

**The Development and Validation of the  
Trait-Map  
Personality Inventory**



**Technical Manual**

## Abstract

This paper presents an in-depth overview of the Trait-Map Personality Inventory, including its Big Five-compatible personality model consisting of 25 traits, its novel ranking questionnaire format, and its distinctive combinatorial optimization method that mitigates the drawbacks of the forced-choice format. A range of analysis is included to examine the questionnaire's psychometric properties and validity, such as means and response distributions, internal consistency, test-retest reliability, and correlation tables, as well as criterion validity based on various studies.

## Rationale for developing Trait-Map

Although the proper application of personality assessments in the workplace is challenging, competent practitioners value the rich data personality assessments can provide for more effective recruitment and development. Their value has been also confirmed by scientific research. In *The Handbook of Industrial and Organizational Psychology* (Hough & Johnson, 2013), the authors review decades of research examining the relationship between personality variables and work performance. They find significant effect sizes and list a wealth of empirical evidence demonstrating the importance of personality assessments. They also advocate for nuanced thinking and careful implementation, and warn about the risk of intentional distortion. Reinforcing this concern, a large-scale meta-analysis (Loy et al., 2025) compared the validity of personality tests across low-stakes settings, like employee assessments, and high-stakes settings, such as applicant testing. The results consistently showed that test validity was higher in low-stakes environments. To improve selection accuracy in light of this, the authors recommend that organizations use faking-resistant assessments, for instance, those with forced-choice formats, and implement complementing measures.

Trait-Map aims to offer such a faking-resistant, forced-choice personality assessment solution based on the well-established Big Five standard, employing an innovative questionnaire format. The assessment evaluates 25 distinct personality traits using a format in which candidates rank five items at a time. This unique methodology makes the assessment process faster and more accurate, and enables an organic integration between the questionnaire and the underlying personality model.

Trait-Map brings a fresh perspective to the debate over Likert-type scales versus forced-choice questionnaire formats. Psychometric assessments typically use Likert-type scales, especially in academic research. Likert-type scales are prone to a range of distortions such as acquiescence bias (some people tend to agree with statements regardless of their content); extremity/midpoint bias (some people prefer to select responses from the extremes, while others from the midpoint); leniency/severity tendencies (some people are too lenient, while others are too strict in self-evaluation); faking attempts (Paulhus & Vazire, 2007; Wetzel et al., 2016); reference group effects (Credé et al., 2010). The use of forced-choice questionnaire formats eliminates these effects (Schulte et al., 2021) and can significantly reduce faking, as the Cao and Drasgow (2019) meta-study authoritatively concludes.

However, forced-choice questionnaires also have inherent disadvantages: reduced internal consistency, as indicated by Cronbach alpha statistics; longer questionnaires; increased negative correlations between scales of the construct or external criteria; and the so-called "ipsative data" phenomenon, i.e., equal total scores, produced by forced-choice questionnaires (the total score is the same for each participant when the raw scores of all scales are summed), which makes the data less suitable for statistical techniques used in academic psychology.

Some forced-choice assessment providers have employed an IRT (Item Response Theory) based scoring system to overcome the ipsative phenomenon. Using this method indeed produces normative results from forced-choice questionnaires, but introduces other issues and achieves mixed results overall (Schulte et al., 2021). The authors of the Trait-Map used a different approach, one that is specific to personality testing. Forced-choice questionnaires comparable to a store where all shoppers have to spend exactly the same amount of money. What they buy represent their individual differences, but there is no difference in the total amount of money spent (the budget is fixed, i.e. "ipsative"). Contrary to academic considerations, the ipsative results (having the same total score when all scales are summed) of Trait-Map can be seen as a suitable model for personality: the authors of Trait-Map consider personality as something that everyone has to the same extent (unlike IQ, EQ knowledge and job skills): there are individual differences in the degree of possessing specific traits, but the sum of scores on all possible traits, the total "amount of personality", is the same for everyone. Therefore, the forced-choice format can be an appropriate method for measuring personality if the included traits offer a comprehensive and balanced representation of human personality. The authors of Trait-Map have embraced the forced-choice questionnaire format and developed innovative measures to minimize its disadvantages. One such measure is the creation of an efficient 5-item/block ranking questionnaire format. Another, less obvious measure is a proprietary combinatorial optimization method used to group the items into the blocks of five. As a result, Trait-Map offers a rare combination of a multi-scale inventory and an efficient forced-choice questionnaire format. It takes only 15-20 minutes to complete Trait-Map and generate a high-resolution personality profile consisting of 25 personality traits (**median test time is 17 minutes**, n=25790). The questionnaire also provides a good experience on the small screens of smart phones.

Another objective in developing Trait-Map was to create an intuitive personality model that is sufficiently detailed yet relatively easy to understand and use. Trait-Map is based on trait theory and is strongly influenced by the "Big Five" personality dimensions, which are widely accepted today as the fundamental structure of human personality traits. The best known Big Five questionnaires are the Revised NEO Personality Inventory (NEO PI-R) and the IPIP-NEO, both of which have a total of 30 traits, 6 traits in each of the Big Five dimensions. Some forms of these questionnaires are used in most academic studies, but they are less popular in work settings because their items are not geared towards work-related item content. The authors of Trait-Map took the IPIP-NEO questionnaire and substantially revised it, replacing the personal content with work-related content to reflect the world of work and contemporary business language. This process also meant trading some factor structure purity for content validity. The Trait-Map model has **a total of 25 traits, 5 in each of the Big Five dimensions**; it is a job-oriented blend of 23 classic traits and 2 innovative traits (Health Management and Observation) to provide rich data for both selection and learning applications, and to enable a wide range of reports: Job Fit, Team Roles, Conflict Styles, and affinity for various skills related to Collaboration, Presentation, Negotiation, and Sales.

### **The history of Trait-Map**

The development of Trait-Map (previously called FiT In and FiT In 5x5) started in Shanghai in 2002. Lead authors: Friedemann Demmer, Gabor Nagy. Scientific advisors: Cynthia Zhang (MA), Lv Xiaojun (Ph.D.), Guo Xiaowei (Ph.D.), Maria Lin (MA). The starting point for the development work was the IPIP-NEO test and the International Personality Item Pool created by Lewis R. Goldberg, a senior scientist at the Oregon Research Institute, and the scoring algorithms of the IPIP-NEO provided by Prof. John A. Johnson. The team modified the IPIP-NEO trait names, definitions and items to better suit the world of work, reduced the number of traits from 30 to 25, and developed an innovative forced choice questionnaire format. The team developed the English and Chinese versions simultaneously, making modifications to

the items and model based on the data collected using item analysis and correlation analysis. The psychometric quality of Trait-Map was certified by the Psychometric Committee of the Shanghai Psychological Association in 2006 and has been continuously monitored and improved over the years.

Besides its intercultural roots, another notable feature of Trait-Map is the corporate environment in which it was developed. It was not developed by using student samples. Instead, the developing authors themselves used it in corporate training and high-stakes hiring and assessment center projects, refining it based on feedback and data gained from workplace application.

Another consideration in developing the Trait-Map questionnaire was to use plain language and common workplace expressions, making it easy to understand and translate. For instance, it does not contain metaphors commonly used by other instruments, such as "I am the life of the party" or "I seldom toot my own horn." High face and content validity were accomplished through feedback from subject matter experts and multiple revisions.

## The Trait-Map Personality Model

Trait-Map is a workplace application of the Big Five model: traits and associated trait descriptions are closely related to work styles and job-fit. You can find a summary table of the Trait-Map personality model below.

The Trait Names and Score Descriptions below are somewhat simplified “pointers”. The traits are explained in full in the User Manual.

	Trait Name	Left Score Pointer	Definition	Right Score Pointer
Emotional Stability	<b>Emotional Control</b>	Expressive	Tendency to avoid outbursts of anger or frustration and maintain emotional restraint	Controlled
	<b>Calmness</b>	Tense	Staying calm under pressure, not overly worrying	Calm
	<b>Health Mgmt</b>	Self-exhausting	Maintaining work-life balance and a healthy lifestyle	Self-protecting
	<b>Optimism</b>	Pessimistic	Positive outlook, tendency to experience positive moods and high spirits	Cheerful
	<b>Confidence</b>	Timid	Freedom from social anxiety, acting confidently in social and ambiguous situations	Confident
Extroversion	<b>Assertiveness</b>	Passive	Acting decisively and taking control of situations	Assertive
	<b>Persuasiveness</b>	Unconvincing	Inclination to talk, debate and being verbally convincing	Persuasive
	<b>Liveliness</b>	Reserved	The degree of wit and vividness in social interactions	Engaging
	<b>Networking</b>	Solitary	The need, search, and enjoyment of the company of others	Contact seeking
	<b>Work Tempo</b>	Measured	A person's speed and energy in work tasks	Fast
Agreeableness	<b>Altruism</b>	Self-focused	Tendency to assist those in need without thinking of own benefits	Altruistic
	<b>Cooperation</b>	Uncompromising	Willingness to sacrifice own interests for the sake of harmony or group unity	Cooperative
	<b>Criticalness</b>	Receptive	The tendency of finding and pointing out faults and judging strictly	Critical
	<b>Individualism</b>	Listening to advice	Tendency to form firm views and stick to them	Unyielding
	<b>Competitiveness</b>	Gentle	The inclination to engage in competition	Competitive
Focus	<b>Drive</b>	Satisfied	Level of ambition and the degree to which one strives towards success	Driven
	<b>Methodicalness</b>	Casual	Being organized and work according to plans	Systematic
	<b>Detail-concern</b>	Overview-oriented	Tendency to focus on details and work precisely	Painstaking
	<b>Dutifulness</b>	Laid-back	Sense of obligation; tendency to persist at tasks until they are completed	Diligent
	<b>Deliberation</b>	Spontaneous	Thinking through possibilities before acting	Deliberate
Openness	<b>Progressiveness</b>	Unadventurous	The attitude towards novelty and change	Change-oriented
	<b>Creativity</b>	Regular	The frequency of generating new and original ideas	Innovative
	<b>Abstractedness</b>	Concrete	Affinity for analysis, working with data, and working with abstract ideas	Abstract
	<b>Aesthetic Sense</b>	Common	Tendency to notice and be inspired by the aesthetic qualities of things	Inspired
	<b>Observation</b>	Detached	Affinity for detecting and decoding nonverbal messages	Sensitive

## **The Trait-Map questionnaire**

Trait-Map has a unique forced-choice questionnaire format which requires the ranking of five items in each block. The questionnaire consists of 25 blocks containing 5 items each, making it 125 items in total. Median test length is 17 minutes (n=25790), and most people (75%) complete it under 25 minutes. It means people can finish it before boredom or fatigue would compromise the accuracy too much.

The ranking format of the Trait-Map questionnaire works this way: five items need to be sorted in order of "More like me", "Less like me". There are 25 such blocks of 5 items. The questionnaire can be compared to a boxing competition, where every competitor (every trait) fights against every other competitor (against 20 other traits from the other four Big Five dimensions to be precise), and the number of victories (when the trait in questions was ranked "more like me" than the others) gets counted. The number of victories reflects the dominance of the trait compared to the other traits.

The forced-choice format makes it impossible passing it with a perfect score, hiding all weaknesses, therefore it is harder to "fake good". The forced-choice format also enables Trait-Map to assess the deeper, core personality trait structure opposed to the surface behavior which is more situational and also influenced by intention, abilities and self-adjustment. The deeper personality structure that Trait-Map measures manifests in long-term patterns such as job fit, or in situations when we react on autopilot, showing our core traits.

As discussed earlier, the forced-choice format has disadvantages as well. Choosing one item results in not-choosing another one: there is interference between the items. The authors of Trait-Map developed a proprietary mathematical method (a combinatorial optimization method inspired by a problem known as Orthogonal Latin Squares) in the questionnaire design to minimize the distortion caused by such interaction between scales. This combinatorial optimization makes the questionnaire highly balanced: the interference that exists between the traits due to the forced-choice format is minimized and spread evenly among the traits. Because of the forced-choice format, Participants sometimes cannot respond in a fully accurate manner; but due to the optimization, in Trait-Map these errors mostly cancel each other out and don't generate significant distortion.

The forced-choice format also reduces internal scale consistency indicated by the Cronbach's Alpha values. The lower alpha values stem from the nature of the questionnaire, and therefore Cronbach's Alpha values should not be used as a measure of reliability for Trait-Map. Internal scale consistency remains an important development objective, but the Cronbach's Alpha values will always be somewhat lower than those of questionnaires with Likert-type items. Therefore, reliability should be assessed using the test-retest method.

## Response Pattern Validity Indices

Verifying the results with other data sources, such as interviews or assessment center exercises, is highly recommended. At the same time, the Trait-Map report provides indicators that help the assessment facilitator determine the validity of the responses.

### Fastest blocks - Slowest blocks - Median block length

These indicators show the processing time of blocks: how long did it take ranking the five items. This data helps to gain insights into the process of responding to the questionnaire. Blocks are numbered from Q01 to Q25. The focus is not on the content of these blocks, the numbers help to identify if something unusual happened at the beginning, or middle, or at the end of the questionnaire.

Fastest blocks: they can reveal if some blocks have been submitted randomly or in too much hurry. It is not possible to carefully consider the accurate order of items in just a few seconds.

Slowest blocks: they can reveal if the process has been interrupted.

The median block processing time of the entire sample (n=25790) is **41 seconds**. The individual's median block length can be compared against this benchmark.

### Centeredness

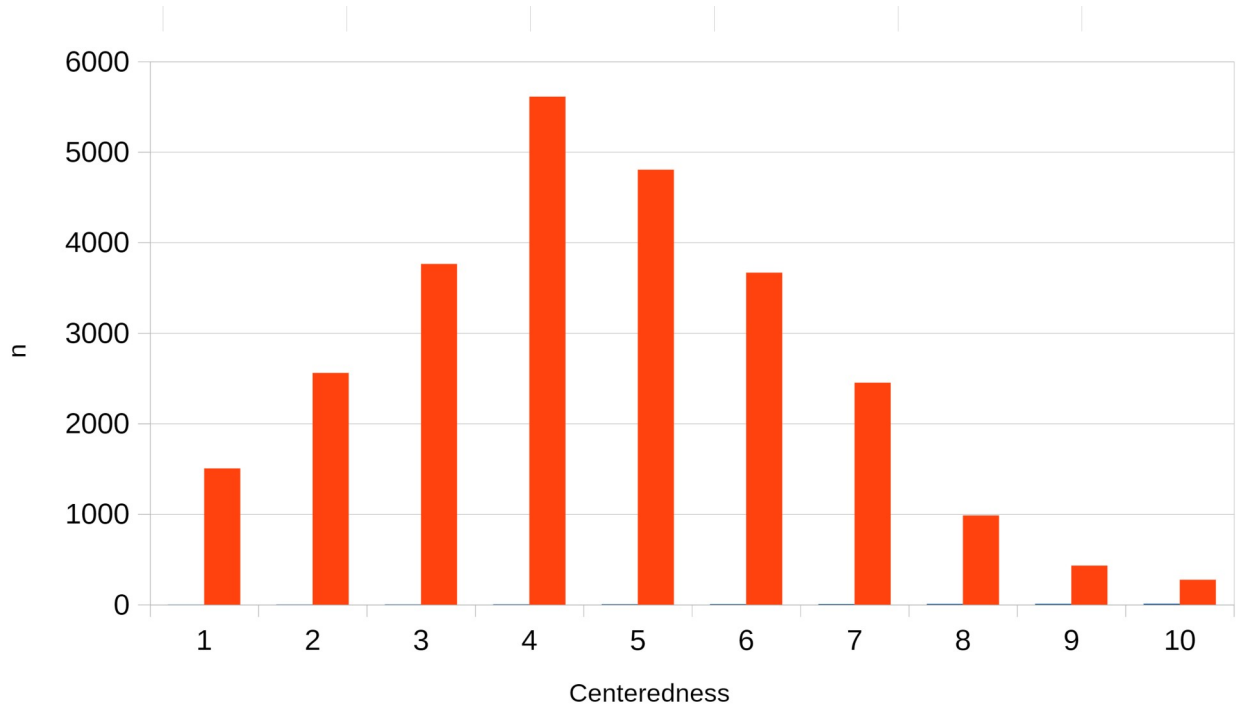
Centeredness is calculated as the inverted variance of the row scores. High Centeredness is caused by low response consistency. The name refers to the "shape" of the Personality Profile: if there are many 5 and 6 standardized scores, and just a few traits with Left or Right scores, then we say the profile is "centered", or Centeredness is high. Centeredness scores between 1 – 7 are normal. Higher Centeredness scores indicate a problem: the information value of the report is low, because the respondent reported few tendencies, dispositions, and inclinations.

High Centeredness can be caused by:

- a) the person is indeed "centered": very much situational, flexible; OR
- b) the person does not know him/herself very well; OR
- c) the questionnaire was done too fast, with too much randomness and without sufficient care and efforts.

Centeredness scores above 5 show too much randomness and insufficient response consistency. The Assessment Facilitator should check with Participants what caused the high score case-by-case. Scores above 7 invalidate the report: **reports with Centeredness scores of 8, 9, and 10 are most likely inaccurate and can be considered invalid.**

The Distribution of standardized Centeredness scores:



Out of 26058 Trait-Map respondents (all responses except responses below 6 minutes), 3.8 % scored 8; 1.7 % scored 9; and 1.1 % scored 10 on Centeredness.

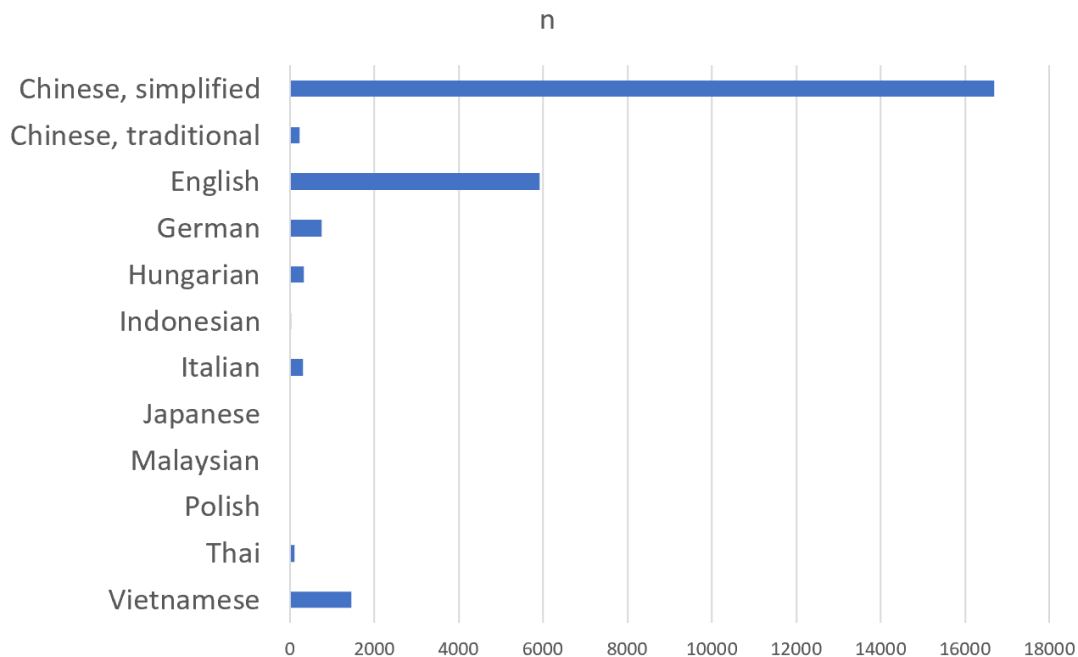
## Descriptive statistics of scales

The exported dataset includes all available Trait-Map responses since 1<sup>st</sup> January 2017 until the 29<sup>th</sup> August 2025 (27221 responses). All data came from professional settings: training programs, recruitment, organizational diagnostics, coaching, and also some client demo and technical testing. The list of user organizations, where the data was obtained from, can be seen in Annex 1.

The first step was data cleaning. The very likely invalid data was deleted from the dataset: responses with completion time below 6 minutes (1163 responses), and responses with Centredness=10, an indicator of lacking response consistency (753 responses). After removing the above described most likely invalid responses (5,3% of the total responses were deleted), in total 25790 responses were left.

Trait-Map is available in the following languages: Chinese, simplified; Chinese, traditional; English; German; Hungarian; Indonesian; Italian; Japanese; Malaysian; Polish; Thai; Vietnamese.

Number of respondents by language version:



This paper focuses on the two largest datasets, the English (n=5918), and the Chinese (n=16702).

### **Cronbach Alpha values, Means, and Standard deviation of traits**

*English data set, n=5918*

The number of items per trait is 5. The range of trait raw scores is 5 – 25. The frequency distribution charts are shown in Annex 2.

#	Traits	SCH		
		Cronbach Alpha	Scale Means	Standard Deviation
1	Emotional Control	0.660	13.9	4.2
2	Calmness	0.530	15.6	3.8
3	Health Mgmt	0.658	13.1	4.5
4	Optimism	0.586	16.0	4.0
5	Confidence	0.720	14.3	4.7
6	Assertiveness	0.614	14.2	4.1
7	Persuasiveness	0.510	12.6	3.9
8	Liveliness	0.619	12.7	4.3
9	Networking	0.623	13.5	4.4
10	Work Tempo	0.672	15.7	4.2
11	Altruism	0.661	16.5	4.1
12	Cooperation	0.566	16.7	3.9
13	Criticalness	0.572	14.8	3.8
14	Individualism	0.571	12.8	4.0
15	Competitiveness	0.658	12.9	4.3
16	Drive	0.680	16.1	4.5
17	Methodicalness	0.683	17.0	4.5
18	Detail-concern	0.683	17.3	4.4
19	Dutifulness	0.581	16.8	4.1
20	Deliberation	0.593	18.4	3.8
21	Progressiveness	0.567	15.8	4.2
22	Creativity	0.692	14.6	4.6
23	Abstractedness	0.481	15.5	4.0
24	Aesthetic Sense	0.640	13.2	4.4
25	Observation	0.672	14.7	4.4

Trait-Map is a forced-choice questionnaire with only five items per scale, therefore the Cronbach's Alpha values are expected to be on the lower side (Rohit, 2024). For comparison, refer to the Cronbach's alpha values of the Sixteen Personality Factor Questionnaire (16PF) or the NEO-PI-R (Rossier et al., 2004).

The somewhat lower alpha values stem from the nature of the questionnaire, and therefore Cronbach's Alpha values should not be used as a measure of reliability for Trait-Map. Internal scale consistency remains an important development objective, but the Cronbach's Alpha values will always be somewhat lower than those of questionnaires with Likert-type items. Therefore, reliability should be assessed using the test-retest method.

*Simplified Chinese data set, n=16702*

The number of items per trait is 5. The range of trait raw scores is 5 – 25. The frequency distribution charts are shown in Annex 2.

#	Traits	ENG		
		Cronbach Alpha	Scale Means	Standard Deviation
1	Emotional Control	0.657	14.8	4.4
2	Calmness	0.604	15.0	4.0
3	Health Mgmt	0.690	14.7	4.7
4	Optimism	0.661	16.0	4.2
5	Confidence	0.666	14.4	4.2
6	Assertiveness	0.675	13.8	4.5
7	Persuasiveness	0.543	11.8	3.9
8	Liveliness	0.686	10.8	4.2
9	Networking	0.638	13.8	4.3
10	Work Tempo	0.518	15.5	3.8
11	Altruism	0.662	13.7	4.3
12	Cooperation	0.573	15.6	4.0
13	Criticalness	0.583	14.7	3.8
14	Individualism	0.569	12.6	3.9
15	Competitiveness	0.591	13.3	4.0
16	Drive	0.555	16.3	4.2
17	Methodicalness	0.610	18.3	4.0
18	Detail-concern	0.617	17.0	4.2
19	Dutifulness	0.559	18.3	3.9
20	Deliberation	0.601	18.1	3.9
21	Progressiveness	0.559	16.4	4.1
22	Creativity	0.670	14.8	4.2
23	Abstractedness	0.582	15.6	4.3
24	Aesthetic Sense	0.644	13.8	4.4
25	Observation	0.640	15.8	4.0

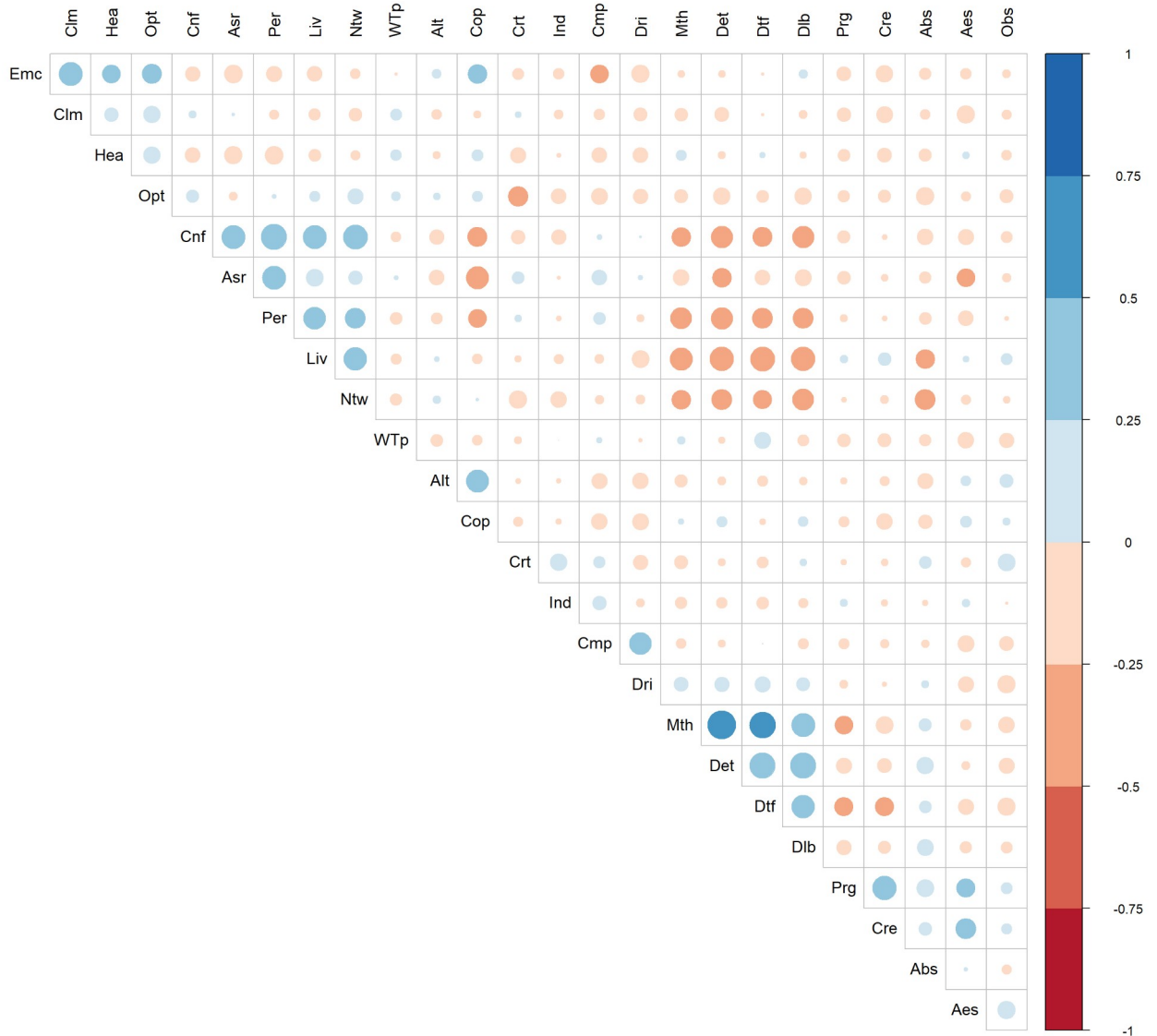
Trait-Map is a forced-choice questionnaire with only five items per scale, therefore the Cronbach’s Alpha values are expected to be on the lower side (Rohit, 2024). For comparison, refer to the Cronbach's alpha values of the Sixteen Personality Factor Questionnaire (16PF) or the NEO-PI-R (Rossier et al., 2004).

The somewhat lower alpha values stem from the nature of the questionnaire, and therefore Cronbach’s Alpha values should not be used as a measure of reliability for Trait-Map. Internal scale consistency remains an important development objective, but the Cronbach’s Alpha values will always be somewhat lower than those of questionnaires with Likert-type items. Therefore, reliability should be assessed using the test-retest method.

## Inter-scale Correlations

Examining the correlations between the scales of the construct based on the data collected provides important information on construct validity. The method of evaluating correlations is to compare observed trends with theoretical expectations. In practical terms, this means evaluating whether the relationships revealed by the correlations make sense.

### Inter-Scale Correlation Matrix – English data set (n=5918)

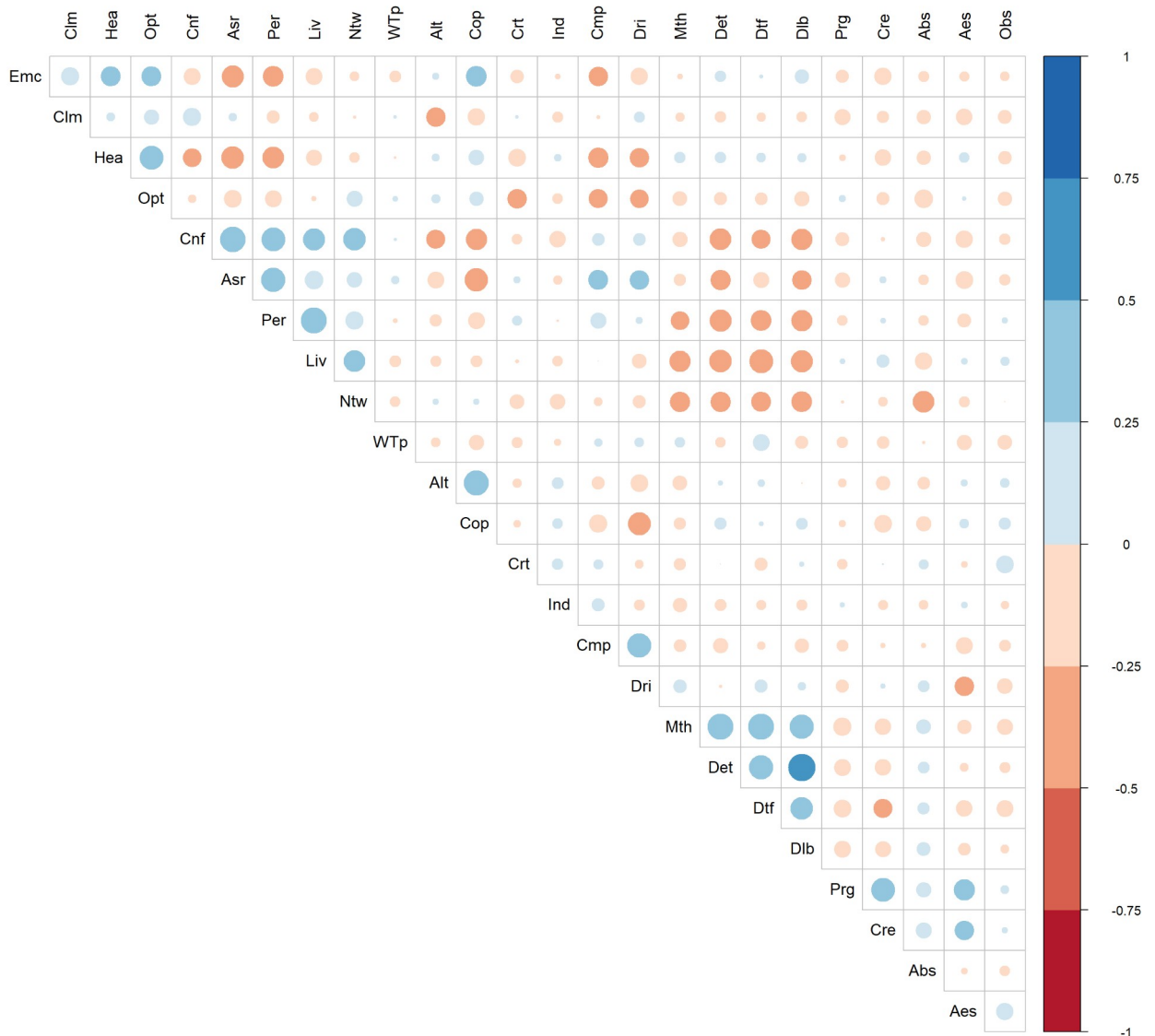


Emotional Control=Emc, Calmness=Clm, Health Mgmt.=Hea, Optimism=Opt, Confidence=Cnf, Assertiveness=Asr, Persuasiveness=Per, Liveliness=Liv, Networking=Ntw, Work Tempo=WTp, Altruism=Alt, Cooperation=Cop, Criticalness=Crt, Independence=Ind, Competitiveness=Cmp, Drive=Dri, Methodicalness=Mth, Detail-concern=Det, Dutifulness=Dtf, Deliberation=Dlb, Progressiveness=Prg, Creativity=Cre, Abstractedness=Abs, Aesthetic Sense=Aes, Observation=Obs

Blue represents positive, red represents negative correlations. Larger disc sizes and darker colors represent stronger correlations. The force-choice questionnaire method of Trait-Map amplifies negative correlations.

The correlations are in line with expectations.

## Inter-Scale Correlation Matrix – Chinese data set (n=16702)



Emotional Control=Emc, Calmness=Clm, Health Mgmt.=Hea, Optimism=Opt, Confidence=Cnf, Assertiveness=Asr, Persuasiveness=Per, Liveliness=Liv, Networking=Ntw, Work Tempo=WTP, Altruism=Alt, Cooperation=Cop, Criticalness=Crt, Independence=Ind, Competitiveness=Cmp, Drive=Dri, Methodicalness=Mth, Detail-concern=Det, Dutifulness=Dtf, Deliberation=Dlb, Progressiveness=Prg, Creativity=Cre, Abstractedness=Abs, Aesthetic Sense=Aes, Observation=Obs

Blue represents positive, red represents negative correlations. Larger disc sizes and darker colors represent stronger correlations. The force-choice questionnaire method of Trait-Map amplifies negative correlations.

The correlations are in line with expectations. The correlations in the different language versions show only small differences, and reveal fundamentally the same structure.

## **Standardization of results**

Standardization means presenting results in a familiar, meaningful, and uniform numerical format that is easy to interpret, even across scales and instruments. This requires converting raw scores into standardized scores using established methods. Trait-Map uses STEN scores, one of the most common methods for standardizing psychometric test scores. Because the trait raw scores are relatively close to the perfect normal distribution, Trait-Map follows the standard method of score conversion to attain STEN scores.

Traditionally, a STEN score represents an individual's performance compared to a large, representative sample, called the "norm group"; thus these norm groups provide a baseline for the comparison. Although STEN scores are used in Trait-Map, the interpretation of the standardized scores differs from the traditional meaning. Due to its forced-choice questionnaire format, Trait-Map's raw trait scores represent trait dominance compared to an individual's other traits. In theory, Trait-Map scoring could be constructed without using parameters from the analysis of larger groups. In practice, however, the authors of Trait-Map use sample segmentation and analysis (by questionnaire language and by gender) to monitor the instrument's psychometric properties, and the scoring system uses different parameters for standardization depending on the questionnaire language. Although the standardization parameters differ slightly by language, these sample groups should not be considered traditional norm groups. The STEN scores should be interpreted as indicators of trait dominance relative to an individual's other traits rather than relative to other people.

## Test-Retest Reliability

The test-retest method assesses a measure's consistency over time by administering the same test to the same group of individuals at two different time points, then calculating a correlation coefficient between the two sets of scores to determine stability. The study was done in cooperation with the Tongji University MBA alumni group. Sample size: n=38. Questionnaire language: Chinese. All participants were working adults, age was between 23-30 years old. There was a 3-week interval break between the first and second test administration. Test-Retest Reliability of primary scales:

	Test-Retest Reliability	Standard Error of Measurement *
Emotional Control	0.84	1.6
Calmness	0.76	1.9
Health Mgmt	0.88	1.4
Optimism	0.83	1.6
Confidence	0.87	1.4
Assertiveness	0.75	2
Persuasiveness	0.82	1.7
Liveliness	0.9	1.2
Networking	0.87	1.4
Work Tempo	0.78	1.8
Altruism	0.77	1.9
Cooperation	0.86	1.5
Criticalness	0.69	2.2
Individualism	0.76	1.9
Competitiveness	0.81	1.7
Drive	0.75	2
Methodicalness	0.82	1.7
Detail-concern	0.7	2.1
Dutifulness	0.81	1.7
Deliberation	0.79	1.8
Progressiveness	0.79	1.8
Creativity	0.85	1.5
Abstractedness	0.83	1.6
Aesthetic Sense	0.93	1
Observation	0.78	1.8
<b>AVG</b>	<b>0.81</b>	<b>1.7</b>

The average primary scale reliability is 0.81; the secondary scale reliability is 0.86.

The test-retest reliability is a number between zero and one, with 1 being a perfect correlation between the test and the retest results, and with 0 showing total irrelevance between the results. Perfection is impossible, and most researchers accept either 0.7, 0.8 or 0.9, depending upon the particular characteristics and application of the instrument. (Martyn Shuttleworth: Test-Retest Reliability. Retrieved Dec 07, 2016 from Explorable.com: <https://explorable.com/test-retest-reliability>).

\*: SEM is estimated with 95% Confidence Interval for STEN scores.

### Test-Retest Reliability of secondary scales:

Secondary scales in Trait-Map are calculated from the combination of primary scales (the 25 traits) using content based bottom-up scale forming method.

Emotional Stability	0.88	Pres Creating Content	0.87
Extroversion	0.94	Pres Preparation	0.86
Agreeableness	0.76	Pres Composure	0.81
Focus	0.86	Pres Reading Audience	0.78
Openness	0.89	Pres Engaging	0.94
Routine-work tolerance	0.86	Pres Arguing Viewpoints	0.9
Communicativeness	0.93	Neg Preparation	0.86
Service Orientation	0.85	Neg Composure	0.81
Sales Orientation	0.9	Neg Reading People	0.78
Obedience	0.75	Neg Convincing Others	0.9
Dominance	0.9	Neg Compromising	0.85
Emotional Expression	0.89	Col Composure	0.81
Adaptability	0.82	Col Proactive Communication	0.93
Stress Tolerance	0.83	Col Reading People	0.78
Burnout Risk	0.87	Col Convincing Others	0.9
Autocratic	0.9	Col Compromising	0.85
Visionary	0.87	WA Stress Tolerance	0.83
Democratic	0.85	WA Service Orientation	0.85
Leadership Orientation	0.91	WA Communicativeness	0.93
Coordinator	0.85	WA Sales Orientation	0.9
Helper	0.87	WA Leadership Orientation	0.91
Promoter	0.9	WA Routine Work Tolerance	0.86
Driver	0.83	WA Following Instructions	0.75
Planner	0.86	WA Adaptability	0.82
Maker	0.78	WA Emotional Expression	0.89
Innovator	0.87	WA Drive and Energy	0.84
Evaluator	0.79	WA Reading People	0.78
Forcing	0.93	WA Personal Organization	0.86
Collaborating	0.82	WA Creative	0.86
Accommodating	0.88	WA Abstract Thinking	0.88
Avoiding	0.88	WA Burnout Risk	0.87
Sales Stress Tolerance	0.83		
Sales Persistence	0.8		
Sales Networking	0.92		
Sales Reading People	0.78		
Sales Influencing	0.9		
Sales Goal Orientation	0.88		
Sales Preparation	0.86		
Sales Creative	0.86		
Sales Customer	0.85	<b>AVG</b>	<b>0.86</b>

The secondary scale reliability is 0.86. The test-retest reliability is a number between zero and one, with 1 being a perfect correlation between the test and the retest results, and with 0 showing total irrelevance between the results. Perfection is impossible, and most researchers accept either 0.7, 0.8 or 0.9, depending upon the particular characteristics and application of the instrument. (Martyn Shuttleworth: Test-Retest Reliability. Retrieved Dec 07, 2016 from Explorable.com: <https://explorable.com/test-retest-reliability>).

## Trait-Map Gender Effects Analysis

Even though nowadays the lifestyle of women is much similar to men's in developed countries, but the gender differences in personality are still questions of great interest. The following study looks into the gender differences captured by Trait-Map in various languages. This study was conducted in March 2024.

### Description of the data

A total of 28718 anonymous Trait-Map respondent responses were exported in March 2024. This data comes from recruitment, training and organizational diagnosis/development projects in Europe and Asia. The respondents are all working adults in white-collar jobs, and the data set represents a wide range of industries and jobs. 876 respondents completed Trait-Map in less than 5 minutes, and their data was removed from the analysis. An additional 728 respondents had Centredness scores between 8 and 10 (indicating low consistency), and their data was also removed from the analysis. The data set contained another 4870 respondents without gender information, and their data was also removed. Only language versions with larger respondent numbers have been included in the analysis. In the end, 5800 responses in English, 791 in German, 13516 in Simplified Chinese, and 952 in Vietnamese were used.

### Methodology

The analysis was performed separately for the four language groups (English, German, Simplified Chinese, and Vietnamese). The means of raw scores were calculated for female and male participants respectively for the 25 traits, and the results are displayed in the columns Means Female and Means Male.

The difference of female and male means was calculated by subtracting the male means from the female means, and the result is shown in the Difference column. If the female means is higher than the male means, the Difference is a positive number; if the male means is higher, then the Difference is a negative number.

The means and differences were calculated from the scale raw scores, not from standardized scores for more accuracy. In order to determine if a difference is small or big, Student's T-Test for Independent Samples was used to calculate the significance of the differences, as shown in column Sig. (2-tailed; equal variance of the compared data sets was not assumed).

The Sig. number indicates the probability that a result (in this case, a difference in means) is a coincidence. Sig. < 0.05 (less than 5% chance of being a coincidence, highlighted with light green) is considered as statistically significant, and Sig. < 0.001 (less than 0,1% chance of being a coincidence, highlighted with vivid green) as highly significant. A smaller Sig. indicates a higher significance of the result, in our case showing that there is a greater difference by gender.

The differences are also displayed as a percentage of the scale standard deviation to illustrate the magnitude of the difference between the gender means. Scale STD represents the standard deviation of the scale and is calculated as the average standard deviation of the male and female data sets. The % of STD column shows the difference as a percentage of the standard deviation, calculated as  $\text{Difference} / \text{Scale STD} \times 100$ .

## English data set

Female=3069, Male=2731, total n=5800. This is a geographically and culturally diverse data set from India, Kenya, Malaysia, Pakistan, Singapore, and the United Kingdom.

Traits	Means Female	Means Male	Difference	Sig.	Scale STD	% of STD
Emotional Control	13.842	13.996	-0.154	0.172	4.222	3.6
Calmness	14.936	15.502	-0.566	0	3.688	15.4
Health Mgmt	13.227	12.527	0.701	0	4.358	16.1
Optimism	17.133	16.204	0.929	0	4.031	23
Confidence	14.266	15.151	-0.886	0	4.755	18.6
Assertiveness	13.867	14.784	-0.916	0	4.123	22.2
Persuasiveness	12.904	13.732	-0.828	0	3.761	22
Liveliness	12.52	12.809	-0.288	0.011	4.269	6.8
Networking	14.145	13.755	0.39	0.001	4.396	8.9
Work Tempo	16.135	14.906	1.229	0	4.089	30.1
Altruism	17.015	16.449	0.566	0	4.071	13.9
Cooperation	17.035	16.522	0.513	0	3.935	13
Criticalness	13.702	14.626	-0.924	0	3.726	24.8
Individualism	12.582	12.736	-0.154	0.131	3.827	4
Competitiveness	11.785	13.583	-1.798	0	4.106	43.8
Drive	16.003	16.497	-0.493	0	4.432	11.1
Methodicalness	17.389	16.53	0.859	0	4.497	19.1
Detail-concern	17.628	16.869	0.759	0	4.342	17.5
Dutifulness	17.101	16.295	0.806	0	3.927	20.5
Deliberation	18.309	18.416	-0.107	0.278	3.783	2.8
Progressiveness	15.802	15.738	0.064	0.558	4.142	1.6
Creativity	14.193	14.93	-0.737	0	4.439	16.6
Abstractedness	15.031	16.002	-0.971	0	3.905	24.9
Aesthetic Sense	13.617	12.528	1.089	0	4.248	25.6
Observation	14.833	13.915	0.917	0	4.438	20.7

The female respondents scored higher by 20% or more of the scale standard deviation in Optimism, Work Tempo, Dutifulness, Aesthetic Sense, Observation. The male respondents scored higher by 20% or more of the scale standard deviation in Assertiveness, Persuasiveness, Criticalness, Competitiveness, Abstractedness.

## German data set

Female=284, Male=507, total n=791. This data set is from Germany.

Traits	Means Female	Means Male	Difference	Sig.	Scale STD	% of STD
Emotional Control	12.535	13.179	-0.644	0.019	3.755	17.2
Calmness	14.268	15.44	-1.172	0	4.019	29.2
Health Mgmt	13.246	12.069	1.177	0	4.14	28.4
Optimism	16.905	16.168	0.737	0.034	4.631	15.9
Confidence	14.827	15.118	-0.291	0.339	4.057	7.2
Assertiveness	14.574	16.41	-1.836	0	4.154	44.2
Persuasiveness	14.31	15.325	-1.016	0.001	3.923	25.9
Liveliness	14.627	13.75	0.877	0.006	4.283	20.5
Networking	12.859	11.992	0.867	0.007	4.32	20.1
Work Tempo	14.775	14.085	0.69	0.016	3.868	17.8
Altruism	16.169	14.89	1.279	0	4.345	29.4
Cooperation	15.317	14.773	0.544	0.091	4.361	12.5
Criticalness	15.923	16.72	-0.797	0.004	3.714	21.5
Individualism	14.109	14.702	-0.593	0.022	3.522	16.8
Competitiveness	10.518	12.442	-1.924	0	3.933	48.9
Drive	14.06	14.933	-0.873	0.005	4.165	21
Methodicalness	16.021	16.321	-0.3	0.364	4.45	6.7
Detail-concern	16.669	16.462	0.207	0.553	4.692	4.4
Dutifulness	16.877	17.059	-0.182	0.527	3.884	4.7
Deliberation	17.035	17.753	-0.718	0.022	4.16	17.3
Progressiveness	16.387	16.004	0.383	0.204	4.127	9.3
Creativity	15.419	14.915	0.504	0.132	4.536	11.1
Abstractedness	14.915	16.471	-1.556	0	4.231	36.8
Aesthetic Sense	15.109	12.692	2.417	0	4.849	49.8
Observation	17.546	15.325	2.22	0	4.221	52.6

The female respondents scored higher by 20% or more of the scale standard deviation in Health Management, Liveliness, Networking, Altruism, Aesthetic Sense, Observation. The male respondents scored higher by 20% or more of the scale standard deviation in Calmness, Assertiveness, Persuasiveness, Criticalness, Competitiveness, Abstractedness.

### Simplified Chinese data set

Female=5416, Male=8100, total n=13516. This data set is from the People's Republic of China.

Traits	Means Female	Means Male	Difference	Sig.	Scale STD	% of STD
Emotional Control	14.241	14.881	-0.64	0	4.399	14.5
Calmness	14.206	14.769	-0.563	0	4.087	13.8
Health Mgmt	15.532	14.168	1.364	0	4.633	29.4
Optimism	16.626	15.318	1.308	0	4.217	31
Confidence	14.247	14.635	-0.388	0	4.288	9
Assertiveness	13.395	14.11	-0.715	0	4.495	15.9
Persuasiveness	11.772	12.068	-0.296	0	3.905	7.6
Liveliness	11.089	10.767	0.321	0	4.265	7.5
Networking	13.919	13.637	0.282	0	4.368	6.4
Work Tempo	16.092	15.239	0.853	0	3.799	22.4
Altruism	13.925	13.886	0.039	0.602	4.261	0.9
Cooperation	16.151	15.439	0.712	0	3.995	17.8
Criticalness	14.653	14.709	-0.056	0.408	3.816	1.5
Individualism	12.538	12.512	0.026	0.704	3.898	0.7
Competitiveness	12.484	13.682	-1.198	0	4.042	29.6
Drive	15.051	17.042	-1.991	0	4.093	48.6
Methodicalness	18.316	18.376	-0.059	0.401	4.015	1.5
Detail-concern	16.787	16.948	-0.16	0.034	4.256	3.8
Dutifulness	18.382	18.299	0.083	0.23	3.912	2.1
Deliberation	17.73	18.557	-0.828	0	3.937	21
Progressiveness	17.076	16.008	1.068	0	4.052	26.4
Creativity	14.472	15.112	-0.64	0	4.331	14.8
Abstractedness	14.637	16.611	-1.974	0	4.217	46.8
Aesthetic Sense	15.05	12.816	2.234	0	4.389	50.9
Observation	16.629	15.41	1.219	0	4.058	30

The female respondents scored higher by 20% or more of the scale standard deviation in Health Management, Optimism, Work Tempo, Progressiveness, Aesthetic Sense, Observation. The male respondents scored higher by 20% or more of the scale standard deviation in Competitiveness, Drive, Abstractedness.

### Vietnamese data set

Female=514, Male=438, total n=952. This data set is from Vietnam.

Traits	Means Female	Means Male	Difference	Sig.	Scale STD	% of STD
Emotional Control	14.29	14.717	-0.427	0.142	4.471	9.5
Calmness	14.352	15.034	-0.682	0.008	3.892	17.5
Health Mgmt	14.399	13.386	1.013	0.001	4.741	21.4
Optimism	15.416	15.164	0.252	0.332	4.003	6.3
Confidence	13.453	13.929	-0.476	0.098	4.422	10.8
Assertiveness	15.909	16.08	-0.171	0.508	3.976	4.3
Persuasiveness	12.634	13.11	-0.475	0.046	3.663	13
Liveliness	12.14	12.199	-0.059	0.847	4.685	1.2
Networking	14.447	13.468	0.979	0.001	4.592	21.3
Work Tempo	16.119	15.865	0.253	0.346	4.134	6.1
Altruism	17.348	16.781	0.567	0.041	4.27	13.3
Cooperation	18.226	17.514	0.712	0.011	4.306	16.5
Criticalness	16.123	16.566	-0.444	0.057	3.593	12.3
Individualism	14.292	13.811	0.481	0.073	4.12	11.7
Competitiveness	11.056	12.226	-1.17	0	3.902	30
Drive	14.014	15.842	-1.829	0	4.544	40.2
Methodicalness	17.097	17.089	0.008	0.975	4.084	0.2
Detail-concern	15.776	15.477	0.299	0.282	4.282	7
Dutifulness	15.969	16.208	-0.239	0.374	4.138	5.8
Deliberation	18.008	18.555	-0.547	0.044	4.184	13.1
Progressiveness	15.411	14.546	0.865	0.001	4.032	21.4
Creativity	13.926	15.189	-1.263	0	4.645	27.2
Abstractedness	13.875	15.121	-1.246	0	4.217	29.5
Aesthetic Sense	14.154	12.221	1.932	0	4.238	45.6
Observation	16.566	14.902	1.664	0	4.228	39.4

The female respondents scored higher by 20% or more of the scale standard deviation in Health Management, Networking, Progressiveness, Aesthetic Sense, Observation. The male respondents scored higher by 20% or more of the scale standard deviation in Competitiveness, Drive, Creativity, Abstractedness.

## Discussion

There are statistically significant differences between the genders in their personality traits as measured by the Trait-Map assessment in all data sets.

It is consistent in all data sets, that women scored higher by 20% or more of the scale standard deviation in Aesthetic Sense and Observation, and men scored higher by 20% or more of the scale standard deviation in Competitiveness and Abstractedness. There are also a number of other traits that consistently show significant differences of smaller magnitudes. Women consistently scored significantly higher also on Health Management, and Networking. Men consistently scored significantly higher also on Calmness, Persuasiveness, and Drive.

The traits with the largest differences are somewhat different in the various data sets. This may suggest that gender differences are influenced by culture. However, we also need to keep in mind that the data doesn't include information about age, position, and company type, and the data may not be an adequate sample to represent cultures.

Even the most statistically significant differences are not large in absolute terms, no difference greater than 53% of the scale standard deviation was found. The degree of gender impact measured by Trait-Map is less than the legal threshold of 80% of one standard deviation for all scales, therefore it can be concluded that Trait-Map would not have a negative impact based on gender if it were used in recruitment and selection.

It can be concluded that there are statistically significant differences between the genders in their personality trait scores. These differences are consistent with the expectations stemming from gender stereotypes. However, the magnitude of these differences is not great.

## Criterion related validity studies

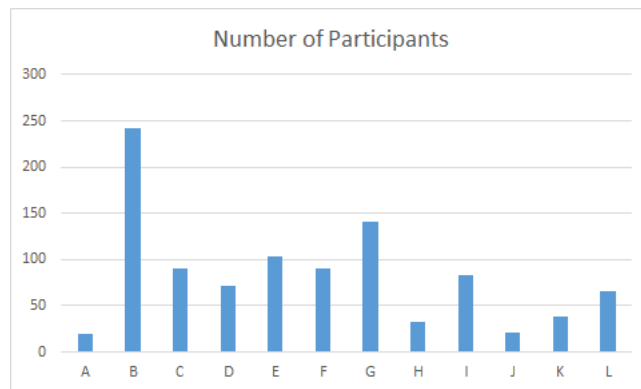
The most important question regarding questionnaire validity is how the measured scales predict critical real-world outcomes. Three validity studies that address this question are presented below.

### Validity Study 1.

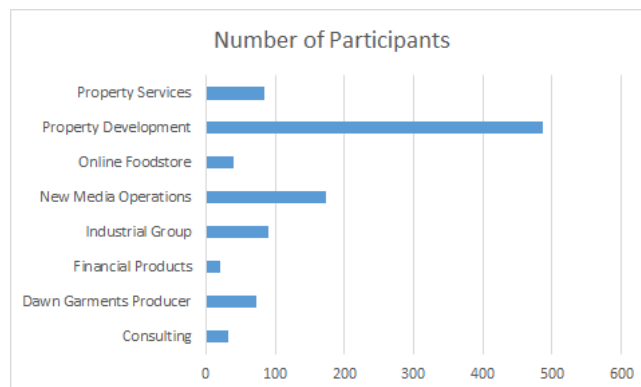
A study from 2023 examined the correlation between personality assessment results and job performance. A consulting firm collected Trait-Map scores and Performance Appraisal results from a total of 1000 employees in 12 companies. Trait-Map along with the OD-Tools Motivation Questionnaire (MQ) was used as part of organizational development projects. The performance ratings were given by the direct supervisor as part of the companies' regular performance management process. Performance ratings were provided in 3 grades: High, Medium, Low. The performance appraisal processes were completely independent of the psychometric assessment results.

### Participant background

Number of employees per company (companies are represented by A, B, C, etc.):



Number of employees per industry:



All data are from Jiangxi Province, China. Company F is government-owned (9% of participants), the other participants work in the private sector. There are 213 managers in the data set, and 560 non-managers (the others are unknown). Additional information such as gender, age, department, job type, or job title were not provided to the authors.

## Results 1.

Based on the full data set, the following scales correlate significantly with performance:

Tool	Scale	r	p
MQ	RECS	0.19	0.000
Trait-Map	Assertiveness	0.16	0.000
MQ	Achievement	0.15	0.000
MQ	Excellence	0.14	0.000
MQ	Autonomy	0.12	0.000
Trait-Map	Competitiveness	0.11	0.000
MQ	IMI	0.11	0.001
MQ	STA	0.10	0.002
MQ	Power	0.10	0.002
MQ	Environment	0.10	0.002
MQ	OPE - Output of Energy	0.10	0.002
MQ	STAS	0.09	0.005
MQ	AutonomyS	0.09	0.007
MQ	JoyS	0.09	0.004
MQ	Challenge	0.09	0.006
Trait-Map	Drive	0.08	0.011
MQ	JobSecurityS	0.08	0.010
MQ	PowerS	0.08	0.007
MQ	AchievementS	0.08	0.014
Trait-Map	Persuasiveness	0.07	0.021
MQ	ExcellenceS	0.07	0.035
MQ	Positivity	0.07	0.019
Trait-Map	Networking	-0.06	0.047
Trait-Map	Cooperation	-0.06	0.041
MQ	EnvironmentS	-0.08	0.017
Trait-Map	Altruism	-0.10	0.002
Trait-Map	Emotional Control	-0.11	0.001
Trait-Map	Health Mgmt.	-0.12	0.000
Trait-Map	Optimism	-0.12	0.000

Pearson correlation values are shown in column r, Significance (2-tailed) in column p.

$p < 0.05$  is considered as statistically significant and  $p < 0.001$  (less than one in a thousand chance of being wrong) as highly significant.

$n=1000$

Scales from the OD-Tools Motivation Questionnaire (MQ, marked in gray) were also included in this list for comparison.

The Trait-Map scales Assertiveness, Competitiveness, Drive and Persuasiveness show significant positive correlations with performance; Networking, Cooperation, Altruism, Emotional Control, Health Mgmt., Optimism show significant negative correlations.

The correlation values are moderate or low.

## Discussion

The table above shows which scales of psychometric instruments predict performance across the board, regardless of job type or company. We find that Assertiveness and Persuasiveness are included, which is consistent with both practitioner experience and theory, as these traits are widely recognized as signs of leadership potential. While the data here includes non-managers as well as managers, it can be argued that affinity for a decisive, persuasive communication style provides benefits across the board. It also seems reasonable that Competitiveness and Drive emerge as predictors, as these traits are signs of professional ambition. The moderate or low correlation values are reasonable because there are many critical factors that affect performance ratings but are not captured in Trait-Map. The high statistical significance shows that certain personality traits, as measured by Trait-Map, unquestionably have a systematic relationship with performance ratings.

The authors do not have a simple explanation for the negative correlations. It may have something to do with the local culture in Jiangxi Province, or it may be an artifact of the forced-choice questionnaire format, which is known to magnify negative correlations between scales and external criteria.

About the MQ scales: MQ's RECS (Satisfaction with Recognition by Direct Supervisor) is expected to be the top correlating scale because the performance ratings are given by the direct supervisors. The MQ scales Need for Achievement, Need for Excellence, Need for Autonomy are signs of intrinsic motivation, and are widely recognized as drivers of performance, so these results are consistent with theory. Seeing more MQ than Trait Map scales in the significant predictor list also seems reasonable because higher motivation is likely to drive performance across the board regardless of job type, whereas personality is known to predict performance mostly through the person-job match, where the predictive traits are job specific.

## Results 2.

In case of the Managers, the following scales correlate significantly with performance:

Tool	Scale	r	p
MQ	RECS	0.26	0.000
MQ	Achievement	0.24	0.000
Trait-Map	Competitiveness	0.23	0.001
MQ	Perceived Control	0.23	0.001
MQ	IMI	0.22	0.002
Trait-Map	Assertiveness	0.21	0.002
MQ	Challenge	0.21	0.002
MQ	Power	0.20	0.004
MQ	Excellence	0.19	0.004
MQ	PowerS	0.18	0.008
MQ	JobSecurityS	0.17	0.016
MQ	Autonomy	0.16	0.016
MQ	JoyS	0.16	0.021
MQ	STAS	0.14	0.038
MQ	Positivity	0.14	0.037
Trait-Map	Altruism	-0.14	0.036

Pearson correlation values are shown in column r, Significance (2-tailed) in column p.

$p < 0.05$  is considered as statistically significant and  $p < 0.001$  (less than one in a thousand chance of being wrong) as highly significant.

n=213

Scales from the OD-Tools Motivation Questionnaire (MQ, marked in gray) were also included in this list for comparison.

The Trait-Map scales Competitiveness and Assertiveness show significant positive correlations with performance; and Altruism show significant negative correlation.

The correlation values are moderate or low.

## Discussion

The table above shows which scales of psychometric instruments predict the performance of managers across different industries. Among the Trait-Map scales, Competitiveness emerged as the top predictor, probably because leadership positions are highly competitive. We also find that Assertiveness is also included as a top predictor, which is consistent with both practitioner experience and theory, since this trait is widely recognized as a sign of leadership potential. Altruism is negatively correlated with performance, which is consistent with many other studies that have found similar traits (typically Agreeableness) to be negatively correlated with managerial performance. Leaders cannot be too nice; beyond a point, that can affect their ability to hold people accountable. The high statistical significance shows that certain personality traits, as measured by Trait-Map, unquestionably have a systematic relationship with managerial performance ratings.

About the MQ scales: MQ's RECS (Satisfaction with Recognition by Direct Supervisor) is expected to be the top correlating scale because the performance ratings are given by direct supervisors. Finding more MQ than Trait-Map scales included in the significant predictor list suggests that motivation plays a critical role in managerial performance across the board regardless of industry or job level. Further discussion of MQ scales is included in the MQ technical manual.

### Results 3.

Comparing the Trait-Map means (standardized ten scores of the 25 traits) for the Manager (n=213) and Non-Manager (n=560) groups. The two groups have statistically significant differences on the following traits:

	Managers	Non-Managers	Difference	p
Emotional Cont	5.51	6.13	-0.62	0.000
Calmness	7.38	6.94	0.44	0.001
Health Mgmt	5.2	5.99	-0.79	0.000
Optimism	5.57	6.15	-0.58	0.001
Confidence	5.76	5.13	0.63	0.000
Assertiveness	5.94	4.45	1.49	0.000
Persuasiveness	4.32	3.96	0.36	0.015
Cooperation	4.64	5.22	-0.58	0.001
Drive	6.68	5.67	1.01	0.000
Methodicalness	6.7	6.37	0.33	0.024
Detail-concern	5.73	6.13	-0.40	0.012
Progressiveness	4.59	5.49	-0.90	0.000
Abstractedness	4.85	4.48	0.37	0.020
Aesthetic Sense	5.38	5.97	-0.59	0.001

The Difference is calculated as Mean Score of Managers – Mean Score of Non-Managers. It is marked with green where Managers > Non-managers, and marked with red where Managers < Non-managers.

Significance (p) was calculated by using t-test for independent samples (equal variances were not assumed).

p < 0.05 is considered as statistically significant and p < 0.001 (less than one in a thousand chance of being wrong) as highly significant.

### Discussion

Based on the data, Managers are significantly more Assertive, Driven, Confident, Calm, and Persuasive than Non-managers; and also somewhat more Abstract and Methodical. These results are consistent with both practitioner experience and theory. The lower optimism score for managers may come as a surprise, but a closer look reveals that the score of 5.57 is in the middle of a scale of ten, and not low in absolute terms. It may well be that being overly optimistic is a hindrance to becoming a manager, and that being realistic is more effective. The other differences marked with red are consistent with common sense: Managers are less emotionally restrained; more workaholic; less “nice and easy to work with”; less detail oriented; more conservative and more pragmatic than Non-managers.

The above differences between the means and the high statistical significance (low p values) provide evidence that Trait-Map does capture meaningful differences in the personality of Managers and Non-managers.

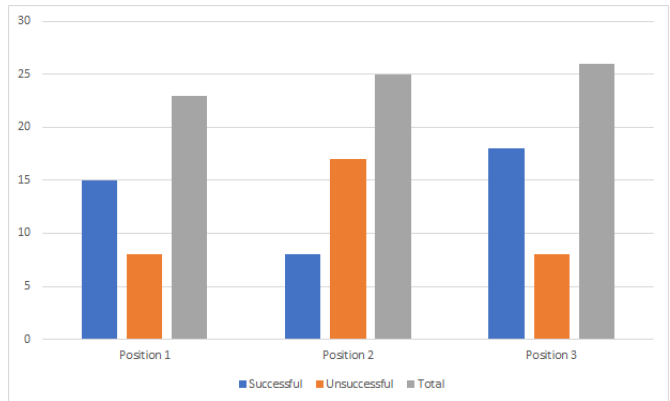
## Validity Study 2.

The data used in this study originates from the service industry, from a global travel management company. The company used Trait-Map to make their recruitment interviews more robust, not as the sole basis for decision making. Over the years, the company accumulated a list of Successful and Unsuccessful employees in their 3 key positions with the highest number of employees. All people included in the study have passed the rigorous selection process of the company and have been hired. Successful means people who adapted to the company's high-performance culture and stayed on the job. Unsuccessful means people who were fired or left the company for whatever reason.

Position 1: B2C Sales type position.  
Successful employees: 15,  
Unsuccessful: 8. Total: 23.

Position 2: B2B Sales type position.  
Successful employees: 8, Unsuccessful: 9.  
Total: 17.

Position 3: Administrative-service type position.  
Successful employees: 18,  
Unsuccessful: 8. Total: 26.







The means and standard deviations of Trait-Map scores have been compared between the Successful and Unsuccessful groups for each of the three jobs. The data analysis was combined with an understanding of the job to create target Job Profiles (personality profiles of the ideal candidate) as shown below. The numbers in the triangles in the graphs below represent the importance of the trait and will be used later as weights in a weighted decision matrix calculation.




Position 1: B2C Sales type position success profile:

▲	Scale	Ideal Range										Definition
4	Assertiveness	1	2	3	4	5	6	7	8	9	10	Frequency and intensity of attempts to express ones desires and to control situations (Trait-Map)
	Networking	1	2	3	4	5	6	7	8	9	10	The need, search and enjoyment of the company of others (Trait-Map)
	Drive	1	2	3	4	5	6	7	8	9	10	Level of ambition and the degree to which one strives towards achievement (Trait-Map)
3	Work Tempo	1	2	3	4	5	6	7	8	9	10	The preferred speed of work (Trait-Map)
	Competitiveness	1	2	3	4	5	6	7	8	9	10	The inclination to, and enjoyment of, competition (Trait-Map)
	Methodicalness	1	2	3	4	5	6	7	8	9	10	The preference of order and standard working procedures (Trait-Map)
	Deliberation	1	2	3	4	5	6	7	8	9	10	Thinking through possibilities before acting (Trait-Map)
	Abstractedness	1	2	3	4	5	6	7	8	9	10	The comfort with, and desire for, doing analysis, working with data and abstract ideas (Trait-Map)
2	Optimism	1	2	3	4	5	6	7	8	9	10	Positive outlook, tendency to experience positive moods and high spirits (Trait-Map)
	Persuasiveness	1	2	3	4	5	6	7	8	9	10	Rhetoric, debate and being verbally convincing (Trait-Map)
	Criticalness	1	2	3	4	5	6	7	8	9	10	The tendency of finding and pointing out faults and judging strictly (Trait-Map)
	Progressiveness	1	2	3	4	5	6	7	8	9	10	The attitude towards novelty and change (Trait-Map)
	Aesthetics	1	2	3	4	5	6	7	8	9	10	Tendency to notice and be attracted to and inspired by beauty (Trait-Map)

Position 2: B2B Sales type position success profile:

	Scale	Ideal Range									Definition	
	Optimism	1	2	3	4	5	6	7	8	9	10	Positive outlook, tendency to experience positive moods and high spirits (Trait-Map)
	Competitiveness	1	2	3	4	5	6	7	8	9	10	The inclination to, and enjoyment of, competition (Trait-Map)
	Assertiveness	1	2	3	4	5	6	7	8	9	10	Frequency and intensity of attempts to express ones desires and to control situations (Trait-Map)
	Networking	1	2	3	4	5	6	7	8	9	10	The need, search and enjoyment of the company of others (Trait-Map)
	Drive	1	2	3	4	5	6	7	8	9	10	Level of ambition and the degree to which one strives towards achievement (Trait-Map)
	Persuasiveness	1	2	3	4	5	6	7	8	9	10	Rhetoric, debate and being verbally convincing (Trait-Map)
	Cooperation	1	2	3	4	5	6	7	8	9	10	Willingness to sacrifice own interests for the sake of harmony or group unity (Trait-Map)
	Individualism	1	2	3	4	5	6	7	8	9	10	Tendency to form firm views and stick to them (Trait-Map)
	Dutifulness	1	2	3	4	5	6	7	8	9	10	Sense of obligation; tendency to persist at tasks until they are completed (Trait-Map)

Position 3: Administrative-service type position success profile:

	Scale	Ideal Range									Definition	
	Criticalness	1	2	3	4	5	6	7	8	9	10	The tendency of finding and pointing out faults and judging strictly (Trait-Map)
	Detail-concern	1	2	3	4	5	6	7	8	9	10	Tendency to focus on details and work precisely (Trait-Map)
	Dutifulness	1	2	3	4	5	6	7	8	9	10	Sense of obligation; tendency to persist at tasks until they are completed (Trait-Map)
	Emotional Control	1	2	3	4	5	6	7	8	9	10	Tendency to avoid outbursts of anger or frustration and maintain emotional control (Trait-Map)
	Assertiveness	1	2	3	4	5	6	7	8	9	10	Frequency and intensity of attempts to express ones desires and to control situations (Trait-Map)
	Persuasiveness	1	2	3	4	5	6	7	8	9	10	Rhetoric, debate and being verbally convincing (Trait-Map)
	Altruism	1	2	3	4	5	6	7	8	9	10	Tendency to assist those in need without thinking of own benefits (Trait-Map)
	Drive	1	2	3	4	5	6	7	8	9	10	Level of ambition and the degree to which one strives towards achievement (Trait-Map)

This study will examine what would happen if Trait-Map were used as the sole basis for making hiring decisions. This is not a recommended use of Trait-Map, rather a theoretical evaluation of the accuracy of a hypothetical decision making based solely on Trait-Map scores. How accurately would it differentiate between Successful and Unsuccessful candidates?

The answer is displayed in the tables below. This kind of table is called table of confusion (sometimes also called a confusion matrix or error matrix). It is a table with two rows and two columns that reports the number of false positives (wrongly hired), false negatives (wrongly rejected), true positives (correctly hired), and true negatives (correctly rejected). In our case it shows the number of wrong (marked with red) and correct hiring decisions (marked with green) both in the Hired and Rejected groups. Based on the above job profiles and Trait-Map scores, a job match score was calculated for each individual, to express the degree of closeness to the success profiles. A Match score is expressed in percentages: 100% is full match, 0% is maximum distance. In this thought experiment, candidates with a Match score higher than a certain Passing score would be hired; those with lower or equal scores would be rejected. The tables below show the decision accuracy depending on what passing scores would be implemented.

Position 1:

Passing score	Wrong	Correct	Decision Accuracy	
80%	5	12	Hired	71%
	3	3	Rejected	50%
85%	3	11	Hired	79%
	4	5	Rejected	56%
90%	2	7	Hired	78%
	8	6	Rejected	43%
95%	0	2	Hired	100%
	13	8	Rejected	31%

Position 2:

Passing score	Wrong	Correct	Decision Accuracy	
85%	5	8	Hired	62%
	0	4	Rejected	100%
90%	4	7	Hired	64%
	1	5	Rejected	83%
95%	0	4	Hired	100%
	4	9	Rejected	69%

Position 3:

Passing score	Wrong	Correct	Decision Accuracy	
80%	5	15	Hired	71%
	3	3	Rejected	50%
85%	2	13	Hired	87%
	5	6	Rejected	55%
90%	2	10	Hired	83%
	8	6	Rejected	43%
95%	0	5	Hired	100%
	13	8	Rejected	38%

It can be concluded that Decision Accuracy would be between 38% and 100%, depending on the Passing score, in another word depending on our tolerance for false negative (wrongfully rejected) or false positive (wrongfully hired). If there are many candidates, it would be possible to optimize the Decision Accuracy using Trait-Map to achieve high probability of success. It is not recommended to use Trait-Map as a single selection tool, but the study suggests that properly identified personality traits, as measured by Trait-Map, have systematic and reasonable relationship with success on the job.

### Validity Study 3.

#### Psychometric Testing Correlated with Assessment Center Results

The following study is based on a dataset of 111 junior and middle managers in an MNC. The managers participated in a leadership Assessment Center to assess their readiness for promotion. The AC consisted of an In-Basket Exercise, a Team Game with Leaderless Group Discussion, a Role Play, a Presentation Exercise, and a Behavioral Event Interview. Participants also completed the Trait-Map. Scoring of the AC was independent of the Trait-Map scores.

#### Method of analysis:

The participants were divided into two groups based on their AC total scores: Lower scoring group (80% of participants); Higher scoring group (the top 20%). The psychometric results' differences between the groups were analysed using independent samples t-tests to determine statistical significance.

The significance value indicates the probability that the difference in means is due to coincidence. The smaller the significance value, the more significant the trend. Depending on the purpose of the study, the cut-off value of significance (the required value used to determine statistical significance) may be set at less than 0.1 (the chance of coincidence is less than 10%); less than 0.05 (the chance of coincidence is less than 5%); or less than 0.001 (the chance of coincidence is less than 0.1%). The following table shows the differences in the Trait-Map standardized (STEN) scores between the two groups (the difference is calculated by subtracting the means of the Lower scoring group from the means of the Higher scoring group), and the statistical significance of the difference. The statistically significant differences (scales where the significance value is less than 0.1) are highlighted in green.

Scale	Means - Lower scoring group	Means - Higher scoring group	Difference of means	Significance
Emotional Control	6.027	5.091	-0.936	0.09266
Calmness	6.91	6.909	-0.001	0.99823
Health Mgmt	5.477	4.455	-1.023	0.06158
Optimism	5.892	4.864	-1.028	0.03006
Confidence	5.577	6.409	0.833	0.07703
<b>Assertiveness</b>	<b>4.919</b>	<b>6.818</b>	<b>1.899</b>	<b>0.00084</b>
Persuasiveness	4.189	5.318	1.129	0.01626
Liveliness	3.802	4.5	0.698	0.14952
Networking	5.505	6.727	1.223	0.01501
Work Tempo	4.73	4.955	0.225	0.58908
Altruism	4.568	3.591	-0.977	0.12563
Cooperation	5.009	4.227	-0.782	0.09403
Criticalness	6	5.864	-0.136	0.76842
Individualism	4.595	4.364	-0.231	0.6574
Competitiveness	5.685	5.909	0.224	0.64578
Drive	6.315	6.955	0.639	0.15208
Methodicalness	6.369	6.5	0.131	0.75036
Detail-concern	6.009	5.409	-0.6	0.23799
Dutifulness	6.396	6.091	-0.305	0.45522
<b>Deliberation</b>	<b>6.775</b>	<b>5.318</b>	<b>-1.457</b>	<b>0.00548</b>
Progressiveness	5.261	4.955	-0.307	0.54185
Creativity	5.423	6.364	0.94	0.05539

Abstractedness	5.261	4.591	-0.67	0.12496
Aesthetic Sense	4.883	4.773	-0.11	0.84614
Observation	6.072	6.5	0.428	0.37538

## Discussion

The result shows that significant differences between the two groups' Trait-Map scores exist.

The Higher scoring group scored higher on the traits related to Extroversion: Confidence, Assertiveness, Persuasiveness, and Networking. This is understandable, since the AC requires a great deal of interpersonal interaction.

The Higher scoring group also scored higher on Creativity. This is understandable since the AC requires participants to perform new tasks under new conditions.

The Lower scoring group scored higher on the traits related to Emotional Steadiness: Emotional Control, Health Mgmt, Optimism (not on Calmness). A possible reason for this may be that the AC exercises have a strict time limit, and participants with a higher sense of urgency tend to perform better. This is consistent with the finding that the Lower scoring group also scored higher on Deliberation, meaning that participants who make quicker decisions tend to perform better on the AC.

## **Conclusion**

The current version of Trait-Map represents more than 20 years of development and numerous iterations, with ongoing refinement. The psychometric considerations detailed here are essential for evaluating any assessment tool. We hope this paper has provided valuable insights into the properties of the Trait-Map Personality Assessment and thank you for your interest.

## References

- Cao M., Drasgow F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, 104(11), 1347-1368. <https://doi.org/10.1037/apl0000414>
- Credé M., Bashshur M., Niehorster S. (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment*, 92(5), 390-399. <https://doi.org/10.1080/00223891.2010.497393>
- Hough, L. M., & Johnson, J. W. (2013). Use and importance of personality variables in work settings. In N. W. Schmitt, S. Highhouse, & I. B. Weiner (Eds.), *Handbook of psychology: Industrial and organizational psychology* (2nd ed., pp. 211–243). John Wiley & Sons, Inc..
- Loy, R.W., Christiansen, N.D., Tett, R.P., Klein, K. and Toich, M. (2025), Personality Test Validity Differs Between Low-Stakes and High-Stakes Employment Settings. *International Journal of Selection and Assessment*, 33: e70018. <https://doi.org/10.1111/ijsa.70018>
- Paulhus D. L., Vazire S. (2007). The self-report method. In Robins R. W., Fraley R. C., Krueger R. F. (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). Guilford.
- Rohit Vishal Kumar (2024). Cronbach's Alpha: Genesis, Issues and Alternatives. *IMIB Journal of Innovation and Management*, Volume 2, Issue 2 July 2024. <https://doi.org/10.1177/ijim.241234970>
- Rossier, J., Meyer de Stadelhofen, F., & Berthoud, S. (2004). The Hierarchical Structures of the NEO PI-R and the 16PF5. *European Journal of Psychological Assessment*, 20(1), 27–38. <https://doi.org/10.1027/1015-5759.20.1.27>
- Schulte, N., Holling, H., & Bürkner, P.-C. (2021). Can High-Dimensional Questionnaires Resolve the Ipsativity Issue of Forced-Choice Response Formats? *Educational and Psychological Measurement*, 81(2), 262–289. <https://doi.org/10.1177/0013164420934861>
- Wetzel E., Böhnke J. R., Brown A. (2016). Response biases. In Leong F. T. L., Bartram D., Cheung F. M., Geisinger K. F., Iliescu D. (Eds.), *The ITC international handbook of testing and assessment* (pp. 349-363). Oxford University Press.

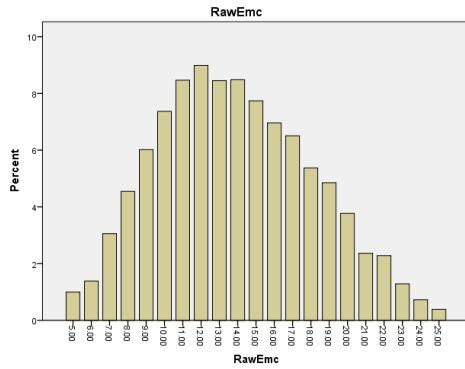
## Annex 1. User organizations:

3M Tax  
Abbott Vascular Deutschland GmbH  
ABSAP & AITS  
ActiveX Consulting  
AHK  
AKG Taicang  
AMSB  
Amway Business Services Asia Pacific Sdn. Bhd.  
ATI  
BCF Group  
B-Coaching  
BERTONI Consulting  
BITGroup  
BLUEX International  
Business Solutions Schwarz  
Carl Götz Consulting  
CeDR Corporate Consulting Sdn Bhd  
Cement Industries of Malaysia Berhad  
Cenviro Sdn Bhd  
Cesvim  
CK Consulting  
CLASSIC INTELLIGENT  
CONBEN Consulting  
Conben Deutschland  
C-TALENTS  
Dakoko  
DaLat Worsted Spinning  
DanhKhoiHR  
DHR PARTNER  
Directions Consulting  
Dr. Franco Ferreri Consulting  
East-West Leadership  
Easun Technology  
Edag  
Engelmann.Training  
EOD VietNam  
FA  
FES Partners  
FESCO Adecco  
fischer (Taicang) fixings Co., Ltd.  
Flamingo Consulting  
Flight Centre Travel Group Ltd (Greater China)  
Fox40 Agency  
GENIO Consulting  
Gentherm  
German Chamber (AHK) Shanghai  
Giovanna Squicciarini Consulting  
GTEX ASIA SDN BHD  
H2O srl  
HAN EAD Consulting  
HappyHR Consulting  
HEIW (Health Education and Improvement Wales)/WCPPE  
Hoài Văn Trinh  
Holland Organics  
HPM  
HR Guru  
HRO Consulting Shanghai  
Humanfield  
INTOMIND Coaching & Consultancy  
Jackie Linh Consulting  
JMGemini Consulting  
Key Talent Consulting  
Krankenpflege Beinert  
Kurt Mosen Coaching Beratung Training  
Madael Consult  
Malaysian Life Reinsurance Group Bhd  
Mindcarter  
MLB Industrial Services GmbH  
MMA Global  
NAM QUOC GROUP  
Nanofilm Technologies International Limited  
Newlead  
Ngọc Thảo Trinh  
Nu-Insight Consulting  
OD Consulting  
ODIS Inc.  
OD-Tools  
OHRC  
Omotenashi  
OMRON  
Orion China  
OTA  
Ovid Assessment Central  
Paramount Corporation Berhad  
Paramount Property Division  
PARSA Haar- und Modeartikel GmbH  
PASONA Thailand  
PAWLIK ASIA  
PE Holdings  
Phi&P Consulting  
PKL International  
Projektforum Rhein Ruhr GmbH  
Qualipac  
QUEST ACADEMY  
ROSSI  
Savage Macbeth  
Schmezer Consulting  
Secret Escapes

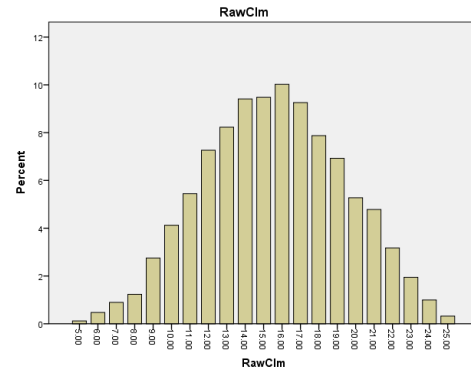
## Annex 2. Frequency distribution charts

### Trait raw scores, English dataset (n=5918)

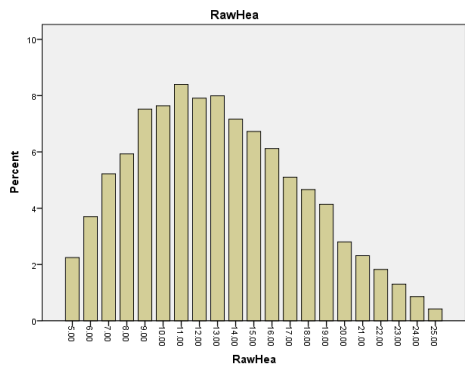
#### Emotional Control



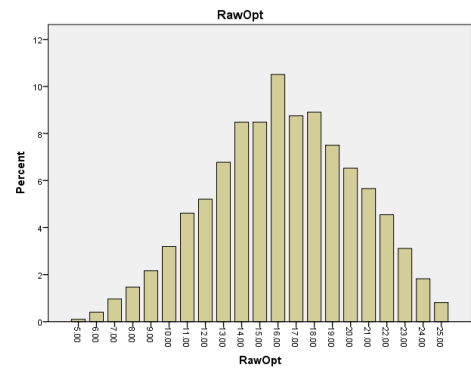
#### Calmness



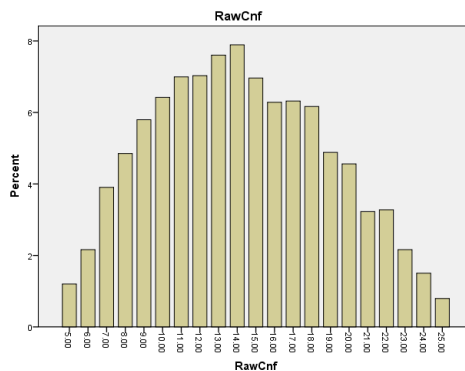
#### Health Mgmt



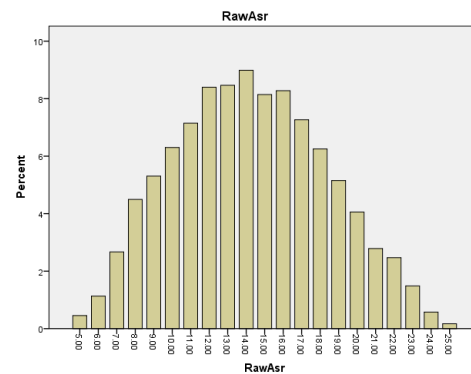
#### Optimism



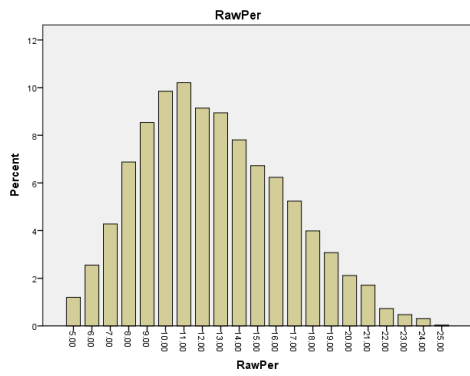
#### Confidence



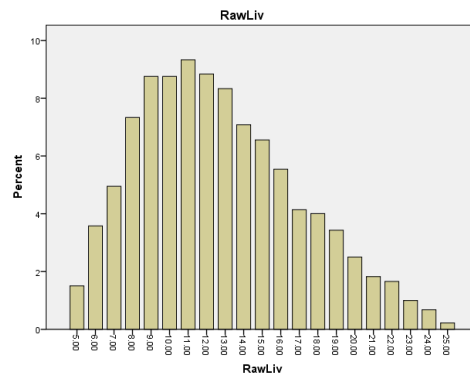
#### Assertiveness



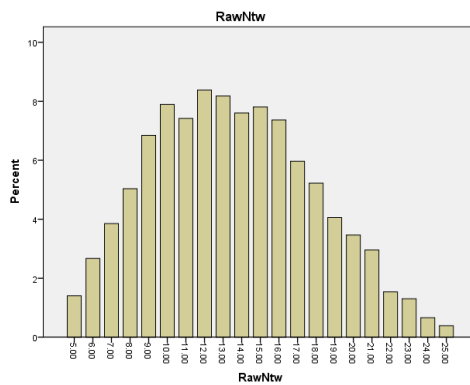
### Persuasiveness



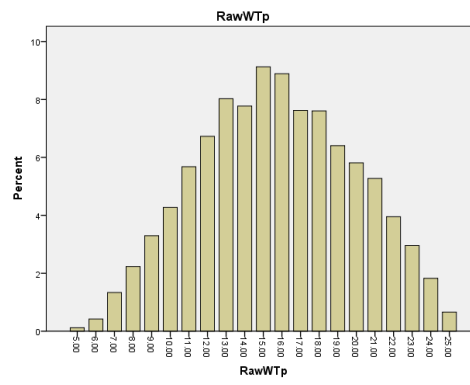
### Liveliness



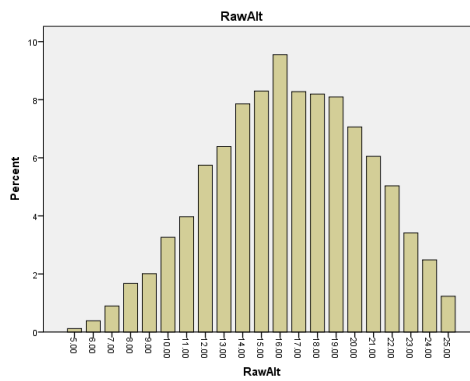
### Networking



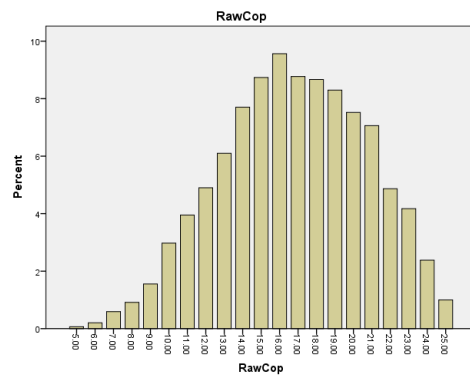
### Work Tempo



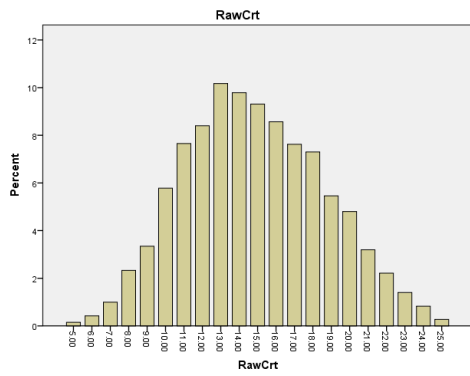
### Altruism



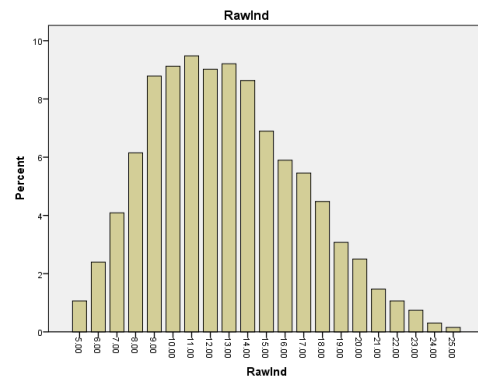
### Cooperation



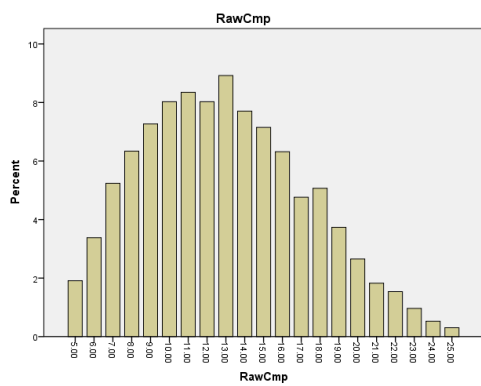
### Criticalness



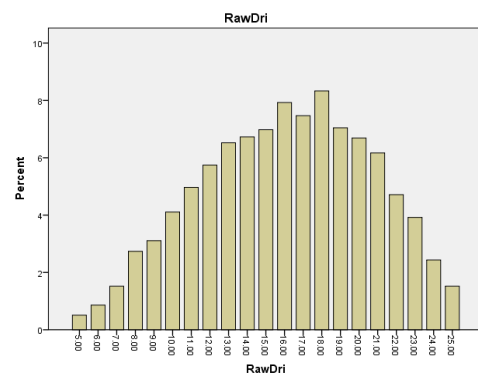
### Individualism



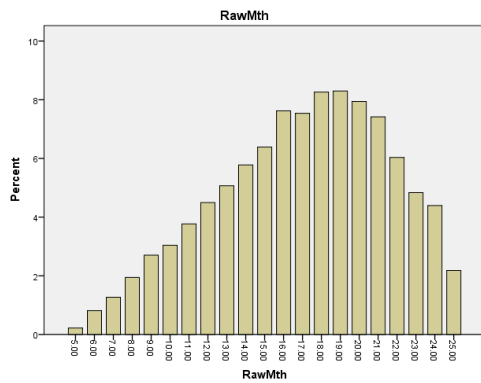
### Competitiveness



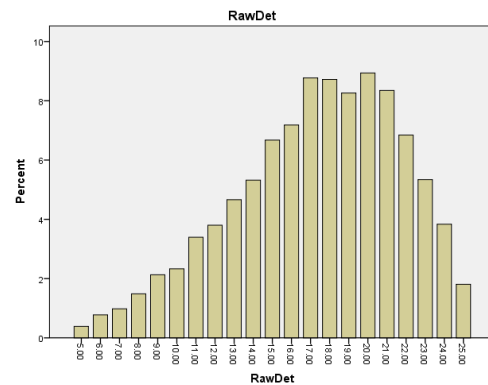
### Drive



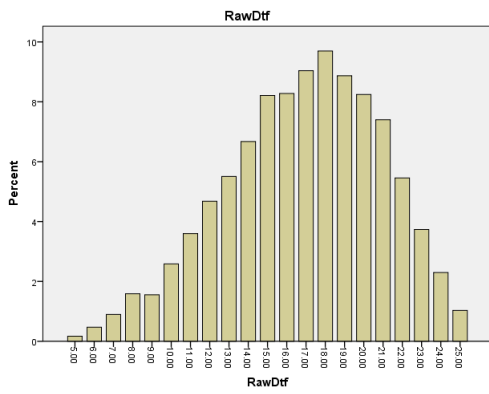
### Methodicalness



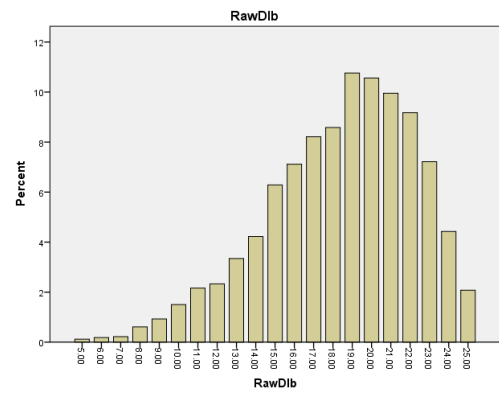
### Detail-concern



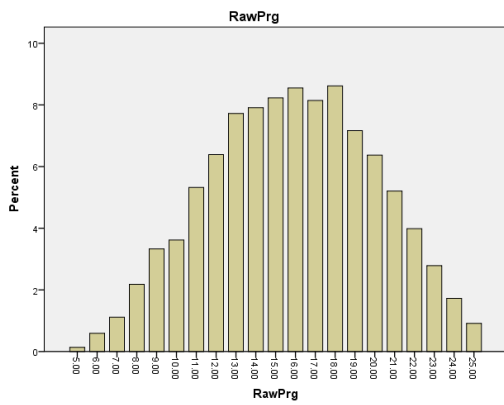
### Dutifulness



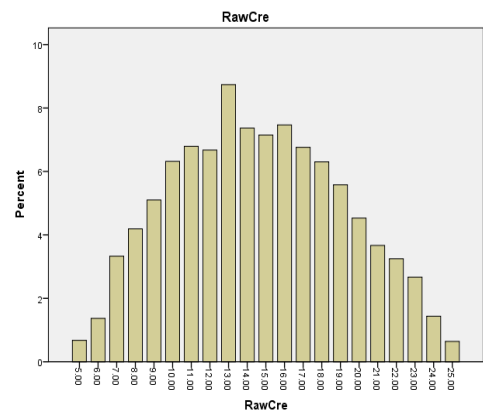
### Deliberation



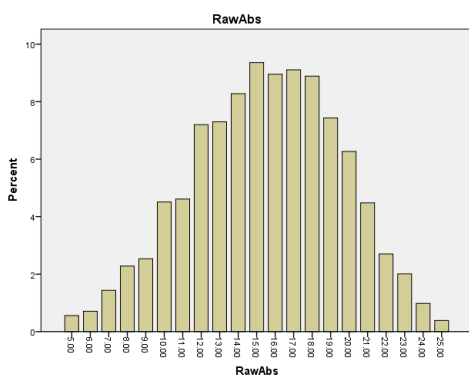
### Progressiveness



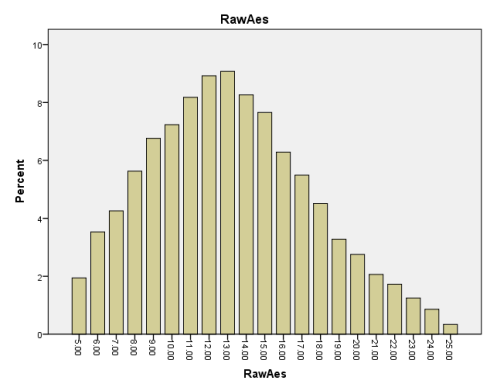
### Creativity



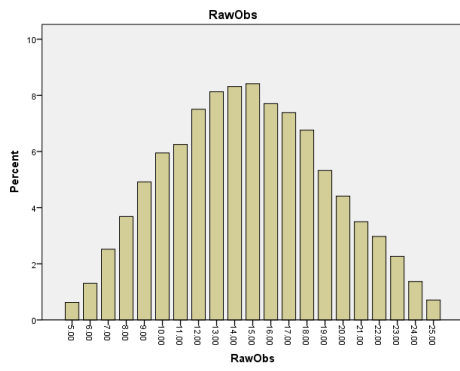
### Abstractedness



### Aesthetic Sense

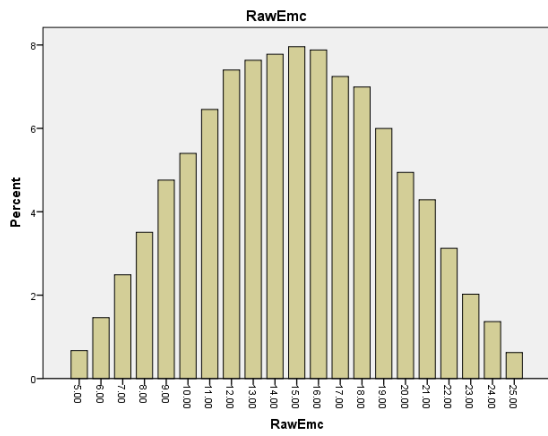


# Observation

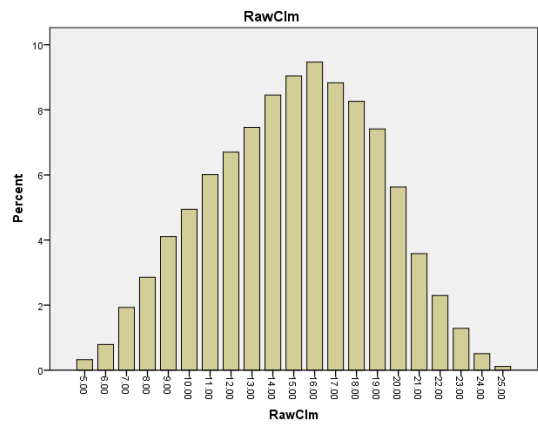


# Trait raw scores, Chinese dataset (n=16702)

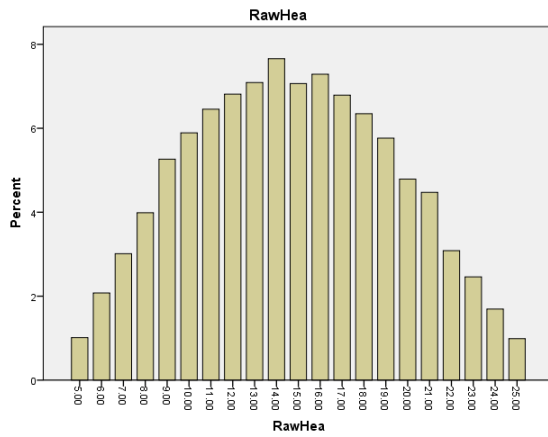
## Emotional Control



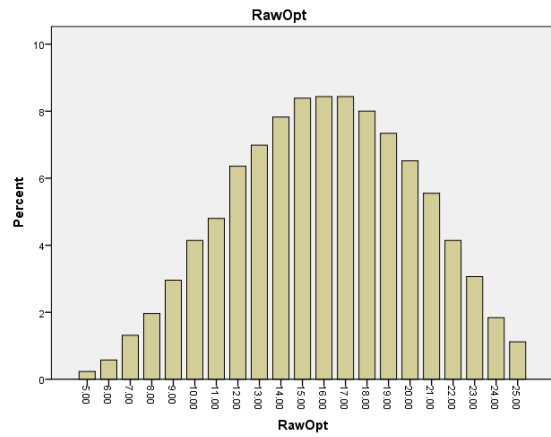
## Calmness



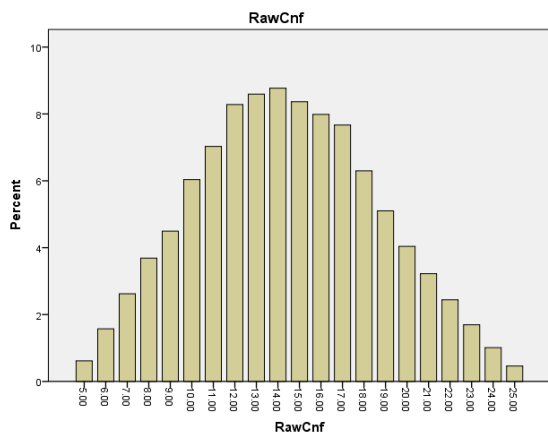
## Health Mgmt



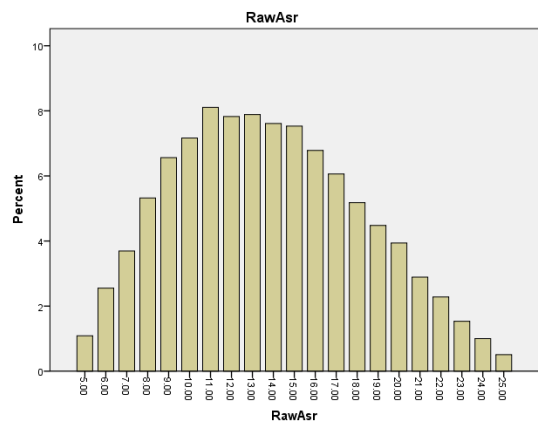
## Optimism



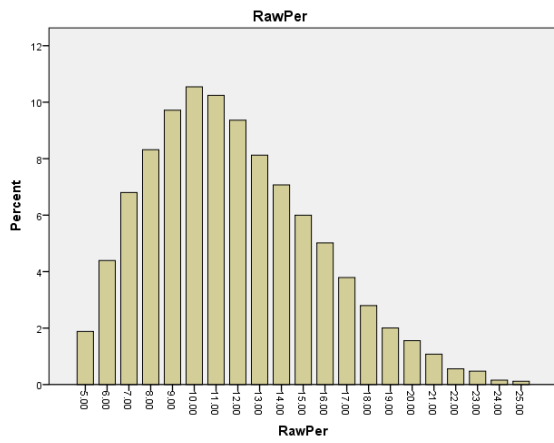
## Confidence



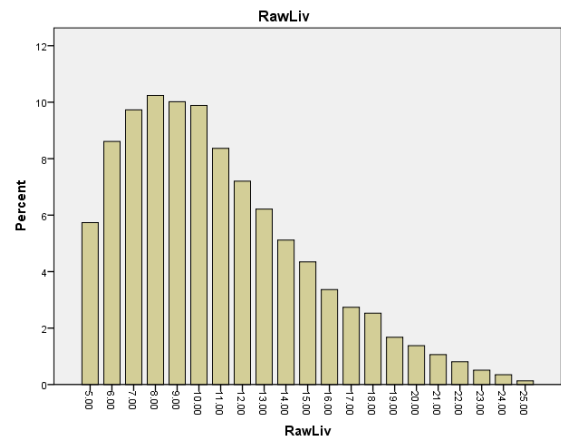
## Assertiveness



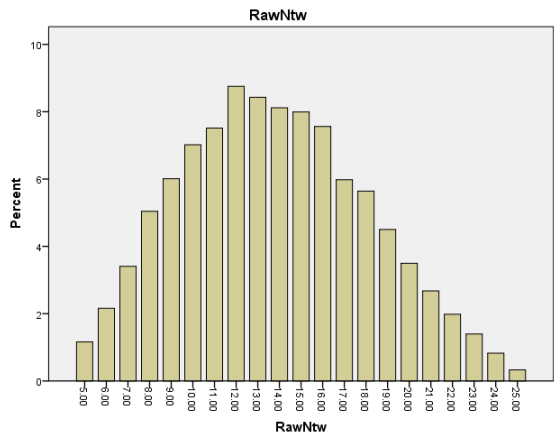
### Persuasiveness



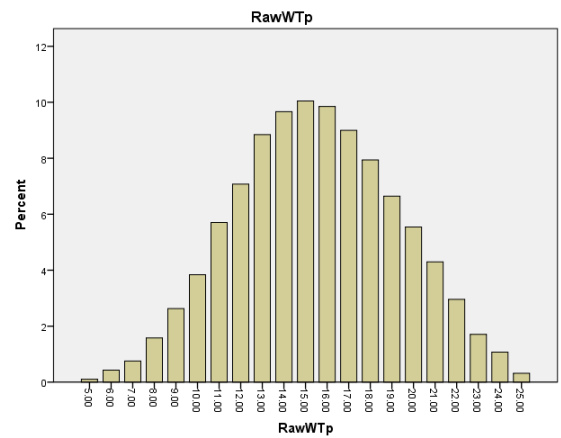
### Liveliness



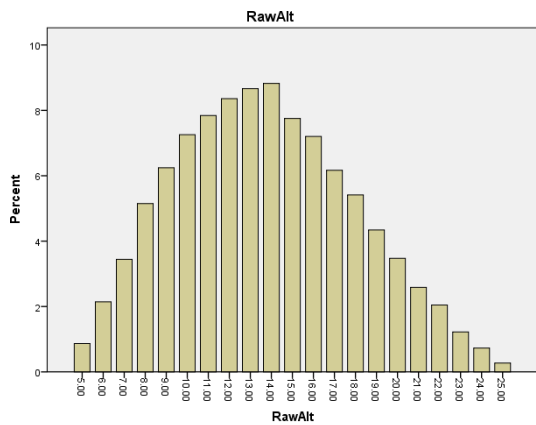
### Networking



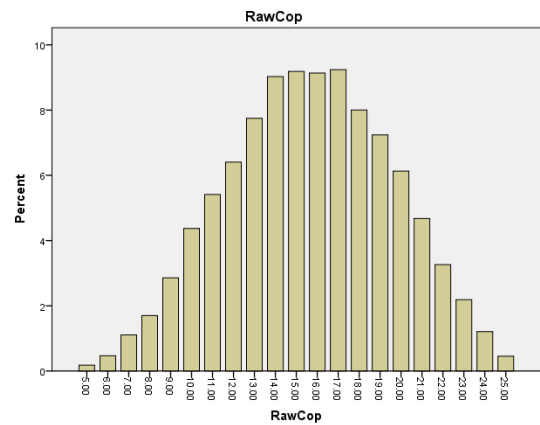
### Work Tempo



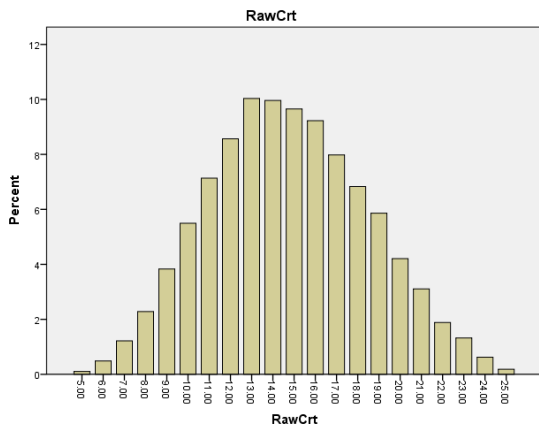
### Altruism



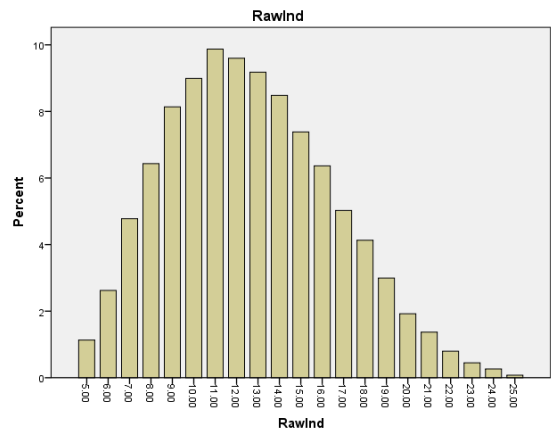
### Cooperation



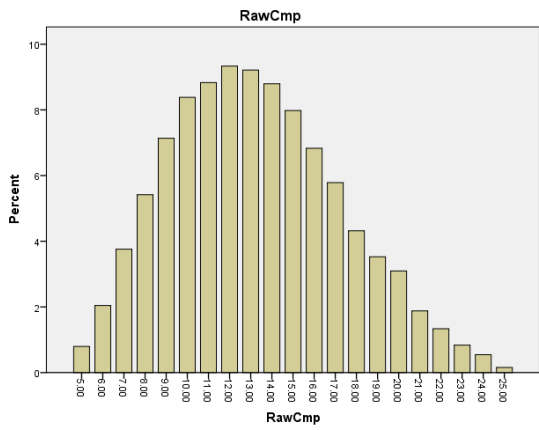
### Criticalness



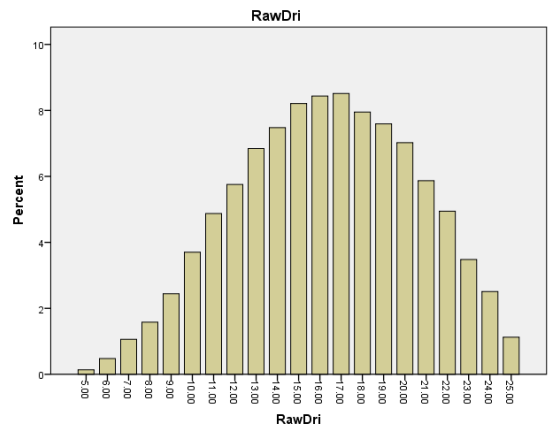
### Individualism



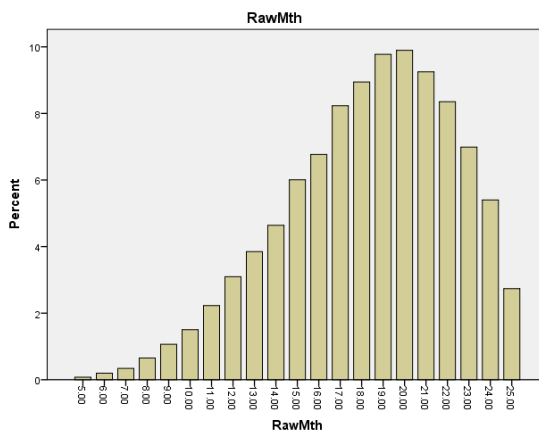
### Competitiveness



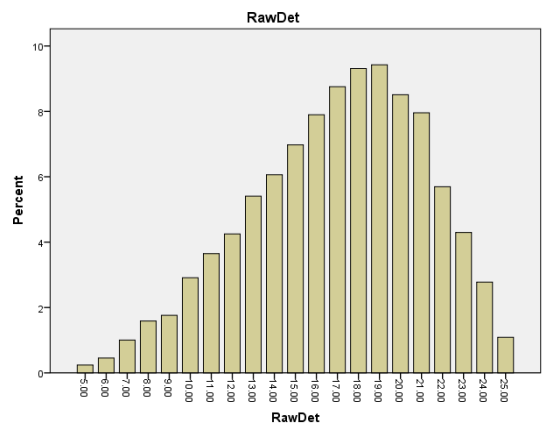
### Drive



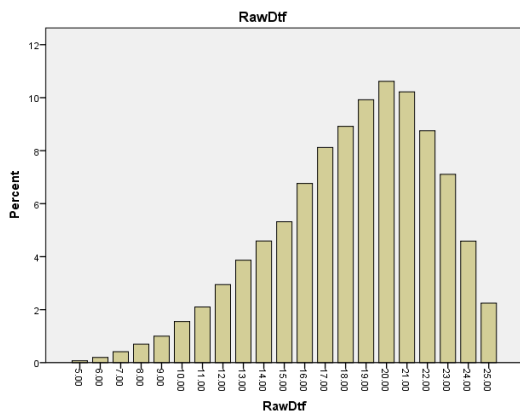
### Methodicalness



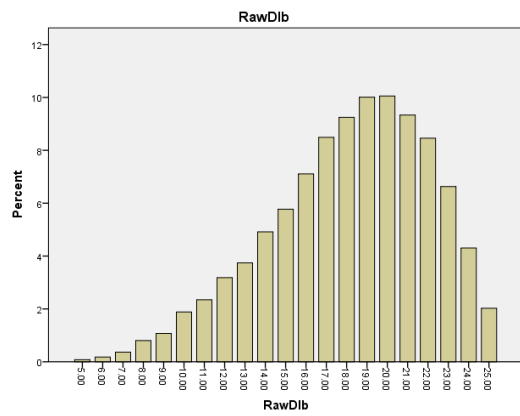
### Detail-concern



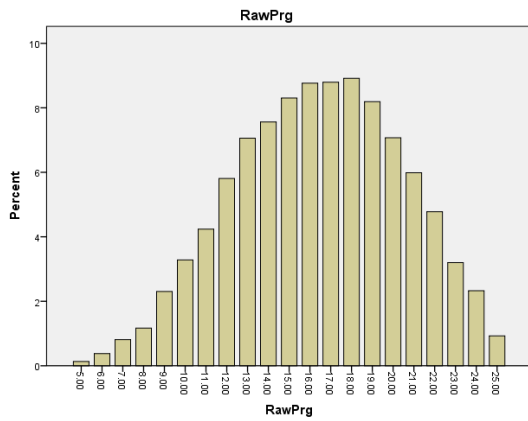
### Dutifulness



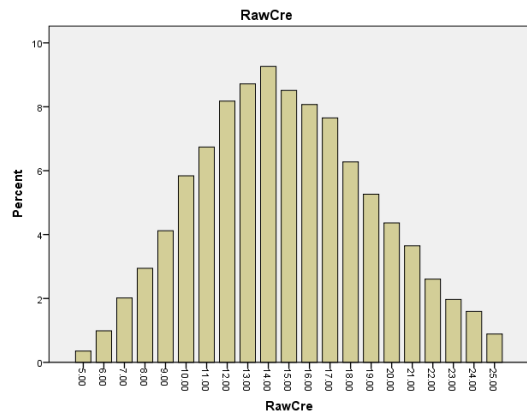
### Deliberation



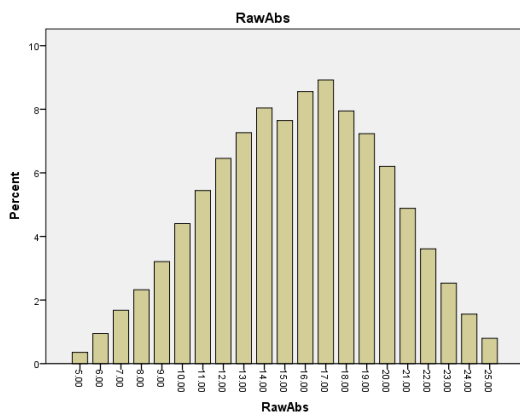
### Progressiveness



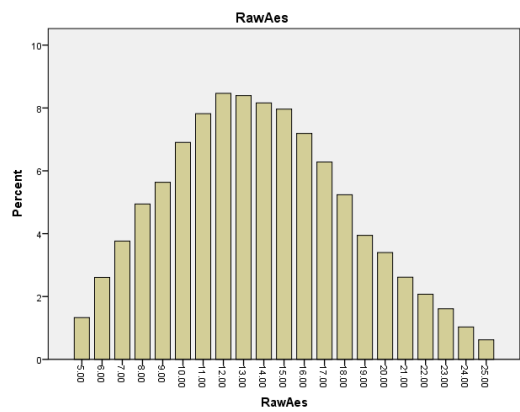
### Creativity



### Abstractedness



### Aesthetic Sense



# Observation

