



MAP Essence

Technical Manual

Version 3.0

HELENE HOPPE REVALD
Director of Psychometrics, Occupational Psychologist

RALUCA IOANA CHIFOR
Psychometrician, Occupational Psychologist

DENNIS HVASS
Psychometrician, Occupational Psychologist

Copyright © 2026 Assessio International AB
Unauthorized copying is strictly prohibited! All reproduction, complete or partial, of the content in this manual without the permission of Assessio International AB is prohibited in accordance with the Swedish Act (1960:729) on Copyright in Literary and Artistic Works. The prohibition regards all forms of duplication and all forms of media, such as printing, copying, digitalization, tape-recording etc.

Table of Contents

1. Introduction & theoretical background	4
2. Scale definitions & interpretations	5
Extraversion (EX)	5
Agreeableness (AG).....	5
Conscientiousness (CO)	6
Emotional Stability (ES).....	6
Openness (OP).....	7
3. Screening lenses	8
Choosing the right lens.....	8
Screening for Integrity	8
Recommended use	10
Ideal candidate profile	10
Screening for Service	10
Recommended use	11
Ideal candidate profile	12
Combining lenses	12
Using trait scores to qualify hiring decisions	12
Construction of screening lenses	13
Validity of screening lenses	13
Mitigating bias and adverse impact	14
Recommended cutoffs to prevent adverse impact	15
4. Instructions for use.....	16
Areas of use	16
Administration and scoring.....	16
Standardized information and instructions	16
Additional information.....	17
Requirements and conditions of testing	17
Environment	17
Target group	18
Computer skills.....	18
Readability.....	18
Other impairments.....	18
Feedback options	19
5. Scale construction.....	20
Trait models	20



Psychometric criteria	21
6. Validity	22
Face validity	22
Content validity	23
Construct validity	24
Consistency.....	24
Unidimensionality	24
Local independence.....	24
Item invariance.....	24
Scale characteristics.....	24
Group differences.....	26
Exploratory factor analysis	27
Convergent validity	29
Criterion validity	29
Validity estimates from a MAP study.....	29
7. Reliability.....	31
8. Standardization.....	32
Score calculation	32
Norm group.....	33
Update.....	33
Sample size.....	33
Composition	34
Group differences & Adverse Impact.....	35
9. Translations and adaptations	37
General procedure for translations and adaptations.....	37
Translational protocol.....	37
Instructions for translators	38
References.....	39



1. Introduction & theoretical background

MAP Essence is a quick tool for personality screening. It was derived from MAP, which is Assessio's signature personality assessment. Although they share the same theoretical foundation, MAP and MAP Essence differ in both length and depth, with MAP providing a much more nuanced description of personality. However, many organizations with high-volume recruitments are faced with a challenge: Screening applicants faster without compromising on quality. We created MAP Essence (MAP-E) to address this need. MAP-E is short, fast, and relies on state-of-the-art science to uncover the core of someone's personality.

Personality drives human behavior, both in private and in the workplace. It shapes a person's strengths and weaknesses, needs and challenges. This makes personality an important screening criterion, no matter what job you are hiring for. It also helps candidates to know how they come across, what their strongest competencies are, and in which jobs they have better chances to thrive.

A crucial step in developing any test is choosing the best theoretical framework for it. On the market, you can find assessments based on different theories, some of them more robust than others. Within the scientific community, most researchers agree that the Five-Factor Model (FFM) is today's best and most robust model for measuring personality (Costa & McCrae, 2008). Regardless of why you want to assess personality (e.g., for personnel selection, team development, counselling, etc.), the FFM provides you with an accurate (and relatively simple) framework. According to the FFM, almost every human behavior (whether work-related or not) is somewhat shaped by five personality dimensions: Extraversion (EX), Agreeableness (AG), Conscientiousness (CO), Emotional Stability (ES), and Openness to Experience (OP). Each of those dimensions are made of several underlying "facets". Different researchers may argue for a different number of facets or give them slightly different names. In MAP, traits are divided into five facets each. With MAP Essence, however, the focus was less on measuring all aspects of personality, and more on providing a quick and reliable snapshot of somebody's "essence". Therefore, we only selected those facets that make up the "core" of each dimension, and which are most predictive of performance.

In this manual, we explain what MAP Essence communicates about a person and how to use it to make better informed personnel decisions. We describe what performance criteria you can predict using MAP Essence. In addition, we provide information on how MAP Essence was constructed, how we standardize the scores to ensure fair comparisons between candidates, and how the instrument performs from a psychometric perspective.

In addition to providing a snapshot of personality, MAP Essence is used with a set of screening lenses for Integrity and Service aimed at predicting performance in service jobs and operational roles. These are described in detail in a separate chapter.



2. Scale definitions & interpretations

This chapter provides information on the five personality dimensions measured by MAP Essence. We describe what being “low” or “high” on a specific dimension means provide some example items.

In principle, each dimension is a spectrum between two opposing tendencies. Therefore, the closer somebody scores to one end of the spectrum, the more they express that tendency. This should always be interpreted in relation to others. “Extreme” scores mean the person has a stronger tendency to do one thing or another compared to other people. These tendencies are more distinct and noticeable, because they stand out from the norm, but they are not necessarily extreme in the sense that they come out as extreme. When someone has an average score, this means they don’t have a distinct tendency and are more likely to blend in with the majority.

Extraversion (EX)

Extraversion measures sociability and the desire for social connections, how comfortable someone is with being the center of attention, and their tendency to express positive emotions.

Individuals high in Extraversion often want to be surrounded by other people. They thrive in situations that are characterized by a high pace, and they enjoy being the center of others’ attention, often taking the role of leaders in different groups. Some of their characteristics are talkative, enthusiastic, lively, social, easy-going, and happy.

Low scores on the Extraversion scale imply a more introverted attitude towards other people and the social environment. Introverted people are often perceived as less social and more reserved, since they more often prefer and/or need seclusion or solitude. Social engagement tires them out, while quiet environments give them strength and energy. An introverted person often prefers to work alone or in smaller intimate groups and looks for projects that require collaboration with only a few other people and on a temporary basis. Low scorers on extraversion does not imply that someone cannot collaborate or engage with others, but they are more likely to shy away from those situations or tire from them.

Example of an item in the Extraversion scale: *I’m often the center of a group.*

Agreeableness (AG)

Agreeableness assesses how a person interacts socially – their level of trust, their inclination to assist others, how much they take other people’s feelings or preferences into account to avoid conflicts.

People with high scores on Agreeableness have a high level of trust in other people. They are helpful, caring, attentive, and care about how others think and feel. They adjust their own behavior and communication to take others’ feelings into account. They want and like to offer their support to others.

People with low scores on Agreeableness are more distant, skeptical, and straightforward in their interactions, and often adopt a more critical approach towards others. They don’t always attach great importance to the feelings, wishes, or preferences of others, and rarely feel the need to adapt their own behavior to please them or to avoid conflict. This does not mean that low scorers cannot come across as kind or friendly; however, they are less likely to do so when in a conflict or when their priorities require them to focus differently. Likewise, a high scorer can set boundaries or appear less accommodating, if the situation requires it, but it may demand more of them.

Example of an item in the Agreeableness scale: *I don’t like to tell the truth if it risks hurting someone.*



Conscientiousness (CO)

Conscientiousness assesses how a person approaches tasks – their dedication to ensuring structure, persistence, thought-out decisions, and alignment with quality standards.

Individuals with high scores are conscientious and meticulous. They place high demands on both themselves and others. They work methodically, systematically, and in a structured manner, even when the work is monotonous, boring, or requires perseverance. They are orderly, thorough, and organized, which means that they are often perceived as trustworthy members of the organization who think before making decisions.

Low scores indicate that a person tends to approach their tasks and commitments in a more relaxed and easy-going way. They often perceive rules, planning, and preparation as restrictive. They would rather follow the spur of the moment, taking each day as it comes. Individuals with low scores on this scale are more relaxed towards obligations and not as bothered by deadlines and commitments, or by not adhering to predetermined plans or procedures. Low scores do not mean that a person cannot work in a structured way or adhere to any rules or standards; however, low scorers are more likely to question procedures and are more likely to feel disengaged when expected to follow procedures strictly. They are also less likely to create structures themselves and more likely to give into feelings of boredom or futility.

Example of an item in the Conscientiousness scale: *I sometimes put off boring tasks. (reversed)*

Emotional Stability (ES)

Emotional Stability measures the extent to which someone reacts emotionally - the tendency to be calm, self-controlled, and rational when faced with setbacks and challenges.

Individuals with high scores will mostly act calmly in stressful situations. They rarely seem affected by adversity and act in a composed way. One of the typical features of emotionally stable individuals is their even temper; they are rarely angry or irritated.

Individuals with low scores on this factor tend to be more temperamental, less composed, and more affected by stressful situations. It will often be more transparent to their surroundings, how they feel and what they think in a given situation. This does not mean that low scorers cannot function in busy environments; however, they are more likely to react more strongly and more visibly in those kinds of situations.

Example of an item in the Emotional Stability scale: *My mood sometimes fluctuates throughout the day. (reversed)*



Openness (OP)

Openness captures a person's overall openness to change, exploration, and innovation – their level of curiosity and preference to gain experience by trying out new things, researching, or engaging with different perspectives.

High scorers on Openness are curious, creative, and likely to come up with new, unconventional ideas. They often have a strong desire to try out what is new and are constantly on the lookout for new experiences. These individuals seek or create situations that satisfy their need for intellectual stimulation.

People with low scores on Openness are practical and down-to-earth, focusing on what is happening here and now. They neither have a great need for, nor do they seek out new experiences. People with low scores usually prefer what is already known and tested over what is uncertain. Therefore, they might be perceived as less flexible. Low scores do not imply that a person opposes change or is incapable of adapting. Low scorers rarely seek out change for the fun or stimulating pleasure of it but will typically engage in it when there is good reason to and a strong argument behind it, without necessarily feeling overly engaged by it.

Example of an item in the Openness scale: *I'm good at imagining things that others might find a bit unrealistic.*



3. Screening lenses

MAP Essence is administered through a set of predefined screening lenses. In this chapter, we outline when and how to use each lens, how they are constructed, and their empirical basis, including the criteria (outcomes) they were designed to predict.

Choosing the right lens

In this section, we describe each screening lens in detail and give recommendations on how to best use them. Table 3.1 below provides a summary of this information.

Table 3.1. Brief descriptions of MAP-E screening lenses and recommendations for use.

Screening lens	Recommended use		Composition
	<i>Job profile</i>	<i>Industry</i>	
Integrity	Following pre-established procedures and demonstrating excellent work ethics	Compliance, security, government, law enforcement	CO, ES, Risk-Taking (<i>reversed</i>)
Service	Working very closely with customers to provide services tailored to their needs	Customer service/support, HORECA, retail	AG, ES, EX, CO

Note. In the last column, the traits are listed in order of their importance.

The following matrix can be used to decide on the proper lens to be used depending on the need for reliability (or dependability) and the level of customer contact in the job:

Customer contact	Need for reliability	
	Low	High
Low	-	Integrity
High	Service	Both

Screening for Integrity

In research, Integrity is defined as complying with rules, respecting the code of conduct, fostering transparency, and working according to the highest ethical and quality standards. Sometimes, literature describes integrity in opposition to counterproductive work behavior (CWB), as these typically have a moderate, negative correlation. Counterproductive work behavior (CWB) refers to intentional employee behavior that harms the organization and its stakeholders (e.g., theft, sabotage, mishandling information) or its other employees (e.g., gossiping, bullying, sexual harassment) (Dalal, 2005).



People with high scores on integrity are more likely to:

- Read, remember, and follow the organization's code of conduct;
- Comply with legal requirements and security protocols;
- Impose high standards on themselves and others;
- Thoroughly document their work;
- Signal when a mistake was made and take responsibility for it;

People with low scores on integrity are more likely to:

- Bypass rules or protocols;
- Act rash;
- Spread fear or gossip, and take out their frustrations on others;
- Gatekeep or mishandle information or documentation;
- Compromise on key quality requirements;
- Overestimate their own competence or make unrealistic promises.

Previous research has shown moderate to strong relationships between integrity and personality traits. The FFM dimensions most strongly related to integrity are Conscientiousness, Agreeableness, and Emotional Stability (Marcus et al., 2007; Catano et al., 2018). People low on integrity are also more likely to engage in CWB (Wing-Man Lau et al., 2023). CWB can be directed towards individuals such as colleagues or managers (CWB-I) or harm the organization as a whole (CWB-O). While CWB-I has been consistently linked to low Agreeableness, CWB-O is more closely predicted by low Conscientiousness. Both categories are relevant to investigate, though CWB-O tends to have more serious and long-lasting consequences for the business. Therefore, we chose to center our measure of Integrity around CWB-O rather than CWB-I. In addition, the opposite of CWB-I, which is helpful, supportive, and proactive behavior, is already captured quite well by the Service lens. We thus excluded AG from the Integrity lens but weighed it more heavily in the Service lens. To further accentuate the difference between Service and Integrity, we included Risk-taking as an extra predictor in the latter. In MAP, Risk-taking was part of EX but not included in the "sociability" factor that is the core of this trait.

Below, we further explain what traits are included in the Integrity lens and why we chose them:

- **Conscientiousness:** Research points to Conscientiousness as the strongest associated trait. This is because conscientious people are keen on following rules and want to do everything right. Therefore, this trait is weighed more heavily in the Integrity score. This relationship is further enforced by the operationalization of conscientiousness in MAP-Essence that focuses more heavily on diligence, structure, quality assurance, and self-discipline rather than determination in the form of personal ambitions and self-efficacy.
- **Emotional Stability:** Furthermore, Emotional Stability is weighed in because it is crucial to impulse control. If someone struggles to self-regulate, they are more likely to lash out when they feel stressed, angry, or frustrated. However, those high in Emotional Stability can inhibit their emotions when necessary, and they think twice before doing something they might regret later. This enforces integrity in the way that it will ensure a larger degree of composure, impulse control and dependability.
- **Risk-Taking:** This characteristic is typically considered a facet of Extraversion. While extraverts are mostly known for their sociable nature, research shows they are also more prone to taking risks. At work, someone with a high score on Risk-Taking may take rash decisions without pausing to consider how those will affect others, accidentally mishandle information, or pay less attention to quality standards and following protocols. These work-specific behaviors are also what we aim to predict. Research has also indicated a reversed pattern for risk-taking compared to other Extraversion facets by showing positive correlations with CWB and negative correlations with Task Performance (TP). Therefore, we excluded the element of risk from EX and reserved the score to predict integrity.



Recommended use

While some level of integrity is necessary in all jobs, this lens is especially relevant when you want to hire for positions in compliance, security, government, or law enforcement, where following pre-established procedures and demonstrating excellent work ethics is crucial.

Some example positions for where Integrity will be extremely valuable as a ranking method could include:

- Warehouse positions
- Inventory positions
- Production positions
- Security positions
- Construction positions
- Maintenance positions
- Machine operator positions
- Housekeeping positions
- Kitchen positions

Ideal candidate profile

The Integrity lens helps you find people who follow the rules thoroughly, cope well with pressure, and refrain from engaging in risky behavior. They make reliable teammates and considerate managers who will likely prioritize quality over speed and carefully weigh the pros and cons of every decision.

Screening for Service

This lens was created to assess service orientation (also referred to as “customer orientation” in research). Service orientation represents a set of characteristics that enables someone to support customers in the best way.

People with high service orientation are more likely to:

- Address issues proactively;
- Stay professional when dealing with so-called “difficult” clients (customer incivility);
- Respond promptly to inquiries;
- Have enough energy to sustain multiple or prolonged interactions;
- Follow communication protocols;
- De-escalate conflicts in a smooth and rapid manner.

People with low service-orientation are more likely to:

- Postpone issues which don't affect them directly;
- Have their patience challenged by so-called “difficult” clients;
- Respond harshly when provoked;
- Feel drained by repeated or prolonged interactions;
- Stick to what they believe and reject compromises.

Research suggests that Agreeableness, Emotional Stability, Extraversion, and Conscientiousness all have links to greater service orientation, as well as performance and satisfaction in service jobs, but their relative importance depends on the work context (Doan et al., 2021; Hurtz et al. 2000; Lounsbury et al., 2012; Motowidlo et al., 2008).



Below, we further explain what traits are included in the Service lens and why we chose them:

- **Agreeableness:** Most studies seem to agree on one thing: treating the customer with respect and doing your absolute best to help them is quintessential to most service jobs. Because of those qualities, we give Agreeableness the highest weight. Adding Agreeableness highlights the need for being accommodating and helpful in nature while avoiding opinionated or too direct communication in client interactions.
- **Emotional Stability:** Dealing with difficult clients is notoriously the worst part of working in customer service. Sometimes, clients react angrily, speak rudely to the employee (or even threaten them), refuse to cooperate, or insist on impossible requests. In such situations, the employee must keep their composure and try to defuse the conflict as smoothly as possible. This is significantly easier for people high in Emotional Stability. Those low on the trait are known to have quick tempers, which makes them more likely to snap back at the customer. Adding Emotional Stability in combination with Agreeableness to the Service lens enforces the need for a calm, collected and accommodating approach to clients even in quite aversive or stressful situations.
- **Extraversion:** In jobs such as waiter, shop assistant, hotel receptionist, or flight attendant, interactions are short and happen face-to-face. The employee must handle multiple customers at a time and often pivots between different tasks, moving around in a physical sense. Therefore, Extraversion is a useful trait to have, as it makes someone sociable and energetic. Studies correlate Extraversion with performance and satisfaction in industries such as HORECA or retail.
- **Conscientiousness:** When the employee engages with customers remotely (e.g., from behind a screen or via phone call), they usually service one customer at a time and their interactions are longer. At times, this category of service jobs is more technical, such as in banking or IT support. While Extraversion might not be as crucial here, Conscientiousness would help the person conform to protocols. It is especially important that the employee does not (accidentally) mislead the customer by giving them inaccurate or incomplete information, nor share sensitive data when they shouldn't. Furthermore, in direct contact with clients, orderliness and professionalism in form of structure is often perceived as more professional, why Conscientiousness in the way it is operationalized in MAP-E also have weight in these kinds of jobs.

Recommended use

Their kind and lively nature makes service-oriented people a great addition to any team. However, this lens is mostly dedicated to people who work very closely with customers. We recommend choosing this lens when you hire for positions in customer service/support, HORECA, or retail.

Some example positions for where Service will be extremely valuable as a ranking method could include:

- Customer service positions
- Front desk positions
- Support positions
- Call center positions
- Counter service positions
- Hospitality positions
- Guest services positions
- Brand ambassador positions



Ideal candidate profile

The Service lens helps you find people who are polite and will respond in a professional manner even when facing rude comments or unreasonable demands. Their social battery almost never runs out, allowing them to have multiple interactions throughout the day without feeling emotionally drained. If communication protocols do exist, those people will make sure to follow them.

Combining lenses

In principle, we recommend using a single screening lens, prioritizing what aligns best with the job description. Users who struggle to prioritize may also use the lenses in combination. However, please be aware that doing so essentially means prioritizing what the two lenses have in common - Conscientiousness and Emotional Stability. Those high on both traits have the best chance of being selected. Thus, we recommend combining lenses when being thorough and maintaining composure are the most important attributes for the job while some level of service orientation (i.e., higher scores on Agreeableness and Extraversion) is needed.

If the job requirements do not align with either of the lenses provided based on MAP-E, consider if MAP (Assessio's extended personality test) might be a better option for your recruitment. While longer, MAP provides a more nuanced assessment of personality and can therefore predict a wider range of criteria.

To get an idea of the relative impact of each trait on each lens, a set of correlations are briefly summarized in Table 3.2 with the strongest correlations highlighted in bold.

Table 3.2. Correlations between MAP-E traits and screening lenses.

Trait	Integrity	Service	Combined
EX	-.01	.36	.19
AG	.05	.62	.37
CO	.83	.47	.73
ES	.62	.59	.68
OP	-.14	.12	-.01
<i>Risk</i>	-.51	-.10	-.34

Using trait scores to qualify hiring decisions

After applying one or both of the lenses to screen and rank candidates, a hiring decision is to be made on the final shortlist of candidates. Besides other relevant criteria, it can be useful to use the trait scores from the assessment to qualify this decision. As Agreeableness, Conscientiousness, and Emotional Stability are well captured by each of the lenses, this decision is likely to center around the traits of Extraversion and Openness.

Although MAP-E is usually applied to lower-level service or strictly operational jobs, even these jobs differ with respect to the level of autonomy or "freedom" to adjust, reinvent, or explore alternative procedures. For jobs with some level of autonomy, moderate to high scores on Openness would be ideal, whereas low scores would be preferred for jobs requiring adherence to a strict protocol. Similarly, moderate to high levels of Extraversion would be preferred for jobs involving frequent social interaction and vice versa for jobs of a more introvert nature.



These recommendations are summarized in the matrix below.

Social interaction	Level of autonomy	
	Low	High
Infrequent	Low-Moderate EX & OP	Low-Moderate EX Moderate-High OP
Frequent	Moderate-High EX Low-Moderate OP	Moderate-High EX & OP

Construction of screening lenses

In designing the screening lenses, the selection of traits was based on three core principles:

- *The "rationality" principle:* The combination of traits should make sense. Considering how the criteria are defined, it should be easy to argue and understand why these personality characteristics are relevant based on examples from real work situations, emphasizing transparency for clients as well as candidates when needed.
- *The "research" principle:* The combination of traits should align with research on what drives performance. We should be able to find peer-reviewed articles by subject matter experts or meta-studies that explain how those behavioral indicators help people thrive and perform in, for instance, service jobs, or jobs that require a high level of integrity.
- *The "data" principle:* Empirical research must show sufficiently strong correlations between the traits and the chosen (or similar) criteria. Those correlations should also be replicated on our data, meaning they should be in the same direction and of comparable strength.

When studies identified different predictors for the same criteria, we chose the ones mentioned most often or that had the strongest relationship with the criteria, and we avoided predictors that showed ambiguous relationships (e.g., positive in some situations and negative in others).

Each trait was assigned a weighting based on its importance. If either data or theory was inconclusive, we used equal weights. In addition, we accounted for adverse impact by ensuring that the final calculation was not discriminating against any genders or age groups. Please note that we report the order of traits for each lens rather than the exact weights, as these are considered intellectual property (IP) of Assessio.

Validity of screening lenses

Furthermore, we wanted to test how well the screening lenses themselves predict the same performance criteria based on the same data. For this analysis, we also included a measure of in-role or task performance (TP), which refers to fulfilling the objectives of the job (Pradhan & Jena, 2017). Our results indicate moderate to strong correlations as shown in Table 3.3.

As expected, OCB was best predicted by Service, whereas CWB and TP were equally well predicted by Integrity and Service. This is mainly due to the shared weightings of traits (especially CO and ES) across the two lenses and the fact that the CWB scale used in the current study had many items relating to interpersonal CWB or CWB-I (e.g., arguing, insulting, making fun of others) rather than CWB-O, i.e. CWB directed towards the organization, the latter of which is usually better predicted by CO (as opposed to AG), and for which the Integrity lens was purposefully optimized to predict.



Table 3.3. Correlations of screening lenses with CWB, OCB, and TP.

Lens	CWB		OCB		TP	
	Correlation (r)	Strength	Correlation (r)	Strength	Correlation (r)	Strength
Integrity	-.41*	Moderate	.37*	Moderate	.43*	Moderate
Service	-.44*	Moderate	.63*	Moderate	.41*	Moderate

Note. *p < .001. CWB = Counterproductive Work Behavior. OCB = Organizational Citizenship Behavior. TP = Task Performance

Mitigating bias and adverse impact

Although the scales in MAP Essence have been carefully designed to reflect real group differences rather than item bias (Differential Item Functioning), these differences can lead to the disfavoring of a demographic group when used for selection. This is commonly referred to as adverse impact.

To address this issue, we conducted simulations of expected adverse impact for both screening lenses using fixed cutoffs ranging from most lenient (50+) to most strict (90+) for gender and age shown below in Tables 3.4 and 3.5. These simulations assume that everyone above a fixed cutoff is hired and compare the proportion of hires between two demographic groups. If the selection ratio for the least represented group is less than 4/5th (80 %) of the selection ratio for the most represented group, this is considered evidence of potential adverse impact.

For Integrity the very strict cutoff of 90+ runs the risk of creating an adverse impact with respect to both gender and age and so should be used with extreme caution. Neither of the more lenient cutoffs are likely to produce any adverse impact, except for the cutoff of 80+ for gender. For Service, the more lenient cutoffs can be applied safely without any impact for gender and younger age groups but runs the risk of impacting the oldest age group (from 40 years and above) if using a fixed cutoff of 70+ or higher. The latter is largely because older individuals scored significantly lower on the traits with most impact on this lens and because the norm group was constructed to best suit younger applicants. Also, a slight impact was observed for gender for cutoffs 70+ and 80+ but not for 90+.

In conclusion, using the screening lenses for hiring decisions does not cause any adverse impact if strict cutoffs are avoided. In most, if not all, practical circumstances, a cutoff of 90 is overly strict, as it reflects only the top 2.3% of candidates. Also, we recommend applying flexible cutoffs by ranking candidates on the lens most important for the job (hence, a score of 60 or 70 is likely enough to be considered for hiring).

Table 3.4. Adverse impact ratios for gender (men/women) on MAP-E screening lenses.

Lens	Dif.	d	50+	60+	70+	80+	90+
Integrity	1.52	0.08	1.00	0.90	0.89	0.73	0.52
Service	1.78	0.09	0.98	0.92	0.76	0.78	0.96

Table 3.5. Adverse impact ratios for age (below/above 40 and 25) on MAP-E screening lenses.

Lens	Main dif.		Above/Below 40					Above/Below 25				
	Dif.	d	50+	60+	70+	80+	90+	50+	60+	70+	80+	90+
Integrity	0.97	0.05	0.99	0.92	0.91	0.95	0.96	0.99	0.99	0.82	0.97	0.80
Service	4.41	0.22	0.82	0.81	0.75	0.65	0.38	0.84	0.87	0.80	0.68	0.62



Recommended cutoffs to prevent adverse impact

Giving firm recommendations on how to prevent adverse impact is difficult, as the level of impact depends on a range of factors that vary across recruitment processes. In general, the level of adverse impact increases with:

- The magnitude of group differences (usually measured with Cohen's d);
- The presence of outliers in each group (unusual high scores within a demographic group;)
- Smaller applicant pools;
- Imbalance of demographic groups within the applicant pool (e.g., 8/10 women);
- Smaller selection ratios (e.g., selecting only the top 5% as opposed to the top 20%);
- Stricter cutoffs (especially when using fixed as opposed to relative cutoffs).

Table 3.6 lists the highest possible cutoffs that can be used for the different lenses to prevent adverse impact on gender and age. Please note that these cutoffs were modelled on a rather large data set constituting the norm group (N = 1 645) and should therefore be applied with large applicant pools (> 100). For smaller applicant pools, these cutoffs likely need to be more lenient to make sure there is no adverse impact. Also, the very lenient cutoffs proposed for the second age group comparison reflect the intended target group of younger candidates for MAP Essence.

To be certain that no adverse impact occurs in the selection process, we make the following additional recommendations:

- Use relative rather than fixed cutoffs (e.g., pick the highest scoring candidates even if their score is below the suggested cutoff).
- Make the final hiring decision by considering criteria other than the match score (such as individual trait scores, interviews, work samples, and reference checks if relevant).
- Monitor the level of adverse impact across recruitments to account for external factors that affect hiring decisions throughout the process, either before the assessment is completed (advertising, self-selection biases) or after (selection ratio, sample size, other criteria).

Table 3.6. Highest acceptable cutoffs for screening lenses to prevent adverse impact.

Demographic group	Highest acc. cutoff		Risk of favoring	
	Integrity	Service	Integrity	Service
Gender (men/women)	70+	60+ or 90+	Men	Men
Age (below/above 25)	90+	70+	-	16-24
Age (below/above 40)	90+	60+	-	16-39



4. Instructions for use

This chapter outlines best practice for administering MAP Essence. Following these guidelines helps ensure that all respondents complete MAP-E under standardized conditions, supporting fair, consistent, and comparable results. It also helps test administrators use the results in a professional, structured, and appropriate way, aligned with the intended purpose of the assessment and the context in which it is used.

Areas of use

MAP-E is designed to measure personality in line with the Five-Factor Model (FFM). It is primarily intended for use in selection contexts, especially screening and other high-volume recruitment processes, where an efficient, standardized, and scalable assessment solution is needed.

Administration and scoring

MAP Essence is administered remotely by sending the respondent a link to the online platform. This offers a flexible and accessible testing experience, allowing respondents to complete the assessment at home or at another suitable location. The administration format is therefore unsupervised. However, this is of limited practical importance, as the testing experience itself is standardized and identical for every respondent.

The test session begins with clear on-screen instructions, after which the respondent completes the assessment independently. Once the test is finished, automatic feedback is displayed on screen for the respondent. Inviting a respondent requires only the person's email address. Respondents are encouraged to answer all items during the same session, which means the assessment should ideally be completed in one sitting. However, the platform allows a respondent to pause and resume the session later if needed.

MAP-E consists of 55 items and takes approximately 10 minutes on average to complete. There is no time limit for completion. Respondents are advised to set aside 10–15 minutes for the assessment. After completion, scoring is handled automatically. Results are calculated and transformed into standardized scores, making administration efficient and minimizing manual handling for the user. Depending on the purpose of the assessment and the broader process in which it is used, the time required for interpretation and feedback may vary.

Standardized information and instructions

Before the assessment begins, respondents receive standardized on-screen information and instructions, including:

- MAP-E measures personality characteristics through statements describing different situations, behaviors, and preferences.
- The respondent's task is to read each statement and indicate the extent to which it applies to them.
- There are no "right" or "wrong" answers.
- Four response options are used: Disagree, Slightly disagree, Slightly agree, and Agree.
- Two practice items are included, and responses to these do not affect the results.
- The respondent should make sure they can complete the assessment without distractions or interruptions.
- There is no time limit.
- It is possible to go back and change responses at any time during the session.
- The respondent will answer 55 statements.
- The expected completion time is 10–15 minutes.



The instructions are designed to be self-explanatory, which means respondents can complete the assessment without assistance. All information required for completion is presented at the start of the test.

Additional information

In addition to the standardized instructions, there is other important information that should be provided before testing. While this information may vary depending on the specific context, best practice is for the test administrator to ensure that respondents know:

- The purpose of testing.
- What type of assessment is being used and why it is relevant in the current context.
- How the assessment will be administered and what is required to complete it (see the section below).
- Whether and how test results will be used and stored, by whom, for how long, and for what purpose. This applies not only to the test platform, but also to how the test administrator or organization handles the data.
- Whether feedback will be provided, when, in what format, and what it will include.
- That the respondent should contact the test administrator or support if any unexpected issues or questions arise during the test session.
- How to contact the test administrator.

Providing this information to all respondents before testing is considered best practice. In many cases, the test administrator may also wish to share further information such as the next steps in a recruitment or development process. Since this type of communication may differ across situations, it is not always delivered in a fully standardized format. If MAP-E is delivered through an ATS, it remains the responsibility of the test administrator to ensure that the respondent receives this information in an appropriate way.

More details about the rights and obligations of test distributors, test administrators, and respondents can be found in international testing guidelines, such as those provided by the International Test Commission, EFPA, and ISO, and may also be available through national psychological associations. Assessio makes sure to strictly follow those standards.

Requirements and conditions of testing

To support high-quality administration and meaningful results, there are several important conditions that test administrators need to consider before administering MAP-E.

Environment

MAP-E should be completed in a non-distracting environment. Public settings, such as internet cafés or public transportation, are not suitable for testing. The assessment can be completed on any device the respondent is sufficiently familiar with, including a computer, tablet, smartphone, or similar device. A stable internet connection throughout the full session is also needed in order to provide the best possible candidate experience.

If the test is used for development purposes, such as a team building workshop, and is administered on-site at the employer's location, then the test administrator is responsible for ensuring a comfortable environment. They should be well acquainted with the assessment from a theoretical, psychometric, and practical perspective, and should be able to communicate in a calm, competent, and reassuring way. It is important to remember that a testing situation may be entirely new to the respondent and therefore may put them in a vulnerable position. Respondents should therefore be given the opportunity to ask questions before testing, and those questions should be answered truthfully and accurately.



Target group

MAP-E is intended for respondents aged 16 years and older. Data used for standardization, norms, and psychometric evaluation are therefore collected from individuals within this age range.

Please note that testing minors, that is, individuals under the age of 18, may in many legal jurisdictions require consent from a legal guardian. A test administrator may manage the consent process, but it is their responsibility to ensure that this is handled correctly and in accordance with relevant legislation and good practice.

MAP-E has been developed for use in organizational contexts, including selection, development, and coaching. It is not intended for clinical settings or for clinical purposes.

Computer skills

Because MAP-E is a web-based assessment, respondents need a basic level of digital familiarity and experience using online tools. They must be able to manage the relevant technical equipment, such as a mouse and/or keyboard. Before testing begins, the test administrator is responsible for ensuring that technical issues do not create unnecessary difficulty for the respondent, as such difficulties may negatively affect the results.

It is therefore important that the test administrator checks whether the respondent has the necessary prerequisites. If there is any uncertainty, the respondent should be given the opportunity to demonstrate, under the supervision of the test administrator, that they have the skills needed to complete the assessment. If a respondent is entirely unfamiliar with the required device or expresses strong reluctance toward completing a computer-based assessment, MAP-E should not be administered.

Readability

From the very beginning of the development of MAP-E, and throughout the process, a central ambition has been to keep all instructions and items as simple, clear, straightforward, and concise as possible. At the same time, the assessment does require a certain level of reading comprehension, although only a basic reading ability is needed to understand the instructions and items.

Even when content is designed to be simple, factors such as severe reading and/or writing difficulties, including dyslexia, or completing the test in a non-native language, may affect how the material is understood and may therefore also affect the results. At present, there is no specific information about whether, to what extent, or in what way scores may be influenced by these factors.

Other impairments

Other impairments, including but not limited to perceptual, visual, and cognitive impairments that may negatively affect test performance, should be identified and taken into account before MAP-E or any other assessment is administered.

The test administrator is responsible for considering how a respondent's specific impairment may influence completion of MAP-E and for determining what adjustments, accommodations, or adaptations may be needed to support the respondent appropriately within the overall assessment process.



Feedback options

MAP-E was designed to provide automatic feedback to the respondent. The test administrator is responsible for deciding whether additional feedback should be provided, and for delivering that feedback where applicable. This is essential to ensure that respondents feel like they are treated fairly, understand the purpose of the assessment, and can complete the assessment under appropriate conditions. If feedback is provided, it should be delivered in a way that is seen as fair, respectful, and non-intrusive.



5. Scale construction

This section outlines the process of item selection and construction of the most recent version of MAP Essence (MAP-E).

When constructing a shorter personality test, a key distinction can be made between broad and narrow measures. Broad measures capture a vast amount of behavioral information able to predict a wide range of criteria. Narrow measures describe behavior more concisely with the aim of predicting specific criteria with greater precision. In research, this distinction is referred to as the “bandwidth-fidelity” trade-off (Soto & John, 2016).

Trait models

In the process of constructing the scales in MAP-E, this trade-off was considered independently for each of the five traits to enhance different psychometric properties. Based on a large data set with scores on the 200 items from MAP (Assessio’s extensive personality test, assessing Big-5 personality traits with a significant layer of nuance by adding 5 facets to further qualify each trait score), five models were tested and evaluated on key psychometric criteria. The models are outlined below:

A) Fully optimized model

Picking the strongest items (based on corrected item-total correlations) using a top-down approach, where the weakest correlating items are removed successively, until reliability cannot be further improved or drops below a set value (e.g., 0.80 or 0.75).

B) Constrained optimized model

Picking the strongest correlating items (as in model A) but with the constraint that at least one item from each facet must be included (ensuring all facets are represented).

C) Fixed model

Picking the strongest items from each facet, ensuring that all facets are represented equally (i.e., have the same number of items).

D) Subfactor-optimized model

Picking facets from a single subfactor only, derived from splitting the overall traits into distinct, yet related subfactors representing different aspects of the construct, e.g., splitting Extraversion into the subfactors of Sociability (Social Need, Social Image, Cheerfulness) and Activity (Work Pace, Risk-Taking).

E) Criterion-optimized model

Picking items from facets that are positively related to key criteria, reducing the impact of weakly correlating facets and excluding facets with reversed prediction patterns (e.g., Risk-Taking in Extraversion).

In general, models A, B, and D have the advantages of maximizing reliability using the lowest number of items. On the other hand, they suffer the drawback of potentially under-representing (or leaving out entirely) certain facets that might be considered “hallmarks” of the trait in question and/or might be important for predicting various outcomes (in other words, maximizing reliability at the expense of construct and/or criterion validity). Conversely, model C ensures equal representation of all facets but does not take into account their importance for prediction, nor does it ensure the strongest reliability with the least number of items. Finally, model E maximizes predictive validity but also potentially represents a narrower construct that best predicts closely matched criteria (e.g., dependability, cooperation) at the expense of others (e.g., adaptive performance).

In deriving the final scales for MAP-E, we relied on a hybrid model encompassing most of the desired attributes, balancing different criteria. In some cases, most models converged into the same solution. For instance, the strongest correlating items for Extraversion were also items from “hallmark” facets of the trait (when conceptualized as “Sociability”) that were uniformly positively related to different criteria (mainly OCB and cooperation). In other cases, the different optimized models were adjusted to account for opposite patterns of prediction, e.g.,



positive relationships with counterproductive work behavior (CWB) for Ambition (in CO) and Confidence (in ES) or weak/negative correlations with task performance (TP) for Unconcern (in ES) or Aesthetics (in OP).

A brief overview of the excluded facets for each trait is shown in Table 5.1 alongside a list of alternative labels (operationalizations) and explanations.

Table 5.1. Overview of operationalizations and excluded facets in MAP Essence traits.

Scale	Alt. label	Excluded facets	Explanation
EX	Sociability	Risk-Taking	Less relevant for the subfactor, reversed patterns of prediction
AG	Kindness	Compassion	Weak correlations with key criteria (OCB, inverse-CWB)
CO	Diligence	Accountability Ambition	Weak correlations with key criteria (TP, inverse-CWB)
ES	Composure	Unconcern Confidence	Weak correlations with key criteria (OCB, inverse-CWB)
OP	Exploration	Aesthetics Self-Reflection	Less relevant for the subfactor, weak correlations with key criteria (OCB, inverse-CWB)

Psychometric criteria

As MAP-E is intended to be a short measure used for screening purposes, the following criteria were prioritized in constructing the final scales:

- **Number of items:** As few as possible (max. 10 per scale).
- **Reliability:** At least .75 (corresponding to SEM < 1).
- **Construct validity:** Trait correlation of at least .80 with the intended trait.
- **Criterion validity:** Uniform patterns of prediction for key criteria:
 - Task Performance (TP)
 - Organizational Citizenship Behavior (OCB)
 - Counterproductive Work Behavior (CWB)
- **Group differences:** Cohen's $d < 0.50$ (ideally < 0.20) for gender and age.
- **Adverse Impact:** No favoring of some groups over others as determined by the 4/5ths rule.
- **Qualitative criteria:** Item content, accessibility, social desirability, relevance for target group.

Upon deciding on a final model, each of the scales was subjected to an extensive psychometric validation procedure assessing the following aspects (elaborated in the sections below):

- 1) Consistency
- 2) Unidimensionality
- 3) Local independence
- 4) Item invariance (no Differential Item Functioning, DIF)
- 5) Scale characteristics
- 6) Group differences & Adverse Impact



6. Validity

While the definition of validity is agreed upon, researchers still debate how many types of validity there are and which research methods are best suited to investigate them. This is mainly due to the fact that, in practice, it can be difficult to determine the type of validity a given study relates to. However, there is a growing consensus that validity is a unitary concept, which can be documented by various forms of statistical and empirical studies.

In the following, validity is categorized and divided into face, content, construct, and criterion validity in accordance with the EFPA test review model (EFPA, 2013).

Face validity

Face validity concerns the extent to which test users and test takers perceive the questionnaire and results as relevant, comprehensive, and reflective of reality. Face validity is thus about whether a test comes across as credible to the ones using the test, which is important to ensure that they are sufficiently motivated to participate in the test and accept the conclusions drawn from it. Face validity is also about the extent that test results resemble how a person perceives themselves or is perceived by others.

To ensure face validity, two key aspects were considered during the development process. First, items with a high degree of transparency as to what is being measured were preferred over dubious and less relevant items (which are often the consequence of selecting items solely based on data). The use of mainly transparent items has the advantage of providing a clear link between the test taker's answers to the individual items and the final scores derived from them. Hence, the likelihood of test takers understanding, recognizing, and accepting their results is increased, which is crucial for self-awareness and general usage of the test for work-related purposes (be it recruitment or personal development). Although this transparency can make items more susceptible to faking or socially desirable responding (i.e., impression management), statistical analyses conducted during item development suggest that this is not a major cause of concern.

Second, items were carefully written to reflect behavior in a working context in keeping with the principle of contextualization, e.g., socializing with or helping colleagues (as opposed to friends or family).

Item examples for each scale can be found in the section above on scale definitions and interpretations.



Content validity

Content validity concerns whether test items and scales are relevant and representative for the theoretical concept (domain) being measured.

As MAP Essence is based on the FFM, the content of the items should clearly reflect the five personality traits in this model. However, as mentioned in the section on scale construction, these traits can be conceptualized more broadly or narrowly. To validate the operationalization of traits in MAP-E, the content was compared to two other trait measures established in previous research: The Big Five Inventory, 2nd version [BFI-2] (Soto & John, 2016) and the Big Five Aspect Scales [BFAS] (DeYoung et al., 2007). This comparison is shown in Table 6.1.

Table 6.1. Comparison of MAP Essence to other Big Five measures used in research.

Trait	Facet	BFI-2	BFAS
EX	Social Need	Sociability	Enthusiasm
	Social Image	Assertiveness	Assertiveness
	Work Pace	Energy level	Assertiveness
	Cheerfulness	Sociability	Enthusiasm
AG	Trust	Trust	-
	Diplomacy	Respectfulness	Politeness
	Helpfulness	Compassion	Compassion/Politeness
	Conflict Aversion	Respectfulness	Politeness
CO	Structure	Organization/Responsibility	Orderliness
	Self-Discipline	Productiveness	Industriousness
	Decision-Making	-	-
ES	Mood Stability	Emotional Volatility (R)	Volatility
	Self-Control	-	Volatility
	Stress Tolerance	-	Withdrawal
OP	Imagination	Creative Imagination	Openness
	Variety	-	Openness
	Mindset	Intellectual Curiosity	Intellect

The BFI-2 is a revision of the initial Big Five Inventory intended to measure prototypical components of each Big Five domain reliably and effectively. The BFI-2 includes 3 facets for each trait that were selected by identifying a “factor-pure” facet (a hallmark of the trait in question that is largely orthogonal to the other four domains), then adding two complementary facets based on previous research and the initial BFI (Soto & John, 2016).

When comparing the facets represented in BFI-2 and MAP-E, they are largely the same, except for the facet Unconcern missing in ES (which corresponds to an inversed combination of Anxiety and Depression in BFI-2), and Aesthetics missing in OP. Please note that the AG facet Compassion corresponds to Helpfulness in MAP-E, and that MAP-E combines the CO facets of Organization and Responsibility from BFI-2 into Structure.

The BFAS measures 10 different aspects of the Big Five and thus represents a level in between the traits and facets. These 10 aspects were derived from factor analyses splitting each trait into two distinct, yet related subdimensions with correlations ranging from .51 to .64 (DeYoung et al., 2007).

Importantly, MAP-E covers all 10 aspects but to different degrees. When compared to the BFAS, the subdimensions of Enthusiasm, Politeness, Orderliness, Volatility, and Openness have the most emphasis as reflected in the alternative labels for MAP-E scales mentioned previously (Sociability, Kindness, Diligence, Composure, Exploration).



Construct validity

Construct validity concerns the agreement between test results and prior theoretical knowledge of the construct being measured. As this validity aspect is quite broad, it is further divided below according to the final scale validation mentioned previously.

Consistency

First, internal consistency was measured using Cronbach's alpha. However, the alpha coefficient suffers the well-known drawback that it relies not only on the correlation between items and the scale but also the number of items (Taber, 2018). Furthermore, studies show that too high levels of alpha can be undesirable and often occur because of redundancy among items (Tavakol & Dennick, 2011).

Therefore, the alpha coefficient was supplemented by corrected item-total (item-rest-correlations), in which each of the items is correlated with the rest score (i.e., the scale score minus the score for the item in question). The higher the correlations, the higher the internal consistency of the scale. Also, consistency was evaluated by ensuring that all inter-item correlations were positive (as recommended by Streiner and Kottner, 2014).

Unidimensionality

Unidimensionality of the scales was investigated using an exploratory factor analysis (Principal Components Analysis, PCA) requiring the extraction of just a single factor. For all scales, a visual scree plot supported a single factor solution, which was confirmed by strong factor loadings, i.e., correlations between scores on the individual items and the score on the extracted factor.

Local independence

As stated above, high inter-item correlations were desired to improve the internal consistency of scales. However, too high values can reflect local dependence, i.e., that the response to one item depends on the response to another item, which inflates overall estimates of scale reliability (Marais, 2012). Ultimately, it can also jeopardize the candidate experience because candidates need to spend more time completing the assessment and experience what feels like answering the same item twice. Items were scrutinized for local dependence and excluded if it was too high.

Item invariance

In addition to the analyses of local independence, partial correlations were used to ensure that items exhibited item invariance, i.e., that there was no Differential Item Functioning (DIF) with respect to gender and age. Hence, response patterns of different demographic groups should be similar at comparable levels of the latent trait. Preventing DIF is a crucial aspect of ensuring a fair and unbiased assessment as well as obtaining estimates of group differences that can be attributed solely to item impact (as opposed to item bias). Items showing signs of DIF were excluded from the pool.

Scale characteristics

Finally, a range of analyses were conducted to investigate the properties of the final scale scores. Specifically, the mean, median, standard deviation (SD), and measures of skewness and kurtosis were computed for each scale. In addition, histograms were inspected visually to ensure that scales were properly normally distributed.



Table 6.2 lists these psychometric properties for each of the MAP-E scales. All scales had great consistency with alphas ranging from .74 to .79 and a median of .77. Item-rest and inter-item correlations were strong, averaging .45 and .27, respectively. In addition, exploratory factor analyses showed strong support for the extraction of a single factor as evidenced by the scree plot in Figure 6.1, with the primary factor explaining 36% of the variance, on average.

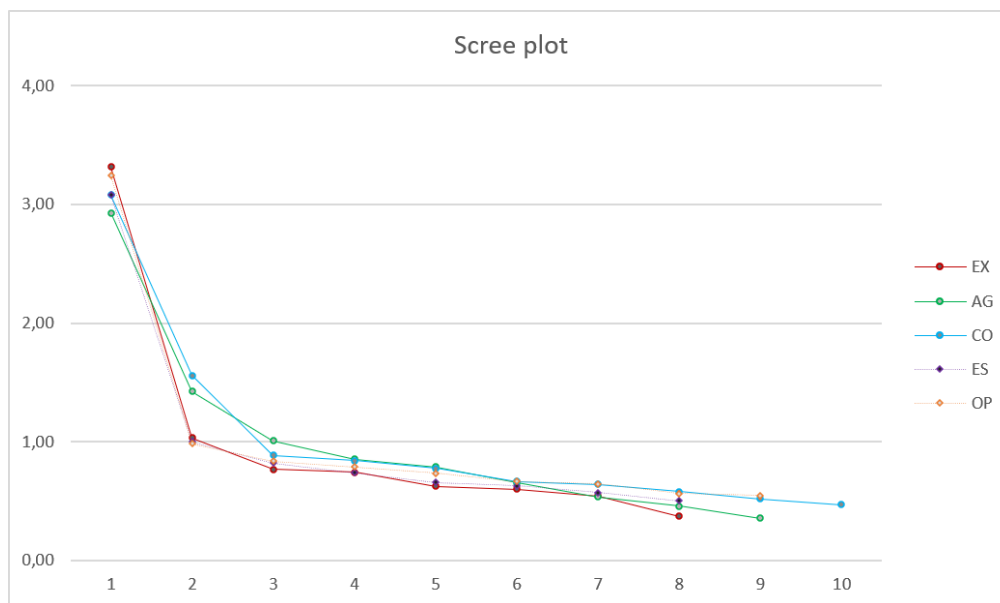
Furthermore, scales showed acceptable levels of local independence with only AG exceeding the recommended mark of .30 for maximum difference between the highest and average partial correlation. In terms of item invariance, partial correlations for any scale with any demographic variable were low, averaging .05 for gender and .07 for age indicating no Differential Item Functioning (DIF). Finally, scales showed a good range of scores with negligible amounts of skewness and kurtosis.

Table 6.2. Psychometric properties of scales in MAP Essence.

Psychometric property	EX	AG	CO	ES	OP	Mean	Median
Consistency							
Alpha	.79	.74	.75	.77	.77	.76	.77
Item-rest correlation (avg.)	.50	.41	.41	.47	.46	.45	.46
Inter-item correlation (avg.)	.30	.23	.23	.31	.28	.27	.28
Unidimensionality							
Factor loading (avg.)	.64	.57	.55	.62	.60	.60	.60
Eigenvalue	3.32	2.93	3.08	3.08	3.24	3.13	3.08
Explained variance (%)	41.5	32.5	30.8	38.5	36.0	35.9	36.0
Local independence							
Partial correlation (avg.)	-.16	-.12	-.10	-.12	-.11	-.12	-.12
Partial correlation (max.)	.15	.45	.20	.01	-.01	.16	.15
Max. difference	.31	.57	.30	.13	.10	.28	.30
Item invariance							
Gender (avg. correlation)	.06	.05	.05	.04	.05	.05	.05
Age (avg. correlation)	.12	.03	.05	.08	.07	.07	.07
Scale characteristics							
Mean	21.26	22.50	29.30	21.76	26.32	24.23	22.50
Median	21.0	22.0	30.0	22.0	26.5	24.3	22.0
SD	4.44	4.32	4.68	4.33	4.48	4.45	4.44
Skewness	-0.03	-0.03	-0.30	0.04	-0.34	-0.13	-0.03
Kurtosis	-0.34	-0.09	0.04	-0.25	0.00	-0.13	-0.09



Figure 6.1. Scree plot from factor analyses of the individual scales in MAP Essence.



Group differences

The validity of MAP-E is further supported by group differences that replicate previous findings in the research literature. Group differences were examined for gender (male/female) and age (below/above 40). When constructing the scales, we aimed to replicate differences to ensure validity but keep effect sizes as low as possible to prevent scores from negatively impacting hiring rates for different demographic groups (elaborated below in the section on Adverse Impact).

Tables 6.3 and 6.4 list the group differences for gender and age, respectively. As for gender, women scored slightly higher on Agreeableness and Conscientiousness, whereas men scored higher on Emotional Stability and Openness. However, effect sizes were negligible ($d < 0.20$) or small ($d < 0.50$) with an average Cohen's d of 0.12.

Regarding age, younger candidates scored slightly higher on all traits, with the largest differences observed for Agreeableness, Conscientiousness, and Openness. However, most differences were negligible ($d < 0.20$) or small ($d < 0.50$) with an average Cohen's d of 0.16.

Table 6.3. Group differences for gender (men/women) in MAP Essence.

Scale	Men			Women			Comparison		
	N	M	SD	N	M	SD	Dif.	t	d
EX	819	21.2	4.2	822	21.1	4.5	0.05	0.23	0.01
AG	819	22.9	4.1	822	23.1	4.0	-0.26	-1.28	0.06
CO	819	30.6	4.7	822	30.4	4.5	0.17	0.76	0.04
ES	819	22.1	4.5	822	21.1	4.5	1.02	4.65*	0.23
OP	819	27.5	4.1	822	26.3	4.3	1.16	5.55*	0.27

* $p < .001$ 

Table 6.4. Group differences for age (below/above 40) in MAP Essence.

Scale	< 40			40+			Comparison		
	N	M	SD	N	M	SD	Dif.	t	d
EX	1 141	21.3	4.4	504	20.8	4.4	0.49	2.07	0.11
AG	1 141	23.3	4.0	504	22.4	4.0	0.86	3.95*	0.21
CO	1 141	30.8	4.4	504	29.9	5.0	0.85	3.48*	0.19
ES	1 141	21.6	4.5	504	21.6	4.5	-0.04	-0.16	0.01
OP	1 141	27.3	4.1	504	26.1	4.5	1.27	5.61*	0.30

* $p < .001$

Exploratory factor analysis

Next, each of the items were subjected to an exploratory factor analysis (Principal Components Analysis) using the orthogonal Varimax rotation to preserve independence of factors.

The scree plot below in Figure 6.2 supported the extraction of five factors accounting for 39% of the total variance. Table 6.5 displays the primary and secondary factor loadings for each of the items and the five extracted factors (suppressing absolute values $< .30$). All but one item loaded the strongest on their intended factors, and only four items had strong cross-loadings ($> .30$) with one secondary factor. These cross-loadings can be attributed to the conceptual overlap between Conscientiousness (especially Self-Discipline) and Emotional Stability.

Figure 6.2. Scree plot from factor analysis of all items in MAP Essence.

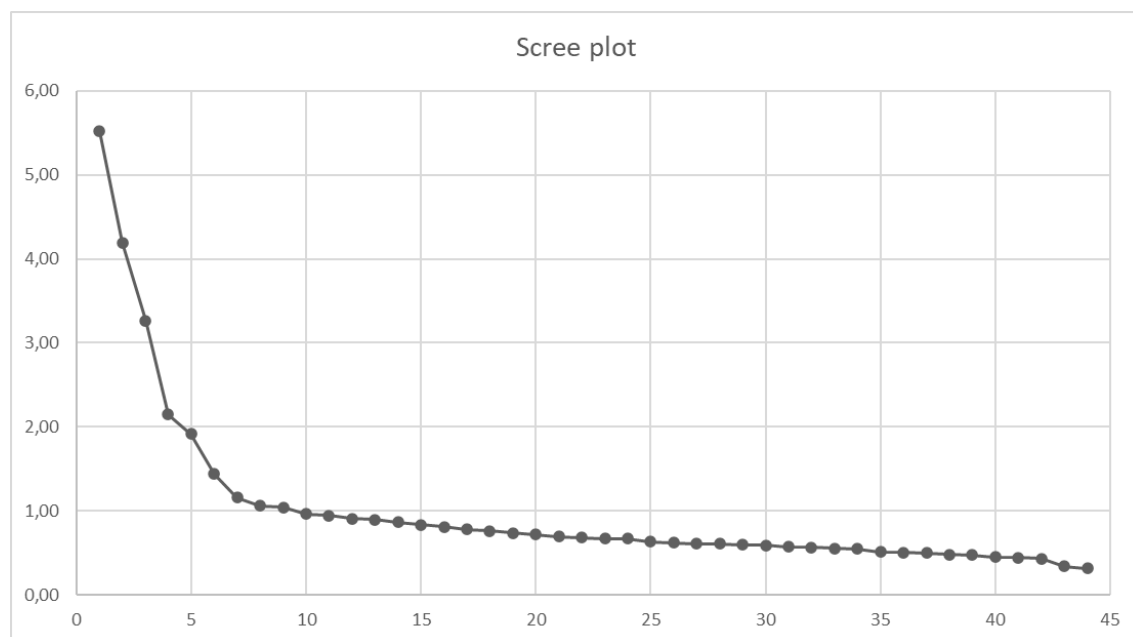


Table 6.5 Factor loading of items on main factors in MAP Essence.

Item	I	II	III	IV	V
EX1			.53		
EX2			.59		
EX3			.69		
EX4			.75		
EX5			.56		
EX6			.49		
EX7			.48		
EX8			.71		
AG1				.53	
AG2				.56	
AG3				.63	
AG4				.52	
AG5				.62	
AG6				.54	
AG7				.62	
AG8				.47	
AG9				.46	
CO1					.68
CO2					.68
CO3					.56
CO4					.40
CO5					.63
CO6	.43				.30
CO7	.39				.44
CO8	.47				.47
CO9	.33				.40
CO10					.60
ES1	.44		.33		
ES2	.64				
ES3	.57				
ES4	.65				
ES5	.68				
ES6	.55				
ES7	.55				
ES8	.66				
OP1		.64			
OP2		.65			
OP3		.57			
OP4		.59			
OP5		.42			
OP6		.44			
OP7		.61			
OP8		.53			
OP9		.56			

Note. Absolute values < .30 not shown.



Convergent validity

Finally, analyses exploring both convergent and discriminant validity of the scales were conducted. Table 6.6 shows the correlations between MAP-E scales and trait scores based on the more extensive personality test, MAP. Numbers on the diagonal display correlations between trait scores in MAP and MAP-E, whereas numbers above the diagonal reflect internal correlations between scales in MAP-E.

In general, MAP-E scales show strong convergent validity with traits correlations ranging from .75 to .85 with an average of .82 and a median of .84. The somewhat weaker correlation for Conscientiousness is caused by the narrow operationalization of the construct, limiting the impact of Decision-Making and excluding the facets of Accountability and Ambition altogether.

As for discriminant validity, correlations between traits were mostly small, with the exception of moderate correlations between Extraversion and Openness (conceptual overlap related to flexibility) and between Conscientiousness and Emotional Stability (conceptual overlap related to inhibition and impulse control).

Table 6.6 Convergent and discriminant validity of MAP Essence traits.

MAP/MAP-E trait	EX	AG	CO	ES	OP
Extraversion	.85	-.02	.12	.21	.41
Agreeableness	-	.82	.03	-.07	.04
Conscientiousness	-	-	.75	.39	.11
Emotional Stability	-	-	-	.84	.06
Openness	-	-	-	-	.85

Note. Numbers in bold display correlations between trait scores in MAP and MAP-E, whereas those above the emboldened diagonal reflect internal correlations between the scales of MAP-E.

Criterion validity

This refers to how well the test predicts relevant performance criteria. Thus, we wanted to estimate how well MAP-E predicts the two screening criteria we have available, i.e., integrity and service-orientation.

Validity estimates from a MAP study

Because no performance data was available for people who completed MAP Essence, we relied on data collected with a previous version of MAP. A total of 254 candidates were included. The scores on the five personality dimensions were calculated as a weighted average between the relevant facets (derived from linear regressions of MAP-E trait scores with MAP facets as predictors), and the weights were corrected to account for the level of equivalence between the previous and updated MAP. While those scores aren't identical to MAP-E, they are similar enough to offer a good estimation on how MAP-E will perform in practice. This section will be supplemented with better and more extensive information as we continue to monitor the quality of MAP-E and conduct validation studies.

Personality scores were correlated with two criteria, a measure of counterproductive work behavior (CWB) and an assessment of organizational citizenship behavior (OCB). When comparing them with our screening criteria, CWB can be understood as the opposite of Integrity. Other research studies also show that those two are largely driven by the same personality traits, only in opposite directions. For instance, a conscientious person is more likely to act with integrity and less likely to be counterproductive, and the opposite is true for someone low on Conscientiousness. OCB is more closely related to service-orientation, because it measures how often someone volunteers to do extra tasks that support the organization, its members, or its collaborators. Below is a description of the instruments used:



- **Counterproductive Work Behavior (CWB):** CWB was measured with 10-item short version of the Counterproductive Work Behavior Checklist (CWB-C) assessing a range of different counterproductive work behaviors such as coming in late without permission or starting arguments with coworkers (Spector et al., 2010).
- **Organizational Citizenship Behavior (OCB):** OCB was assessed with a 25-item questionnaire (Staufenbiel & Hartz, 2000) assessing the facets of altruism (helping others), conscientiousness (following rules and instructions), sportsmanship (handling changes and unforeseen events without complaining), and civic virtue (attending conferences, improving skills, representing the company well outside of daily work, etc.).

Our data shows that even shortened personality scales can produce a solid prediction of performance criteria. The results of a correlation analysis are presented in Table 6.7. For simplicity, only the strength of significant correlations is presented. As Openness did not correlate with any criteria, we decided not to include it in the screening algorithms for Integrity and Service. However, Openness is an excellent predictor of other forms of success, such as adaptive performance. If we expand the repertoire of lenses in the future, this trait may be included in other screening algorithms.

Table 6.7. Correlations of MAP-E scales with CWB and OCB.

Trait	CWB		OCB	
	Correlation (r)	Strength	Correlation (r)	Strength
EX	-.04	-	.45**	Moderate
AG	-.38**	Moderate	.56**	Strong
CO	-.41**	Moderate	.35**	Moderate
ES	-.28**	Small	.47**	Moderate
OP	.05	-	.12	Small

Note. *p < .05, **p < .01. CWB = Counterproductive Work Behavior. OCB = Organizational Citizenship Behavior.



7. Reliability

Reliability is defined as the consistency with which an instrument measures a construct.

An often-used measure of internal consistency is Cronbach's alpha (Cronbach, 1951), which is listed below for each of the MAP scales in Table X. The final column of the table contains the Standard Error of Measurement (SEM) defined as:

$$\text{SEM} = \text{SD} * \sqrt{(1-r)}$$

Where SD represents the standard deviation, and r refers to the reliability of the scale in question.

As shown in Table 7.1, all scales have acceptable levels of reliability with alphas ranging from .74 to .79 and an average of .76. The Standard Error of Measurement (for the raw score) ranged from 2.01 to 2.30 with an average of 2.11. When normed and converted to C-scores (with an SD of 2), the average SEM was only 0.97, i.e., less than one C score point.

In sum, these estimates show great consistency and measurement accuracy of MAP-E.

Table 7.1. Reliability and SEM of scales in MAP Essence.

Scale	Items	Alpha	SD	SEM	SEM (C score)
EX	8	.79	4.44	2.01	0.92
AG	9	.74	4.32	2.06	1.02
CO	10	.75	4.68	2.30	1.00
ES	8	.77	4.33	2.15	0.96
OP	9	.77	4.48	2.04	0.96
Mean	8.8	.76	4.45	2.11	0.97
Median	9.0	.77	4.44	2.06	0.96

Note. SEM = Standard Error of Measurement.



8. Standardization

Standardization refers to the procedure of design and testing that leads to a standardized test. Standardization thus says something about the way in which the test is constructed, thoroughly tried, and tested. There are several ways to standardize, where the best known are the normative and ipsative methods. MAP is a normative test, which means that the test result is compared to a relevant norm group.

One of the advantages of normative tests is that they are quick and straightforward to complete. Although there are typically more questions compared to other types of tests, normative tests still take a short time to complete because the questions are easier to answer. Another strength of the normative method is that the test scales are completely independent of each other. Because the scales are measured one at a time, normative tests show more nuances and make the measurements more accurate.

Most importantly, normative tests are suitable for comparing individuals. Normative tests not only provide answers to what is characteristic of individuals but also what is characteristic of test persons in relation to others. Normative tests thus measure interpersonal differences (differences between people), where the person's response is compared to the responses of others. Therefore, the normative approach is the preferred method when a tool is to be used for selection purposes and is also an ideal tool for development purposes (providing insight as to how the individual differs from others).

Score calculation

First, the responses to all items (answered on a 4-point Likert scale with the options Disagree/Slightly disagree/Slightly agree/Agree) in each scale are summed to a raw score, which is then compared to the answers from a norm group. These raw scores are then converted to z-scores by subtracting the raw score from the mean and dividing by the standard deviation of scores in the norm group. Then, z-scores are converted to C scores with a mean of 5 and a standard deviation of 2. Those are displayed as the results. The interpretations, z-score ranges, percentages, and percentiles for C scores are shown below in Table 8.1.

Table 8.1. C score distributions and percentiles.

Category	C score	z-score	Percentage	Percentile
Low	0	-2.75;2.25	1	1
	1	-2.25;-1.75	3	4
	2	-1.75;-1.25	7	11
	3	-1.25;-0,75	12	23
Moderate	4	-0.75;-0-25	17	40
	5	-0.25;0.25	20	60
	6	0.25;0.75	17	77
High	7	0.75;1.25	12	89
	8	1.25;1.75	7	96
	9	1.75;2.25	3	99
	10	2.25-2.75	1	100



Norm group

At Assessio, we are committed to offering norms of the highest quality based on quality standards derived from various international standards, including EFPA, COTAN, and ITC guidelines. In short, these guidelines set out criteria for various aspects of the norm group:

- **Update:** When was the norm group last updated?
- **Sample size:** How large is the norm group? Is it sufficiently large to ensure representativeness?
- **Composition:** How is the norm group composed with respect to different demographics?
- **Subgroup differences:** Are group differences sufficiently small to prevent adverse impact?

Update

Over time, what is considered normal behavior changes. Major life changes and historical events impact the way people behave, and new generations may also challenge the existing standards. Therefore, with respect to assessments, it is very important to update norm groups regularly and thus make sure that all people assessed are compared with a norm group representing what is currently considered “normal” behavior. This will provide the most valid assessment. In addition, updating the norm group keeps scores balanced and avoids too many candidates getting either very high or very low scores. In other words, norm updates allow for better differentiation of candidates, which in turn leads to better recruitment decisions. According to EFPA and COTAN guidelines, a norm of the highest quality should not be older than 10 or 15 years, respectively. At Assessio, however, we check whether norms need updating every 2 years.

The MAP-E norm group was updated in 2026 based on data collected from MAP in a high-stake setting (recruitment) from May 2025 to May 2026.

Sample size

A good norm group consists of many people, as a high number provides greater representation and statistical certainty. The dominant view is that the larger the sample, the better the norm group. While that is true, it very much depends on sampling procedures as well as composition with respect to different demographic characteristics. In general, norm groups that are too small run the risk of underrepresentation (e.g., too few people with a certain occupation or education level), whereas too large norm groups risk overrepresentation (e.g., too many people of a certain age or nationality). According to EFPA, a sample size of at least 1,000 constitutes an excellent norm group (in some cases, smaller norm groups may also be sufficient depending on composition, target groups, and intended applications). For high-stake purposes, a norm group consisting of 400-999 people is considered a good sample size (EFPA, 2013).

The norm group consists of 1,645 people selected through stratified randomization from a total of 59,712 people aged 16-70 who completed MAP in a high-stake setting. Dropout analyses showed only a small selection bias towards lower mean scores for Emotional Stability ($d = 0.26$) and higher mean scores for Openness ($d = 0.24$). Across scales, Cohen's d averaged 0.13.



Composition

To ensure that a norm group is representative of all target groups and is appropriate for all intended applications, key demographic characteristics must be carefully weighed and balanced, especially those that can lead to potential score differences between subgroups.

To obtain a sample representative of the work segment for MAP Essence, the data from MAP was stratified with respect to four key aspects: country of residence, education level, age, and gender. The key variables were nested such that an equal amount of people were randomly selected and included in each of the strata to balance out all variables simultaneously. Each country was constrained to account for no more than 8 % of the total norm group (although a lower percentage could be desired, this was needed to balance remaining key variables and ensure a sufficient sample size). Education levels were rebalanced aiming for a 70/30 split between lower education levels, i.e., less than a bachelor's degree (ISCED levels 0-5) and higher education levels, i.e., a bachelor's degree or higher (ISCED levels 6-8). The same logic was applied for age, such that 70% of the norm group was covered by the youngest age groups aged 16-39. Finally, main genders were balanced to have an equal distribution, and other genders were included as well.

As a result of these iterations, a total of 100 countries and 97 nationalities were included, with the highest percentage being 9.3% and 7.7%, respectively. Distributions for the remaining demographics ended up as expected. A short breakdown of the demographic composition is shown in Table 8.2. Due to this composition, please note that the norm group is less suitable for purposes other than selection (recruitment) and candidates with a higher educational level (bachelor's degree and above).

Table 8.2. Composition of the global norm for MAP Essence.

MAP Essence: Global norm	
Update	2026
Data collection	May 2025- May 2026
Sample size	1,645
Purpose	Selection
Gender	Men: 49.8% Women: 50.0% Other: 0.2%
Age	M = 34.5, SD = 10.5 16-24: 20.5% 25-39: 48.8% 40-54: 26.3% 55-70: 4.3%
Nationalities	97 Max.: 7.7%
Countries	100 Max.: 9.3% Native residents: 61.2%
Education level (ISCED)	
0-1. Not completed lower secondary education	1.0%
2. Lower secondary education	6.4%
3. High school or vocational education	40.7%
4-5. Short-cycle higher education	21.0%
6. Bachelor's degree or equivalent	12.8%
7. Master's degree or equivalent	17.3%
8. PhD or doctoral degree	0.7%



Group differences & Adverse Impact

When using an assessment to make important decisions with a great impact on individuals (such as selection, promotion, and hiring decisions), a key requirement is to ensure fairness and mitigate Adverse Impact, defined as “a substantially different rate of selection in hiring, promotion, or other employment decisions which works to the disadvantage of members of a race, sex or ethnic group” (Uniform Guidelines on Employee Selection Procedures, Equal Employment Opportunity Commission, 1978). The “Four-Fifths rule” can be used to determine whether an assessment has Adverse Impact. Usually, a selection rate for any demographic group that is less than four-fifths (or 80 percent) of the selection rate for the group with the highest rate (majority group) is considered evidence of Adverse Impact. The level of impact depends both on the magnitude of group differences (e.g., between males and females) and the selection ratio, i.e., the number of people hired compared to the total number of applicants.

The first step in preventing Adverse Impact is to ensure that differences between demographic groups reflect true differences (item impact) and not item bias (or Differential Item Functioning, DIF). That is, including items with a uniform bias for one demographic group would inflate group differences, thus creating a disadvantage for the lowest scoring group. Ensuring that items are not subject to item bias ensure a report of true group differences. However, true group differences can still produce Adverse Impact making it important to study further.

Simulations of expected Adverse Impact were conducted at different selection rates for gender (men/women) and age (above/below 40 and above/below 25). Please note, however, that these calculations are based on the assumptions that 1) candidates are selected based on a single score only, 2) the assessment is used as the sole basis for selection, and 3) a fixed selection rate is applied (i.e., hiring everyone scoring above a predefined cut-off). In practice, Assessio recommends basing recruitment decisions on a combination of assessments, scales, and other information relevant to the job in question (i.e., KSAOs) to consider both job, team, and organization fit.

Tables 8.3 and 8.4 list the standardized mean difference (Cohen's d) between groups alongside the simulated Adverse Impact Ratio (AIR; selection rate of the least represented group compared to the most represented group) for gender (men/women), and age (above/below 40 as well as above/below 25), respectively. The calculations are based on three fixed selection ratios (SR): Strict (C score 7-10), Moderate (C score 6-10), and Lenient (C score 5-10), equivalent to the top 23, 40, and 60%, respectively. For any given scale, we aimed for an AI ratio above 0.80 for a lenient selection ratio as suggested by the Four-Fifths rule.

For most scales and selections ratios, the AIR was well below the .80 cut-off suggested by the four-fifths rule. Only exceptions were Emotional Stability and Openness slightly favoring men if using a strict selection ratio, and Openness favoring younger applicants below 40 for strict and lenient selection ratios. For the remaining comparisons and lenient selection ratios altogether, no evidence of adverse impact was found for any scale with respect to any of the demographic groups.



Table 8.3. Adverse impact ratios for gender (men/women) at different selection ratios.

Scale	Dif.	d	Strict	Moderate	Lenient
EX	0.05	0.01	0.98	0.93	1.00
AG	-0.26	0.06	0.90	0.89	0.96
CO	0.17	0.04	0.89	0.94	0.99
ES	1.02	0.23	0.79	0.84	0.86
OP	1.16	0.27	0.66	0.80	0.86

Table 8.4. Adverse impact ratios for age (below/above 40 and 25) at different selection ratios.

Scale	Main dif.		Below/Above 40			Below/Above 25		
	Dif.	d	Strict	Moderate	Lenient	Strict	Moderate	Lenient
EX	0.49	0.11	0.73	0.79	0.89	0.81	0.91	0.94
AG	0.86	0.21	0.74	0.81	0.87	0.82	0.80	0.85
CO	0.85	0.19	0.93	0.90	0.88	0.91	0.91	0.86
ES	-0.04	0.01	0.80	0.84	0.90	0.94	0.96	0.98
OP	1.27	0.30	0.78	0.77	0.84	0.74	0.75	0.81



9. Translations and adaptations

The following section outlines the process for translations and adaptation of items.

General procedure for translations and adaptations

When translating an existing instrument from the original (source) language into a new (target) language, it is crucial to consider two different aspects: linguistic similarity and psychological similarity. Linguistic (or semantic) similarity is often established using literal translations with the purpose of developing a test as close to the original as possible with respect to language characteristics such as wording, phrases, and connotations. In the literature, this method is variously termed “translation” or “application” (van de Vijver & Poortinga, 2005).

However, enhancing linguistic similarity sometimes occurs at the expense of psychological similarity, defined as individual test items having the same meaning, interpretation, and level of difficulty across languages. The process of changing certain items to ensure psychological similarity is commonly referred to as “adaptation” (van de Vijver & Poortinga, 2005). These changes span the range of minor adjustments (such as changing the currency on items of numerical skills) to entire revisions, i.e. rewriting items completely to capture the intended meaning (as opposed to opting for a literal, yet non-idiomatic translation).

Translational protocol

When translating/adapting our assessment tools, we aim at enhancing both linguistic and psychological similarity. Wherever possible, literal translations of items are preserved with only minor adjustments made when needed. When literal translations do not suffice, alterations are made to adapt to the specific target language and culture. In this regard, our procedures closely follow expert recommendations on conducting adaptations and are aligned with the criteria of EFPA's test review model (van de Vijver & Poortinga, 2005; EFPA, 2013).

Our translations/adaptations follow a strict protocol with the following steps:

- 1) Selection of translators and reviewers**
Translators and reviewers are certified native language translators residing in the country of the target language, which ensures high quality authorized translations.
- 2) Briefing of translators**
Translators are given thorough instructions on translation procedures and principles with respect to wording, style, and localization (e.g., selecting proper currencies, metrics, names, idiomatic language, etc.). The full list of instructions is outlined below.
- 3) First translation**
The first translation is always based on the English source version (or multiple source versions, if possible).
- 4) First proof reading**
A proof reading of the first translation is made by a different certified translator.
- 5) Back-translation**
The most recent version is translated back into the source language aimed specifically at ensuring linguistic similarity and identifying necessary adaptations (i.e., discrepancies between the target and source language versions).
- 6) First revision**
Any discrepancies and adaptations made to the target language versions are reviewed by our own team of subject matter experts. Then, adjustments are made in close



collaboration with the initial translator to ensure equivalence with the source language version and that ordinary, idiomatic language is used in the target language version.

7) First implementation

The first version is implemented in the test platform.

8) Functional test

From the test platform, a certified native-speaking translator makes a proof reading and suggests adjustments in relation to the target context.

9) Second revision

Based on the functional test, the translator suggests adjustments that are discussed with our team of subject matter experts to arrive at the second version.

10) Second implementation

The second version is implemented in the test platform and made available to candidates.

11) Statistical validation

Finally, the translation is statistically validated upon data collection, and the final adaptation is made.

12) Final version

Based on the results of the statistical analyses, the translation is finalized.

Instructions for translators

When translating items to a new target language, translators are given the following instructions:

- The translation should follow the original wording as closely as possible. Adaptation is only allowed if necessary for making the text comprehensible, meaningful, or accurate in the target language. It can, however, be necessary to make slight adaptations to make the item sound natural to the native speaking test person.
- The text should not be changed unnecessarily – especially, negations must not be removed or changed if this alters the coding or “direction” of the answers.
- Phrases, idioms, or other kinds of figurative speech should be omitted, as these may be unknown to or interpreted differently by different test persons.
Source versions: The first translation is always based on the English source version and a second version from the same or adjacent linguistic families (if possible).
- **Forms of address:** By default, the form of address should be informal, unless this is considered unprofessional, rude, or offensive (or in opposition to common usage) in the target language/culture.
- By default, forms of address should be neutral, unless it opposes common usage or is not a viable alternative in the target language.
- Non-gender specific terms are used, or feminine endings are added (e.g. in parentheses) whenever possible. If this is not possible in the language in question, our subject matter experts are involved in the decision.



References

- Catano, V. M., O'Keefe, D. F., Francis, R. E., & Owens, S. M. (2018). Construct-based approach to developing a short, personality-based measure of integrity. *International Journal of Selection and Assessment*, 26(1), 75-92.
- Costa, P.T. and McCrae, R.R. (2008) The Revised NEO Personality Inventory (NEO-PI-R). In: Boyle, G.J., Matthews, G. and Saklofske, D.H., Eds., *The SAGE Handbook of Personality Theory and Assessment: Volume 2—Personality Measurement and Testing*, SAGE Publications, 179-198.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), 297-334.
- Dalal, R. S. (2005). A Meta-Analysis of the Relationship Between Organizational Citizenship Behavior and Counterproductive Work Behavior. *Journal of Applied Psychology*, 90(6), 1241-1255. <https://doi.org/10.1037/0021-9010.90.6.1241>
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between Facets and Domains: 10 Aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880-896.
- Doan, T., Kanjanakan, P., Zhu, D., & Kim, P. B. (2021). Consequences of employee personality in the hospitality context: a systematic review and meta-analysis. *International Journal of Contemporary Hospitality Management*, 33(10), 3814-3832.
- European Federation of Psychologists' Associations [EFPA] (2013). Review model for the description and evaluation of psychological and educational tests. ver. 4.2.6. EFPA Board of Assessment.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85(6), 869-879.
- Lounsbury, J. W., Foster, N., Carmody, P. C., Young Kim, J., Gibson, L. W., & Drost, A. W. (2012). Key personality traits and career satisfaction of customer service workers. *Managing Service Quality: An International Journal*, 22(5), 517-536.
- Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five, or one in addition? *Personnel Psychology*, 60(1), 1-34.
- Motowidlo, S. J., Brownlee, A. L., & Schmit, M. J. (2008). Effects of personality characteristics on knowledge, skill, and performance in servicing retail customers. *International Journal of Selection and Assessment*, 16(3), 272-280.
- Mróz, J., & Kaleta, K. (2016). Relationships between personality, emotional labor, work engagement and job satisfaction in service professions. *International journal of occupational medicine and environmental health*, 29(5), 767-782.
- Pradhan, R. K., & Jena, L. K. (2017). Employee performance at workplace: Conceptual model and empirical validation. *Business perspectives and research*, 5(1), 69-85. <https://doi.org/10.1177/2278533716671630>
- Soto, C. J. & John, O. P. (2016). Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *Journal of Personality and Social Psychology*. Advance Online Publication.
- Spector, P. E., Bauer, J. A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *Journal of Applied Psychology*, 95(4), 781-790.
- Staufenbiel, T. & Hartz, C. (2000). Organizational Citizenship Behavior: Entwicklung und erste Validierung eines Meßinstruments. *Diagnostica*, 46(2), 73-83.



- Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing* 70(9), 1970–1979.
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
- Van de Vijver, F. J. R. & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, & C. Spielberger (Eds.). *Adapting Educational and Psychological Tests for cross-cultural assessment* (pp. 39-64). Lawrence Erlbaum Associates, Inc.. Publishers.
- Wing-Man Lau, R., Chan, D. K. S., Sun, F., & Cheng, G. H. L. (2023). Predictive validity of integrity tests for workplace deviance across industries and countries in the past 50 years: A meta-analytic review. *Journal of Pacific Rim Psychology*, 17, 18344909231171729.

