17
E	4.74	1.05	2.80	8.50	5.70	0.12
C	5.55	1.10	2.80	7.80	5.00	0.13

42-49 (n = 34)

R	4.91	1.11	2.80	7.00	4.20	0.19
I	5.28	1.09	3.80	8.00	4.20	0.19
A	4.45	0.97	1.80	6.50	4.70	0.17
S	5.26	1.12	2.80	7.50	4.70	0.19
E	4.70	1.06	2.50	6.80	4.30	0.18
C	5.56	1.06	3.50	7.30	3.80	0.18

50+ (n = 72)

R	4.87	1.29	1.00	7.50	6.50	0.15
---	------	------	------	------	------	------

I	5.52	1.37	2.00	8.80	6.80	0.16
A	3.88	1.21	1.50	7.30	5.80	0.14
S	5.20	1.32	1.00	7.50	6.50	0.16
E	4.67	1.12	2.50	8.80	6.30	0.13
C	6.02	1.30	2.50	8.80	6.30	0.15

Table 8. *Factor score norms by study level*

Factor	Mean	SD	Min	Max	Range	SE
Level 1 (n = 81)						
R	4.69	1.27	2.00	8.00	6.00	0.14
I	5.37	1.10	2.80	8.30	5.50	0.12
A	4.36	1.18	1.80	8.00	6.20	0.13
S	5.39	1.22	2.30	8.00	5.70	0.14
E	4.78	1.05	2.50	8.50	6.00	0.12
C	5.55	0.99	3.50	8.30	4.80	0.11
Level 2 (n = 77)						
R	4.88	1.13	1.50	7.50	6.00	0.13
I	5.26	1.27	1.80	8.50	6.70	0.14
A	4.29	1.27	1.30	7.80	6.50	0.14
S	5.33	1.39	2.30	8.80	6.50	0.16
E	4.92	1.20	2.30	7.50	5.20	0.14
C	5.48	1.07	3.50	8.00	4.50	0.12
Level 3 (n = 36)						
R	4.87	1.02	2.80	7.30	4.50	0.17
I	5.43	1.28	2.00	8.00	6.00	0.21
A	4.33	1.41	1.50	8.30	6.80	0.23
S	5.10	1.46	2.50	8.50	6.00	0.24
E	4.68	1.25	2.50	8.80	6.30	0.21
C	5.75	1.52	2.50	8.80	6.30	0.25

Level 4 (n = 65)

R	5.00	1.20	1.50	7.30	5.80	0.15
I	5.23	1.24	2.50	7.80	5.30	0.15
A	4.31	1.16	1.50	7.50	6.00	0.14
S	5.26	1.25	2.80	7.50	4.70	0.15
E	4.64	1.03	2.50	7.30	4.80	0.13
C	5.71	1.09	2.80	7.80	5.00	0.14

Level 5 (n = 13)

R	5.11	0.83	3.50	6.30	2.80	0.23
I	5.00	0.99	4.00	7.30	3.30	0.27
A	4.58	1.38	2.80	7.00	4.20	0.38
S	5.19	1.41	1.00	7.00	6.00	0.39
E	4.55	0.92	2.80	6.00	3.20	0.26
C	5.66	0.93	4.00	7.00	3.00	0.26

Student norms (n = 8)

R	5.49	1.10	3.50	6.80	3.30	0.39
I	4.71	1.18	3.80	7.30	3.50	0.42
A	4.40	1.75	1.30	6.30	5.00	0.62
S	5.44	1.45	4.00	8.30	4.30	0.51
E	5.31	1.32	4.30	8.50	4.20	0.47
C	4.81	0.79	3.80	6.30	2.50	0.28

4.3. Reliability

Reliability refers to the extent to which an assessment is free from error and the degree to which items on a scale are measuring the same construct. 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.

In this manual, we therefore report both composite reliability, which reflects the internal consistency of items within each factor, and marginal reliability, which reflects the precision of ability estimates across the range of the trait.

4.3.1. 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's 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.¹³ Internal reliability values vary from 0 to 1, with higher values more desirable. As a rule of thumb, a reliability of 0.70 or higher is recommended, with values greater than 0.90 (excellent); 0.70 to 0.90 (good); 0.60 to 0.70 (acceptable); 0.50 to 0.60 (poor); and values less than 0.50 (unacceptable).¹⁴

The formula for Composite Reliability (CR) is:

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

The composite reliability for **ORIENTATION** was computed on the whole sample ($N = 300$). Table 9 describes the composite reliability of **ORIENTATION**. All factors showed composite reliability above 0.70. In addition, **ORIENTATION** had an overall composite reliability of 0.86, indicating that the scale has very good internal consistency and is highly reliable.

¹³ 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. <https://doi.org/10.1111/bjop.12046>; Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173–184. <https://doi.org/10.1177/01466216970212006>; Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>

¹⁴ Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Table 9. *Composite reliability for factors*

Factor	CR
R	0.94
I	0.86
A	0.91
S	0.89
E	0.90
C	0.78

4.3.2. *Marginal reliability*

Marginal reliability is a measure of the precision of ability (θ) estimates in an Item Response Theory (IRT) framework. Unlike traditional internal consistency indices that focus on the correlations among items, marginal reliability evaluates the extent to which the latent trait estimates are free from error across the population of test takers. It is mathematically defined as the ratio of the variance of the true scores to the total variance of observed scores, incorporating both true score variance and measurement error.

In practice, marginal reliability is derived from the distribution of ability estimates and their associated standard errors, providing an index of the average information provided by the test. This method is particularly appropriate for IRT-based assessments, as it accounts for measurement precision across different trait levels, rather than assuming equal measurement error across the scale.¹⁵ Marginal reliability values vary from 0 to 1, with higher values being more desirable. Following the same rule of thumb as for other reliability indices, values of 0.70 or higher are recommended, with values greater than 0.90 (excellent); 0.70 to 0.90 (good); 0.60 to 0.70 (acceptable); 0.50 to 0.60 (poor); and values less than 0.50 (unacceptable).¹⁶ The formula for Marginal Reliability (MR) is:

¹⁵ Thissen, D., & Wainer, H. (2001). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates; Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21(4), 347–360. <https://doi.org/10.1111/j.1745-3984.1984.tb01039.x>

¹⁶ Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

$$MR = \frac{Var(\hat{\theta})}{Var(\hat{\theta}) + Mean(SE^2)}$$

where $Var(\hat{\theta})$ is the variance of the estimated ability distribution, and $Mean(SE^2)$ is the mean squared standard error of the estimates. The marginal reliability for **ORIENTATION** was computed on the whole sample ($N = 300$). Table 10 describes the marginal reliability of **ORIENTATION**. All factors showed marginal reliabilities above 0.70, with values ranging between 0.78 and 0.83.

Table 10. *Marginal reliability for factors*

Factor	MR
R	0.83
I	0.82
A	0.83
S	0.83
E	0.82
C	0.78

4.4. Validity

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

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.¹⁷ 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. **ORIENTATION** 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.¹⁸ This is an important aspect to take into account in order to obtain the candidate's cooperation and commitment. Face validity and attitudes towards tests influence candidates' motivation, concentration and test-taking anxiety.¹⁹ As a result, the question of "Does **ORIENTATION** 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 **ORIENTATION**.

4.4.1. 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.²⁰ 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) developed with insufficient research. Therefore, in order to develop a scale that

¹⁷ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

¹⁸ Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287–293. <https://doi.org/10.1111/j.1745-3984.1985.tb01065.x>

¹⁹ McCarthy, J. M., & Goffin, R. D. (2003). Improving the validity of letters of recommendation: An investigation of three standardization methods. *Canadian Journal of Behavioural Science*, 35(3), 184–197. <https://doi.org/10.1037/h0087199>

²⁰ Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>

measures occupational and vocational preferences, the construct must reflect known theories that define interests, identify the components that it is composed of, and pinpoint 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 **ORIENTATION** define and classify occupational and vocational preferences in a way that represents theoretical and practical understanding?” For **ORIENTATION**, construct validity was established using two approaches: inter-factor correlations and confirmatory factor analysis, as detailed in the sections below.

4.4.2. Inter-factor Correlations

The factors in **ORIENTATION** 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 11 shows the inter-factor correlations.

Table 11. Inter-factor Correlations.

Factor	R	I	A	S	E	C
R	1.00					
I	-0.12*	1.00				
A	0.02	-0.16**	1.00			
S	-0.31****	-0.28****	-0.44****	1.00		
E	-0.29****	-0.24****	-0.22***	-0.03	1.00	
C	-0.30****	-0.19***	-0.24****	-0.04	-0.15**	1.00

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

The inter-factor correlations in Table 11 show both consistencies with and departures from the patterns predicted by Holland's RIASEC model. As expected, some opposite domains displayed negative associations, such as R with S ($r = -0.31$) and A with S ($r = -0.44$).

At the same time, several adjacent pairs, such as R with C ($r = -0.30$) and I with A ($r = -0.16$), showed weaker or negative correlations where positive ones might have been expected. Similarly, S and E, which are adjacent in the hexagon, were only weakly related ($r = -0.03$).

Overall, these results suggest that while the general structure of the RIASEC model is reflected in some correlations, the boundaries between certain interest domains may be less rigid in practice. This may reflect the way modern career paths often combine elements from multiple domains, blurring distinctions that appear sharper in the theoretical model.²¹

4.4.3. Confirmatory factor analysis

A confirmatory factor analysis (CFA) with continuous indicators was conducted to test whether the observed data fit the theoretical assessment model. CFA, a hypothesis-driven form of structural equation modelling (SEM), specifies the expected number of factors and examines how well the indicators represent the latent constructs using goodness-of-fit indices.²²

Model fit was evaluated using multiple indices to capture different aspects of fit. The chi-square (χ^2) test provides a measure of absolute fit, indicating how well the proposed model reproduces the observed covariance matrix; however, because χ^2 is sensitive to sample size and violations of distributional assumptions, relative fit indices are also recommended.²³

For the validation of *ORIENTATION*, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker–Lewis Index (TLI) were selected, as they are least affected by sample size.²⁴ Using a combination of these indices ensures a more comprehensive evaluation of model fit, as each reflects a different dimension of model accuracy.²⁵ The RMSEA, an absolute fit index, assesses residual error, with values of 0.05 or below indicating a good fit. The CFI and TLI are incremental fit indices that compare the tested model with an independence model in

²¹ Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science*, 23(2), 98–103. <https://doi.org/10.1177/0963721414522812>

²² 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. <https://doi.org/10.3200/JOER.99.6.323-338>

²³ Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

²⁴ Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA; Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on SEM fit indices. *Structural Equation Modeling*, 6(1), 56–83. <https://doi.org/10.1080/10705519909540119>

²⁵ Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60.

which observed variables are uncorrelated; values of 0.90 and above are generally considered acceptable to excellent. However, a CFI or TLI above 0.90 does not always correspond to an RMSEA of 0.05 or lower.²⁶

4.4.3.a. Thurstonian IRT

ORIENTATION is based on an ipsative forced-choice structure, which violates many of the assumptions of traditional CFA based on Likert-style responses. To address this, the Thurstonian Item Response Theory (TIRT) model is used.²⁷

TIRT is specifically designed for forced-choice assessments, where respondents choose between options rather than rate them independently. It accounts for the comparative nature of responses, dependencies between item triads, and the latent structure underlying choice data, making it more appropriate and psychometrically robust for this type of assessment. TIRT integrates with the CFA framework by modelling the latent traits that govern response behavior across pairwise items. Instead of treating each item as independent, TIRT models each triad of items (blocks) as a function of underlying trait differences. This ensures more accurate parameter estimation (e.g., factor loadings, trait estimates) as well as allows us to assess model fit using conventional SEM indices such as RMSEA, CFI, and TLI.

In order to analyse the data using TIRT modelling, the raw responses were converted into binary outcomes following the procedure described by Brown and Maydeu-Olivares for item response modelling of forced-choice formats.²⁸ In **ORIENTATION**, each item consists of a triad of statements from which respondents select the option that is most like them and the one that is least like them. Each triad therefore implies three pairwise comparisons (a vs. b, a vs. c, and b vs. c), which are used to generate binary outcomes representing the relative preference between statements. These binary outcomes serve as input for the TIRT model.

The models were estimated in *R* using the *thurstonianIRT* package in conjunction with *lavaan*.²⁹ Estimation employed the unweighted least squares mean and variance adjusted (ULSMV) algorithm, which is appropriate for binary forced-choice data and provides robust model fit indices. Table 12 shows the goodness of fit indices and model results.

²⁶ Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

²⁷ The Thurstonian Item Response Theory (TIRT) model extends Thurstone's (1927, 1931) comparative judgment framework to modern item response modelling for forced-choice assessments. See Brown, A., & Maydeu-Olivares, A. (2011). *Item response modeling of forced-choice questionnaires*. *Educational and Psychological Measurement*, 71(3), 460–502. <https://doi.org/10.1177/0013164410375112>

²⁸ Brown, A., & Maydeu-Olivares, A. (2018). *Modeling forced-choice response formats: A comparison of Thurstonian and item response theory approaches*. *Psychological Methods*, 23(4), 764–783. <https://doi.org/10.1037/met0000155>

²⁹ Thissen, D., & Brown, A. (2019). *thurstonianIRT: Thurstonian IRT models for forced-choice questionnaires* [R package]. <https://cran.r-project.org/package=thurstonianIRT>; Rosseel, Y. (2012). *lavaan: An R package for structural equation modeling*. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>

Table 12. *Confirmatory factor analysis using thurstonian IRT*

N	χ^2	df	χ^2/df	p	RMSEA	p ≤ 0.05	CFI	TLI
300	8004.70	6925	1.16	0.00	0.023	1.00	0.671	0.661

Table 12 shows the model fit results for the **ORIENTATION** factor model. Given the large sample size, the chi-square test is not a reliable indicator of fit. Instead, greater weight is placed on the RMSEA, which is less sensitive to sample size and provides a robust measure of absolute fit. The RMSEA value of 0.023 indicates an excellent fit to the data (values ≤ 0.05 are considered good). Although the CFI and TLI values were below conventional thresholds, RMSEA is considered the most appropriate index in this context. Taken together, the results support the validity of the **ORIENTATION** factor model.

5. ORIENTATION 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. **ORIENTATION** 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)

5.1. 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.³⁰ 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 whether 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.

5.1.1. 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.³¹ 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 but a disadvantage at the lower end. 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 is used to calculate the probability of endorsing an item. Figures 5 and 6 show examples of item characteristic curves for uniform and non-uniform DIF.

³⁰ Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72(1), 19–29. <https://doi.org/10.1037/0021-9010.72.1.19>; McLeary, R. (2014). Measuring social mechanisms. *Annual Review of Sociology*, 40, 39–59. <https://doi.org/10.1146/annurev-soc-071913-043137>

³¹ 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. <https://doi.org/10.1111/j.1745-3992.1998.tb00619.x>; Hanson, B. A. (1998). Uniform and nonuniform differential item functioning. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 135–156). Mahwah, NJ: Lawrence Erlbaum Associates.

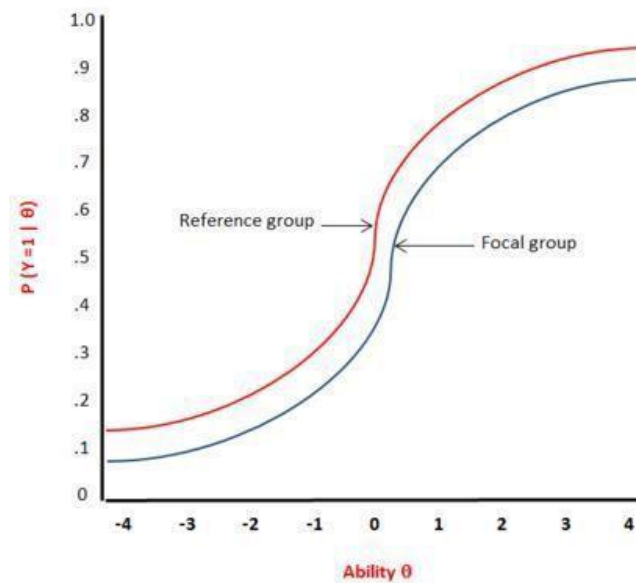


Figure 5. Item characteristic curve with uniform DIF

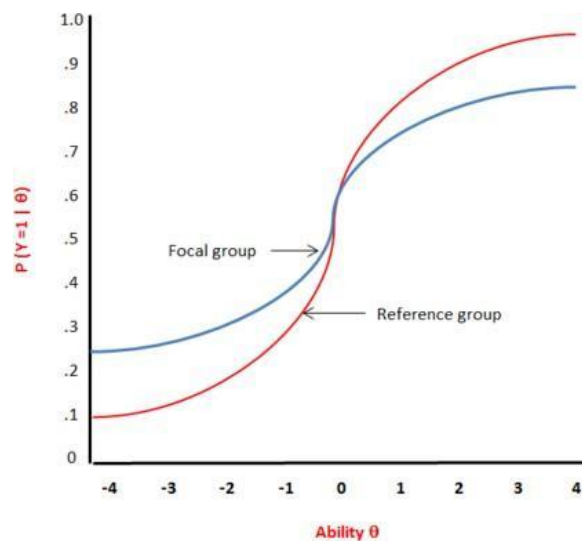


Figure 6. Item characteristic curve with non-uniform DIF

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).

5.1.1.a. DIF Analysis: Gender

DIF analysis was performed across gender in order to ensure that there was no bias toward either males or females, and therefore that **ORIENTATION** is a fair test for both genders. For this analysis, the norm sample was divided according to gender: 146 females and 150 males. The analysis identified 0 out of the 120 items with a DIF effect for gender. According to Zumbo and Thomas' (1997) classification of effect size, any R^2 under 0.13 is considered negligible.³² Out of the 120 items, 0 showed an effect size of 0.13 or greater; therefore, it can be considered that **ORIENTATION** does not contain any item that functions differently across genders.

5.1.1.b. DIF Analysis: Age

DIF analysis was performed based on age in order to ensure that **ORIENTATION** was a fair test for people of all ages. For this analysis, two age groups were defined: 141 participants aged 18–35 ($M = 27.13$, $SD = 4.85$) and 155 participants aged 36 or over ($M = 51.70$, $SD = 12.65$).

Table 13. Coefficient and effect size for DIF items

Item	χ^2	ΔR^2	Interpretation
A3	9.868	0.024	Negligible
A26	10.118	0.025	Negligible
A36	11.644	0.032	Negligible
S3	17.807	0.046	Negligible
S38	11.435	0.029	Negligible

The analysis identified five out of the 120 items with a DIF effect for age. Table 13 shows the χ^2 statistic as well as the ΔR^2 effect size. None of the five items flagged for DIF showed an effect size of 0.13 or greater; therefore, it can be considered that **ORIENTATION** does not contain any item that functions differently across age groups.

³² Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying differential item functioning (DIF)*. Prince George, Canada: University of Northern British Columbia.

5.2. T-tests and ANOVAs

The DIF analysis demonstrates that **ORIENTATION** has high consequential validity. Although a few items were flagged with DIF, there was a negligible impact on the overall assessment, meaning that the assessment does not bias toward 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 across genders, age groups, and study levels (see Tables 14 –18).

5.2.1. Gender

In order to analyse differences in scores between genders, an independent-samples *t* test was conducted to compare all **ORIENTATION** factors between females ($n = 146$) and males ($n = 150$).

Table 14. Gender *t* tests for factor scores

Factor	Female	Male	df	t	p
R	4.81	4.91	293.54	-0.73	0.47
I	5.36	5.22	293.30	1.01	0.31
A	4.41	4.22	292.45	1.32	0.19
S	5.38	5.26	293.21	0.76	0.45
E	4.45	5.10	291.77	-5.25	0***
C	5.74	5.45	290.28	2.29	0.02*

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

The *t* tests found significant differences between males and females on the factors of E and C.

Males were significantly more interested in E ($M = 5.10$) than females ($M = 4.45$), $t(291.77) = -5.25$, $p < 0.001$. This could be due to the societal norms that men are encouraged to take charge, compete and lead. Jobs in the Enterprising category, such as sales and entrepreneurship, may become of more interest to men than women as it fits the image set by cultural expectations.³³

³³ Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135(6), 859–884. <https://doi.org/10.1037/a0017364>

Females were significantly more interested in C ($M = 5.74$) than males ($M = 5.45$), $t(290.28) = 2.29$, $p = 0.02$. This could also be due to societal norms where jobs that involve organisation and people-focused aspects, such as office or administrative work, have been associated with the attention to detail and reliability qualities that women are often encouraged to develop.³⁴

5.2.2. Age Groups by Factors

In order to analyse differences in **ORIENTATION** factor scores between age groups, participants were separated into five groups according to their age: 18–25 ($n = 62$), 26–33 ($n = 56$), 34–41 ($n = 72$), 42–49 ($n = 34$), and 50 and above ($n = 72$).

Table 15. ANOVA for factors by age group

Factor	18-25	26-33	34-41	42-49	50+	df	F	p
R	4.82	4.68	5	4.91	4.87	4	0.59	0.67
I	5.15	5.22	5.24	5.28	5.52	4	0.93	0.45
A	4.45	4.51	4.42	4.45	3.88	4	3.04	0.02*
S	5.44	5.51	5.21	5.26	5.2	4	0.71	0.59
E	4.87	4.9	4.74	4.7	4.67	4	0.5	0.73
C	5.4	5.33	5.55	5.56	6.02	4	4.12	0***

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

Table 16. Post hoc t test for Artistic by age group

Artistic				
Age group	18-25	26-33	34-41	42-49
26-33	1.00	-	-	-
34-41	1.00	1.00	-	-
42-49	1.00	1.00	1.00	-
50+	0.077	0.041*	0.090	0.275

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

³⁴ Lippa, R. A. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology*, 74(4), 996–1009. <https://doi.org/10.1037/0022-3514.74.4.996>

Table 17. *Post hoc t test for Conventional by age group*

Conventional				
Age group	18-25	26-33	34-41	42-49
26-33	1.00	-	-	-
34-41	1.00	1.00	-	-
42-49	1.00	1.00	1.00	-
50+	0.011*	0.004**	0.097	0.416

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

The ANOVA for the RIASEC factors by age groups found significant differences for A and C. Post hoc analysis was conducted using pairwise t tests with Bonferroni correction to determine where these differences occurred.

Post hoc t tests for factor A revealed a significant difference between the 26–33 age group ($M = 4.51$) and the 50 and above age group ($M = 3.88$), $p = 0.041$. In addition, the 18–25 age group ($M = 4.45$) scored higher than the 50 and above group ($M = 3.88$), with the difference approaching significance, $p = 0.077$. This suggests that younger people may be more interested in artistic endeavours. One possible explanation is that artistic fields are becoming increasingly popular and accessible as careers. Creative professions such as design and illustration are more desirable than ever, with organisations recognising their value. As a result, young people may be more exposed to these opportunities and show greater interest in pursuing them.³⁵

Post hoc t tests revealed significant differences between the 18–25 age group ($M = 5.40$) and the 50 and above age group ($M = 6.02$), $p = 0.011$, as well as between the 26–33 age group ($M = 5.33$) and the 50 and above age group ($M = 6.02$), $p = 0.004$. These results suggest that older individuals tend to show stronger interest in conventional fields compared to younger individuals. This may be because conventional roles, which often involve structured tasks, stability, and clear career progression, are especially valued among those with more career experience. As people age, the appeal of predictability and security in the workplace may increase, whereas younger individuals might be less focused on tradition and more attracted to flexible or unconventional career paths.

³⁵ Hoff, K. A., Briley, D. A., Wee, C. J. M., & Rounds, J. (2018). *Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies*. *Psychological Bulletin*, 144(4), 426–451. <https://doi.org/10.1037/bul0000140>

Consequently, older groups are more likely to express interest in conventional environments and practices.³⁶

5.2.3. Study Level Groups by Factors

In order to analyse differences in **ORIENTATION** factor scores between study levels, participants were separated into six groups according to their study level: Level 1 ($n = 81$), Level 2 ($n = 77$), Level 3 ($n = 36$), Level 4 ($n = 65$), Level 5 ($n = 13$), and Student ($n = 8$).

Table 18. ANOVA for factors by study level

Factor	Level 1	Level 2	Level 3	Level 4	Level 5	Student	df	F	P
R	4.69	4.88	4.87	5.00	5.11	5.49	5	1.10	0.36
I	5.37	5.26	5.43	5.23	5.00	4.71	5	0.74	0.60
A	4.36	4.29	4.33	4.31	4.58	4.4	5	0.13	0.98
S	5.39	5.33	5.10	5.26	5.19	5.44	5	0.29	0.92
E	4.78	4.92	4.68	4.64	4.55	5.31	5	0.96	0.44
C	5.55	5.48	5.75	5.71	5.66	4.81	5	1.26	0.28

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

The ANOVA for the RIASEC factors by study level found no significant differences for any of the factors.

³⁶ Leuty, M. E., & Hansen, J.-I. C. (2014). *Teasing apart the relations between age, birth cohort, and vocational interests*. *Journal of Counseling Psychology*, 61(2), 289–298. <https://doi.org/10.1037/a0035341>

CONTACT US

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

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



www.centraltest.com

© 2025 by Central Test – Empowering Talent.
All rights reserved.