MATCH-V

Technical manual
Version 2.0

HELENE HOPPE REVALD
Director of Psychometrics, Occupational Psychologist

DENNIS HVASS
Psychometrician, Occupational Psychologist

Assessio

Zero Talent Waste.

Table of contents

Table of contents

1. Introduction

Match-V is an 85-item assessment measuring nine basic human values. The assessment can be used for a range of different purposes, including selection (culture-fit) and development. This technical manual outlines the theoretical background, interpretation, and construction of the scales as well as psychometric properties in terms of validity, reliability, and norms.

2. Background

Theoretical framework

Understanding the values and motivational goals of humans has attracted much attention from philosophers, psychologists, and organizational leaders. In recent years, understanding the values of employees has become of much importance to organizational leaders given that it is estimated that 66% of the workforce are disengaged at work, 51% are dissatisfied with their job, and 90% would gladly take a pay cut to work in a meaningful role (Achor, Reece, Kellerman, & Robichaux, 2018; Gallup, 2019; Pew Research Center, 2016). Sadly, the evidence indicates that most organizations are failing to inspire and engage their employees, and their organization's productivity and growth is lacking because of it.

If organizations are to successfully attract, attain, and retain a talented workforce, they need to better understand what drives their people, assign them to tasks and projects that appeal to them, and build a culture that reflects and enables the expression of such values.

It is for this reason the MATCH-V was developed. The MATCH-V can be used to measure an individual's basic values, alongside measuring an organization's culture if deployed to an entire workforce. The MATCH-V can be used by I/O psychologists, practitioners, and leaders to inform talent management strategies, improve coaching and development, and support organizational change and strategic initiatives. The MATCH-V offers scientific and data-driven insights about an organization's talent to help people live more meaningful lives, make the workplace a more satisfying environment, and make organizations more productive.

The nature of values

Within research, there has been a great interest in understanding human values and how they relate to other constructs such as motivation, needs, interests, and preferences – terms that are often used interchangeably. Clearly, these constructs overlap but are not identical (i.e., they share some, but not all, characteristics).

Based on an extensive literature review, we posit that these constructs can be placed on a continuum from Universal (common) to Particular (individual). That is, reading from left to right, the concepts are increasingly individual and shared by less people (i.e., all people have needs, a lot of people have similar values but very few people have the exact same interests).

At the far left of this continuum, we find Needs at the most fundamental level. In the literature, psychological needs are universal in nature and thus shared by all individuals across all cultures.

Maslow suggested a hierarchy of needs, whereas Self-Determination Theory (SDT) outlines three basic needs for autonomy, competence, and relatedness, the fulfilment of which fosters intrinsic motivation, psychological growth, and well-being (Ryan & Deci, 2000; Deci & Ryan, 2010; 2012). At the far right, we would find terms such as Motives (or Drivers), Interests, and Preferences; concepts that are highly similar and share many theoretical overlaps. All of these are characterized by being highly individual and useful for evaluating person-job or person-team fit as opposed to person-organization fit (McClelland, 1985; Holland, 1959; Furnham, 2001; Belbin, 2017).

In between these two, we find Values and Goals, which have a universal nature but are not shared by all individuals to the same degree and varies more between cultures (than do Needs, for instance). In other words, values are crucial for explaining the choices we make and how we organize ourselves socially and personally. Thus, they form the motivational bases of attitudes and behavior. Research suggests that values and their structure are universal because they are structured in similar ways across different culturally diverse groups. However, individuals and groups differ substantially in the relative importance they attribute to these values. That is, individuals and groups have different value "priorities" or "hierarchies" (Schwartz, 2012).

Schwartz' theory of basic human values

Schwartz' theory of basic human values identifies universally common, yet motivationally distinct types of values as well as the dynamic relations among them. Some values may conflict with one another (e.g., Connection and Status) whereas others are more compatible (e.g., Conformity and Security).

When we think of our values, we think of what is important and ideal to us. Each of us holds numerous values with varying degrees of relative importance. Also, a value may be very important to one person but less so to another. Generally, values are characterized by the following characteristics:

- Values are beliefs: When values are activated, they become infused with feeling. People become aroused if their values are threatened, feel despair when they are helpless to protect it, and are happy when they can enjoy it.
- Values refer to desirable goals: Values motivate action. We are motivated to pursue the values we find important.
- Values transcend specific actions and situations: What we value in one context usually applies for multiple contexts.
- Values serve as standards: The way in which we evaluate actions, policies, people, and events as good or bad, justified, or illegitimate, worth doing or avoiding are guided by our values. These evaluations are rarely conscious but become apparent when in conflict with a cherished value.
- Values are ordered by importance: People's values form an ordered system of priorities that characterize them as individuals and guide their actions as well as aversions.
- The relative importance of multiple values guides action: Attitudes or behaviors typically have implications for more than one value. The tradeoff among relevant, competing values guides the individual's attitudes and behaviors. Values influence action when they are both relevant in the context and important to the actor.

The extent to which they influence behavior is dependent on how easily a value is activated (Verplanken & Holland, 2002). The more an individual places importance on a given value, the more readily activated it is, and likely to result in influencing behavior (Schwartz, 2006). Finally, the more an individual can live according to these values, and have them satisfied, the more likely they are to experience positive affect and engagement at work (George & Jones, 1996; Rich, Lepine, & Crawford, 2010). Altruists who value social justice are more likely to find that value activated in organizations dedicated to social justice than investment banking.

Organizational culture & person-environment fit

Organizational culture may be defined as the shared basic assumptions, values, and beliefs that characterize a setting and are taught to new incumbents as the proper way to think and feel, communicated by the myths and stories people tell about how the organization came to be the way it is as it solved problems associated with external adaptation and internal integration (Schneider, Ehrhart, & Macey, 2013). In other words, organizational culture describes "how things get done around here". Organizational cultures differ between organizations and industries, and ideally, culture reflects and supports the organization's strategy. Accordingly, culture is often found to explain why some organizations are more productive and effective than others (S. J. Hogan & Coote, 2014; Ogbonna & Harris, 2000). Continuing this, 94% of executives believe that distinct workplace culture is important to business success, only 19% of executives believe their company has the 'right culture' (Bersin, Geller, Wakefield, & Walsh, 2016). It is therefore no surprise that organizations invest heavily on building, improving and maintaining cultures.

Organizational culture is not just important for incumbents. In fact, culture is a powerful mechanism that externally communicates the organization's values, thereby shaping its reputation — thereby influencing perceptions of stockholders, potential clients, and job applicants. From such a perspective, organizational culture can be understood as communicating the firm's brand (Hatch & Schultz, 1997). A strong and desirable culture that is communicated externally increases stock prices (Chamberlain, 2015), decreases cost per hire (Guletkin, 2011), improves the number of qualified individuals applying for jobs, attracts diverse candidates, and drives employee referrals (Kennedy & Hill, 2016).

Value congruence

Given that organizational culture describes a group's shared goals, assumptions, and beliefs, culture can be understood through the lens of values theory. Specifically, a culture reflects the aggregated values of a group. At this level, shared values create established social norms around appropriate and inappropriate behaviors, prototypical groups roles, and a clear in-/out-group identity that enforces a sense of social connection while also continuing to orientate, influence and motivate individual behavior towards similar goals (Ashforth & Mael, 1989; Cornelissen, Haslam, & Balmer, 2007; Haslam, Postmes, & Ellemers, 2003). From this perspective, if an individual is to "fit" within a group and its culture, there must be value congruence (Kristof, 1996).

Value congruence can be described as the compatibility of values between individuals and other organizational entities such as supervisors, interviewers, coworkers, work group, and the entire organization (Chatman, 1989; Kristof, 1996). Value congruence is often examined alongside theories of person-organization fit (Westerman & Cyr, 2004), which postulates that if there is a congruence between an individual's values, personality, and expertise and the organization's

culture and strategy, there will be increased levels of job performance, engagement, intrinsic motivation and citizenship behaviors, alongside reduced turnover and counterproductive work behaviors (Bao, Dolan, & Tzafrir, 2012; Chatman, 1989; De Cooman et al., 2009; Kristof, 1996; Westerman & Cyr, 2004). Influenced by this research, it is now best practice for hiring managers and leaders to maximize "fit" when selecting new employees, building teams, and undergoing organizational transformation.

Supplementary fit vs. complementary fit

Although the notions of person-environment fit and value congruence have often focused on hiring people that fit within the existing team or organization (usually termed supplementary fit), research distinguishes a second type of fit known as complementary fit. Unlike supplementary fit (focusing on congruence), complementary fit focuses on adding values, traits, skills, and behaviors that are missing from the existing team or organization as a means to develop and transform the organization to meet new business needs, adjust to market trends and adapt to societal and environmental changes (Malinowski et al., 2006).

Although not stated in a published scientific article, in a TED talk on work life, renowned organizational psychologist Adam Grant advocates the need for striking a balance between two fundamental dimensions of results vs. relationships and rules vs. risk. In short, overfocusing on results and achievements above anything else runs the risk of Toxicity – one of the strongest predictors of turnover (Sull et al., 2022) – whereas overfocusing on relationships leads to Mediocrity, i.e., focusing on getting along as opposed to achieving excellence. Conversely, Bureaucracy arises from relying excessively on rules and adherence to established protocols (thus hindering innovation and growth), whereas uncontrolled risks and the absence of rules leads to Anarchy. Jointly, Grant refers to these extremes as "the four deadly sins" (Grant, 2022).

Company values

Many companies and organizations have formulated their company values and communicates them internally as a guiding tool for all employees in their interactions internally, with the company's clients and when they make decisions on behalf of the company.

These values may describe the company's existing cultural attributes, but they may also describe the values perceived as ideal from HR and top management to live up to the company's goals and as needed to fulfill the strategy, thus supplementary or complementary to the current company culture.

When hiring new employees based on their value congruence with company values, it is important to be mindful if that entails a supplementary or complementary fit and thereby the support needed for the new hire.

In development, the value congruence between company values and individual values for each employee can help guide the development needs of that employee and the support needed to engage in the perceived optimal company values.

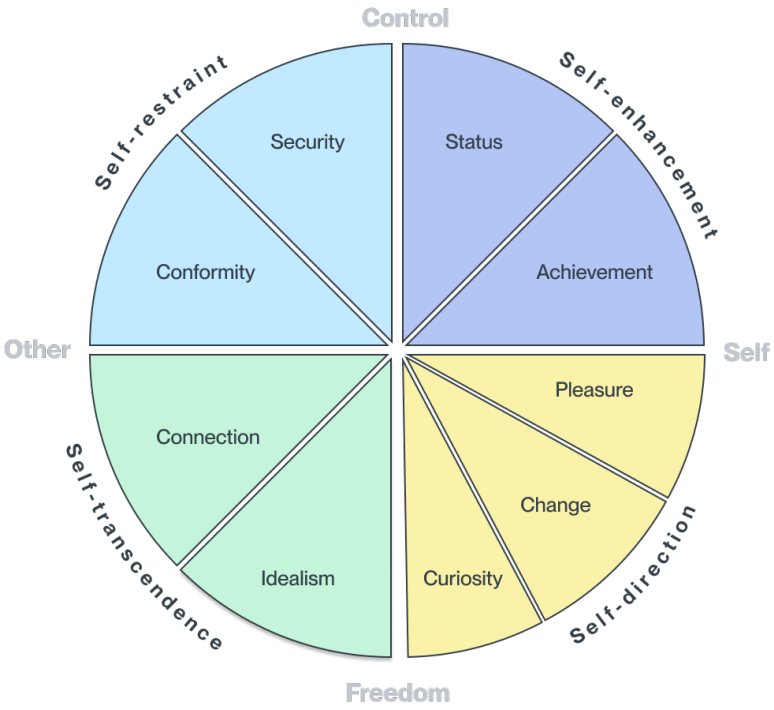
The MATCH-V model

MATCH-V is a contextualized, psychometric assessment measuring values inspired by Schartz' theory of basic human values. MATCH-V includes nine distinct, yet internally related values relevant to organizational characteristics, purpose, and culture.

The model is structured around a pattern of relations of conflict and congruence among values arranged in two conflicting dimensions and four congruent domains. Values opposing each other in the model includes conflicting beliefs, whereas values located in a single domain share the same broad motivational goal. Of course, people can and do pursue competing values, but not in a single act. They do so through different acts, in different situations.

In the first competing dimension, the model includes 'Self-enhancement' vs. 'Self-transcendence' values. This dimension captures the conflict between values that emphasize pursuit of one's own interests and relative success and dominance over others versus values that emphasize concern for the welfare and interests of others.

In the second conflicting dimension, we have 'Self-restraint' vs. 'Self-direction' values. This dimension captures the conflict between values that emphasize the need for stimulation, independence of thought, action, and feelings and readiness for change versus values that emphasize order, conservation of what is well-known, and preservation of social expectations or norms.



Self-Enhancement

In the self-enhancement domain, you will find the values of Status and Achievement. These values are based in qualities such as ambition, aspiration, and competition. People that prioritize values in this domain typically assign great weight to their personal results and tend to measure them in a traditional way, e.g. by means of personal recognition and prestige.

The Self-Enhancement domain consists of two values:

1) Status: Striving for power and influence
2) Achievement: Striving for results and acknowledgement

Self-Direction

In the self-direction domain, you will find the values of Pleasure, Change and Curiosity. These values are based in qualities such as stimulation, interest, and excitement. People that prioritize values in this domain typically require flexibility, appreciate alternative ways of thinking, and assign little value to tradition.

The Self-Direction domain comprises three values:

3) Pleasure: Valuing fun and informality
4) Change: Valuing novelty and innovation
5) Curiosity: Valuing immersion and new perspectives

Self-Transcendence

In the self-transcendence domain, you will find the values of Idealism and Connection. These values are based in qualities such as kindness, inclusiveness, and philanthropy. People that prioritize values in this domain typically assign great weight to goals that benefit multiple stakeholders and with a humanistic purpose.

In the Self-Transcendence domain, we find two values:

6) Idealism: Prioritizing altruism and inclusion
7) Connection: Prioritizing helpfulness and comradeship

Self-Restraint

In the self-restraint domain, you will find the values of Conformity and Security. These values are based in qualities such as stability, commitment, and respect. People that prioritize values in this domain typically assign great weight to ensuring harmony and refrain from actions or impulses likely to pose major risks or to upset others or violate social expectations and norms.

The Self-Restraint domain includes two values:

8) Conformity: Appreciating rules and standards
9) Security: Appreciating safety and insurance

Please note that compared to the original Schwartz framework, the value of Tradition has been omitted. These and other adaptations of the theory are elaborated in the section on Content validity.

3. Use of MATCH-V

Applications of the product

By obtaining scientifically validated data and unbiased insights, the MATCH-V can be used for multiple purposes at the individual, team, and organizational level:

- Understand individuals' value hierarchies. Knowing what drives the individual and what they value the most leads to a deeper understanding and management of employees.
- Improve employee selection decisions. By measuring individual values, you can assess candidates' fit to the team as well as organization, i.e., expanding from a sole focus on person-job fit to person-environment fit.
- Empower change through self-awareness. MATCH-V enables the individual to gain insight into their basic values, helping target effort in the most appropriate areas and identifying potential for personal development.
- Shape team composition and mitigate conflict potential. Although benefiting cognitive diversity, bringing together people with different values and value priorities can give rise to different conflicts. By knowing the composition of the team, managers can anticipate and mitigate conflicts within the team, fostering better team cohesion and cooperation.
- Understand the values of an organization. As Peter Drucker put it, "culture eats strategy for breakfast". In other words, even the best of business strategies will fail if the company culture is not suited for implementing and executing it. Hence, knowing the company's values is crucial to understanding company culture and fulfill strategic objectives.
- Measure organizational culture to ensure culture fit. An added benefit of measuring company values is that it paves the way for assessing culture fit. As previously mentioned, this fit can be supplementary as well as complementary, that is, hiring employees with values similar to or different from existing company values, respectively.

Interpretation of the scales

Below the nine values in MATCH-V are described. Interpretations of high and low scores are emphasized and hypotheses for the attitudes, goals, and behaviors the respondent is striving for depending on their individual score on each value is elaborated, focusing on the kinds of environments the respondent will thrive in, what they will contribute and to what they will respond negatively.

1. Status

The Status scale describes the extent to which individuals are striving for classical career goals, leading positions and being in charge, as opposed to valuing consensus and building equitable relationships. Individuals who score high on this scale can be described as someone who is ready to make decisions on other people's behalf and who is drawn to positions of power and who often has a strong desire to be recognized and gain social prestige. This person wants to get ahead rather than get along and supports and practices hierarchical relationships and structures at work. Individuals who score low on this scale can be described as someone who wants to fit in and get along with others and preserves relationships over self-interests. This person operates in a way that is more modest and inclusive to others and does things without seeking for authority. High scorers run the risk of being dominant, whereas low scorers risk losing influence in the company of high scorers.

From a cultural perspective, organizations characterized by high levels of status will often be hierarchical, top down driven, or prestige focused. Organizations with low levels of status can often be described as egalitarian, down to earth, and driven by consensus.

2. Achievement

The Achievement scale describes the extent to which individuals are striving for achieving ambitious goals, gaining mastery, and becoming an expert, as opposed to being satisfied, laid-back, and more interested in other pursuits. Individuals who score high on this scale can be described as someone who is goal-driven and ambitious and is intentional about how they invest their time and energy. This person enjoys challenges and pushing themselves and invests heavily into their expertise, quality of work, and deliverances. Individuals who score low on this scale can be described as someone who is driven to lead a more relaxed and satisfied life and is unlikely to overly challenge themselves unless motivated by reasons other than intrinsic need to improve. This person is happy to follow the direction and opinion of others and has interests beyond their career and professional development. High scorers run the risk of putting too much pressure on themselves and expect the impossible of themselves and their surroundings.

From a cultural perspective, organizations characterized by high levels of Achievement will often be driven, ambitious, and highly goal oriented. Organizations with low levels of Achievement can often be described as less result-driven, laid-back, and content.

3. Pleasure

The Pleasure scale describes the extent to which individuals are striving for enjoyment, fun, and playfulness, as opposed to being restrained, formal, and solely task-oriented. Individuals who score high on this scale can be described as someone who is fun, informal, light-hearted, and pursues joyful activities. This person has a balanced approach to work, insisting on leveling pleasure and effort. Individuals who score low on this scale can be described as someone who prefers to separate fun from work and who prioritizes professional appearance and appropriate behavior at all times. This person is usually focused on the task at hand and not easily distracted by other more fun and enjoyable opportunities. High scorers run the risk of insulting someone, whereas low scorers can be perceived as hard to integrate socially in an organization.

From a cultural perspective, organizations characterized by high levels of Pleasure will often be informal, playful, and social. Organizations with low levels of Pleasure can often be describes as serious, formal, and professional.

4. Change

The Change scale describes the extent to which individuals are striving for and seeking out change, novelty, and variety, as opposed to tradition, continuity, and familiarity. Individuals who score high on this scale can be described as someone who enjoys change and experimental environments and is comfortable in times of change and uncertainty. This person is willing to take risks and try new things and supports creativity and innovation at work. Individuals who score low on this scale can be described as someone who practices and upholds traditions and prefers predictability and certainty. This person analyzes potential consequences, prefers what is known to be true, and establishes stability in the organization. High scorers run the risk of initiating too

many new things without the proper foundation for them, whereas low scorers run the risk of slowing down the development of the company.

From a cultural perspective, organizations characterized by high levels of Change will often be spontaneous, flexible, and adaptive. Organizations with low levels of Change can often be describes as traditional, conventional and stable.

5. Curiosity

The Curiosity scale describes the extent to which individuals are striving for acquiring deep knowledge, understanding the world and continuously learning new things, as opposed to being practical about their own efforts and priorities, and focusing on what is truly necessary. Individuals who score high on this scale can be described as someone who is highly curious and values time to immerse and investigate a wide range of dispersed topics. This person enjoys widening their own perspective and likes to learn new things. Individuals who score low on this scale can be described as someone who prefers to rely on what they already know and understand and who adopts a practical approach to knowledge. This person needs good reasons to immerse and values knowing how their effort will benefit their current situation before accepting to invest the time and effort. High scorers run the risk of wasting time on irrelevant information, whereas low scorers run the risk of lacking perspective or overlooking avenues for refinement or improvement.

From a cultural perspective, organizations characterized by high levels of Curiosity will often be expert-oriented, knowledge-driven, and curious. Organizations with low levels of Curiosity can often be describes as practical, frugal, and narrow-minded.

6. Idealism

The Idealism scale describes the extent to which individuals are striving to live closely to their own moral principles, displaying altruism, and holding themselves and others to high social standards. This is opposed to insisting on accepting the harsh realities of the world, being realistic as to what is morally and ethically practical, and doing what is expedient. Individuals who score high on this scale can be described as someone who is altruistic, valuing a strong sense of purpose in their work. This person holds themselves and others to high social standards and has a clear sense of the right and wrong way of doing things. Individuals who score low on this scale can be described as someone who is likely to look after one's own and chooses to do what is profitable and efficient rather than what is the altruistic or sustainable way to do things. This person is willing to challenge the "right" attitudes and is likely to put their own and the organization's wellbeing before others'. High scorers run the risk of deprioritizing profitability whereas low scorers risk overlooking higher purposes.

From a cultural perspective, organizations characterized by high levels of Idealism will often be respectful, principled, and purpose driven. Organizations with low levels of Idealism can often be described as pragmatic, liberal, and willing to do what it takes to achieve its goals.

7. Connection

The Connection scale describes the extent to which individuals are striving to help others and work closely with other people, assigning great importance to the well-being of their colleagues, as opposed to being independent, self-reliant, and valuing clear and separate responsibilities.

Individuals who score high on this scale can be described as someone who enjoys working cooperatively and being part of a close-knit team. This person sees helping other people as part of their responsibility and may find it hard to say no, even when it might compromise their own priorities, getting caught up in other people's problems. Individuals who score low on this scale can be described as someone who works best independently and left to their own resources and prefers having their own time and space. This person puts their own goals and priorities in front of those of others'. High scores run the risk of neglecting financial targets in favor of individual considerations, whereas low scorers risk compromising shared results in favor of their own interests.

From a cultural perspective, organizations characterized by high levels of Connection will often be collaborative, team oriented, and helpful. Organizations with low levels of Connection can often be described as individualistic and focused on individual responsibilities.

8. Conformity

The Conformity scale describes the extent to which individuals are striving to comply to shared standards, building a fair work environment and ensuring alignment by managing processes and rules, as opposed to believing in individual freedom and influence as the foundation for business excellence, accepting misalignment or even chaos. Individuals who score high on this scale can be described as obedient, respectful, and compliant. This person expects everyone to follow the same rules and will feel offended by disobedience or if someone violate rules or social norms. Individuals who score low on this scale can be described as someone who is flexible and solution oriented, relying on common sense and what they think is right in any given situation. This person can be self-serving, ignoring rules or acting as if they do not apply to them. High scorers run the risk of being inflexible and inhibit personal initiative, whereas low scorers risk introducing anarchy.

From a cultural perspective, organizations characterized by high levels of Conformity will often be strict, rule-driven, structured, and bureaucratic. Organizations with low levels of Conformity can often be described as anarchistic, free, and chaotic.

9. Security

The Security scale describes the extent to which individuals are striving to limit uncertainty, risk, and rash decisions, promoting quality assurance, as opposed to valuing bravery, efficiency, and the courage to take risks. Individuals who score high on this scale can be described as cautious and is typically well-prepared, safe-guarding safety and quality. This person hates making mistakes and feels insecure when dealing with uncertainty and unpredictability. Individuals who score low on this scale can be described as someone who is willing to take risks and able to handle adverse situations calmly. This person may ignore mistakes or show up unprepared. High scorers run the risk of being overly cautious and risk aversive, whereas low scorers risk overlooking clear warning signs.

From a cultural perspective, organizations characterized by high levels of Security will often be cautious, compliant, and thoughtful. Organizations with low levels of Security can often be described as fastmoving, courageous, and unrestrained.

4. Scale construction

As described earlier, MATCH-V is a contextualized assessment of the Schwartz basic human values framework adapted to the working text. This section outlines the process of item development, data collection, item selection, and contextualization.

Item development

Based on the theory and research presented above, a large pool of items for each scale were produced by our internal team of psychologists and psychometricians. In this process, careful consideration was given to the relevance and appropriateness of items in a working context (i.e., items of a personal or intimate character were omitted). For each value, a number of facets were identified to capture all aspects of any given value based closely on Schwartz' descriptions of the values, which were then used to guide item development and check that each of the facets consistently mapped to the value in question.

In addition, item development was guided by the notion of Schwartz (2012) that values are linked to affect, refer to desirable end states or outcomes, and serve as standards or criteria by which to evaluate and select different courses of action, people, jobs, etc. To reflect these properties of basic human values, we used different types of items to measure each value as accurately as possible. These item types are described below (with examples in parentheses):

- Behavior & Motivation: Statements on what the person usually does or is motivated to do (e.g., trying something new).
- Attitudes & Beliefs: Statements containing opinions on what is right or wrong (e.g., respecting authorities).
- Ideals & Lifestyle: Statements about one's ideals and preferred ways of living/working (e.g., being the best or serving the greater good).
- Aversions: Statements that trigger dislikes, i.e., opposes the person's values (e.g., strict rules or boredom).
- Preferred colleagues: Statements describing the types of colleagues (or managers) the person prefers to work with (e.g., someone to have fun with).
- Work content: Statements referring to the kind of work one values (e.g., immersion or benefiting others)

The best items (in terms of wording as well as content) for each scale were then tested and reviewed in successive fashion, with each scale requiring a different number of iterations to reach the final scale.

In total, 486 items were produced, of which 307 were submitted to testing.

Data collection

Data on items submitted to testing was collected from October 2023 to March 2024. For all items, a minimum of 500 cases were collected prior to statistical analyses and psychometric scale validation. Demographic criteria were similar to those of the norm group, i.e., people aged 20-65 years of age who completed the test in a high-stake setting (selection and development) and in their native language.

Item selection

Upon data collection, items were subjected to a series of statistical analyses and reviewed by at least two reviewers from our internal team of psychologists and psychometricians. At the first stage, items were evaluated with respect to internal consistency (by means of Cronbach's alpha and corrected item-total correlations) as well as response distributions to detect lack of discriminatory power or socially desirable responding. Items were screened out if more than 90 % of answers were in the same direction (i.e., response options 1 and 2 or 3 and 4).

Items were selected based on the corrected item-total correlations using a top-down procedure, which has proven one of the best methods for item selection (Zijlmans et al., 2019). In this procedure, all items are included at the first step. Based on the corrected item-total correlations, weakly correlating items are then removed successively, until the scale can no longer be improved (i.e., the point at which the exclusion of any item would not enhance the internal consistency of the scale). Although higher values were desired for the final scales, a guiding cut-off for the correlation were set at .20 during development phases to avoid discarding items with relevant content and sufficient psychometric quality.

Final scale validation

Following the various number of development iterations, each of the preliminary scales were then subjected to an extensive psychometric validation procedure assessing five different aspects:

1) Consistency
2) Unidimensionality
3) Local independence
4) Item invariance (no Differential Item Functioning or DIF)
5) Scale characteristics

Each of these aspects is elaborated in the sections below on construct validity.
In addition to the psychometric validation, the final scales were also examined for qualitative aspects such as readability, comprehensibility (negations), grammatical redundancy, and balancing the number of positively and negatively worded (reversed keyed) items.

In total, 85 items were included with each scale consisting of 7-10 items.

5. Validity

Despite a clear definition of validity, it is a topic of debate among international researchers and experts, how many types of validity there are, and which research methods are most suitable for shedding light on what. 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 subjects perceive the questionnaire and test results as relevant, comprehensive, and reflective of reality. Face validity is thus about whether a test comes across as credible to the test person, which is important to ensure that a test person is sufficiently motivated to participate in the test and accept the conclusions drawn from it. It is also about recognizability of test results to both the test person and to 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 items with less recognizable relevance (which is often the consequence of using a solely data-driven item selection process such as empirical keying). The use of mainly transparent items has the advantage of providing a clear link between the test subject's answers to the individual items and the final scores derived from them. Hence, the likelihood of the test subject understanding, recognizing, and accepting test 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, to ensure acceptance on behalf of the test taker, careful consideration was given in defining and labelling each of the different values. A key emphasis was placed on reflecting the actual content of the individual items (i.e., being honest about what the items actually measure) whilst refraining from focusing solely on the positive or negative aspects of the attribute in question.

The relevance of items is demonstrated in Table 1 with item examples alongside the definitions of each of the MATCH-V scales. Items with the strongest correlation to the scale were chosen as examples (please note that for two scales, these items were reversed keyed, i.e. opposing the value in question).

Table 1. Definitions and item examples for MATCH-V scales.

Scale	Definition	Item example
Status	Striving for power and influence	"I really enjoy being in charge"
Achievement	Striving for results and acknowledgement	"It's unambitious if you don't strive for perfection"
Pleasure	Valuing fun and informality	"I do what I can to create a fun atmosphere at work"
Change	Valuing novelty and innovation	"I would be comfortable with a job that requires me to take risks"
Curiosity	Valuing immersion and new perspectives	"I can't think of anything better than having to learn a lot of new things"
Idealism	Prioritizing altruism and inclusion	"It's not my problem that some people are less privileged than me" (R)
Connection	Prioritizing helpfulness and comradeship	"Sometimes I'm so preoccupied with other people's needs that it compromises achieving my own goals"
Conformity	Appreciating rules and standards	"I would rather use my common sense than follow a rigid set of rules" (R)
Security	Appreciating safety and insurance	"I get anxious when I don't know what's going to happen"

Content validity

Content validity concerns whether test items and scales constitute a relevant, representative sample of the aspects that define the theoretical concept (domain) being measured. As previously mentioned, MATCH-V is based on the framework of basic human values by Schwartz (2012). The original definitions and corresponding MATCH-V scales are listed below in Table 2.

To have a contextualized assessment suitable for work-related purposes, a few adaptations were made to the original framework. First, the value of Tradition was omitted as the emphasis on religious values were deemed inappropriate in a working and selection context. In addition, respect and acceptance of cultural ideas and customs overlaps greatly with what is already measured in Conformity. Interestingly, in the original Schwartz model, Conformity and Tradition is placed in the same sector (part) of the circle, suggesting a greater conceptual overlap than any of the other values in the model, with "subordination of self in favor of socially imposed expectations" as the common denominator. Furthermore, Schwartz (2012) points out that the "circular arrangement of values represents a motivational continuum" and "dividing the domain of value items into ten distinct values is an arbitrary convenience" (p. 10). In other words, all values could just as well be arranged into the four broader domains or even divided into 19 more narrowly defined values as done in Schwartz' later work. As a final argument, Tradition is placed in the right side of the sector as this value is in greater opposition to Stimulation (termed Change in MATCH-V) than is Conformity. That is, most aspects of Tradition are represented in MATCH-V by a combination of high scores on Conformity and low scores on Change.

Regarding Pleasure (termed Hedonism by Schwartz), several aspects relevant to a measure of personal gratification (such as eating, drinking, partying or even sexual behaviors) were deemed

unacceptable to ask about in a working context. Instead, a range of different more work-related facets were identified and tested in the development process. These facets include valuing aesthetics, having fun with colleagues at social events, being joyful and spreading a good mood in the workplace, being informal and joking with colleagues (sometimes being somewhat inappropriate), prioritizing fun over duties, and valuing self-interest, i.e. avoiding tasks or job-related activities that the person finds boring. From the perspective of culture fit, the aspects of having fun and being informal with colleagues and spreading a good mood (versus acting more serious, formal, and professional) were deemed the most relevant, as scores on opposite ends of this scale pose a great risk of conflicts and culture mis-fit. Furthermore, these were the aspects most clearly related to values and separable from other constructs such as personality (e.g., Extraversion). Statistical analyses confirmed that these facets could be meaningfully combined into a unidimensional and consistent scale, which was not the case for the other facets tested.

Finally, Self-direction seems the most comprehensive and weakly defined value in the Schwartz model, relating to aspects of both freedom ("independent thought and action") and immersion ("choosing, creating, exploring"). Statistical analyses conducted in the early stages of scale development showed that these aspects represented distinct concepts rather than a single, unitary construct. Therefore, the Curiosity scale was operationalized to capture the aspect of immersion and being curious about new perspectives. The aspect of freedom, defined as valuing independence and breaking free from strict social rules, standards, and protocols, can be found in low scores on Conformity (notably, this value is diametrically opposite the Self-direction domain in the Schwartz model).

Table 2. Adaptation to the Schwartz model of basic human values.

Value	Schwartz definition	MATCH-V scale
Power	Social status and prestige, control or dominance over people and resources	Status
Achievement	Personal success through demonstrating competence according to social standards	Achievement
Hedonism	Pleasure or sensuous gratification for oneself	Pleasure
Stimulation	Excitement, novelty, challenge in life	Change
Self-direction	Independent thought and action - choosing, creating, exploring	Curiosity
Universalism	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature	Idealism
Benevolence	Preserving and enhancing the welfare of those with whom one is in frequent personal contact (the 'in-group')	Connection
Tradition	Respect, commitment, and acceptance of the customs and ideas that one's culture or religion provides	-
Conformity	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms	Conformity
Security	Safety, harmony, and stability of society, of relationships, and of self	Security

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 is also affected by the number of items in the scale (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 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).

Cronbach's alphas, average corrected item-total correlations, and average inter-item correlations for each scale are listed below in Table 3.

All scales have sufficient or excellent levels of consistency, with alphas ranging from .70 to .81 with an average of .74. For any scale, a minimum requirement of .70 was set for alpha, as this is considered adequate by EFPA (2013). In addition, average corrected item-total correlations are all above the desired value of .30 (He & Wang, 2015; Shen et al., 2018). Furthermore, all inter-item correlations are positive but not too high, suggesting high internal consistency but minimal redundancy of items.

Table 3. Internal consistency of MATCH-V scales.

Scale	Alpha	Item-Rest cor.	Inter-item cor.
Status	.81	.55	.37
Achievement	.70	.37	.19
Pleasure	.78	.47	.30
Change	.72	.40	.21
Curiosity	.70	.39	.22
Idealism	.72	.38	.20
Connection	.73	.40	.24
Conformity	.73	.41	.26
Security	.78	.49	.31
Mean	.74	.43	.26
Median	.73	.40	.24

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.
An example Scree plot from the Status scale is shown below in Figure 1. Table 4 displays the Eigenvalue and amount of explained variance (%) for the single factor solution alongside the items' average factor loading for each scale.

Across scales, the single factor extracted explained a minimum of 29.5 % of the total variance, and average factor loadings were strong, ranging from .53 to 68. For each scale, the factor loading of any item exceeded .30 with several items having a loading of more than .40, which is well above the values recommended for sufficient primary factor loadings (Howard, 2016). In sum, these analyses support the unidimensional nature of the constructed scales.

Figure 1. Example Scree plot from the Status scale.

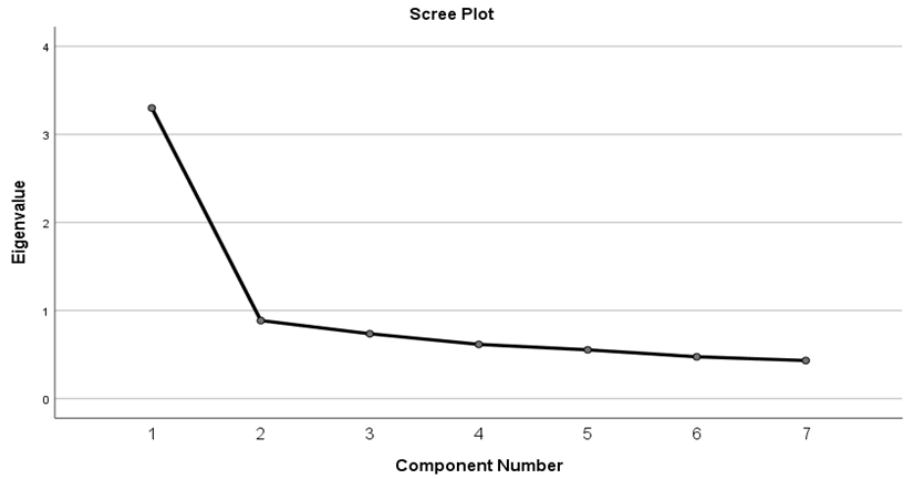


Table 4. Unidimensionality of MATCH-V scales.

Scale	Eigenvalue	Expl. Var. (%)	Factor loading
Status	3.30	47.1	.68
Achievement	2.69	29.9	.54
Pleasure	3.34	37.1	.60
Change	2.82	31.3	.56
Curiosity	2.57	32.1	.56
Idealism	2.95	29.5	.53
Connection	2.86	31.8	.56
Conformity	2.95	32.8	.56
Security	3.23	40.4	.63
Mean	2.97	34.7	.58
Median	2.95	32.1	.56

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.

Local independence was inspected using partial correlations, where each of the items are correlated whilst controlling for the scale score (i.e., the sum of items). Hence, partial correlations are expected to be weaker than inter-item correlations and close to zero or negative (van Bork et al., 2018).

As the magnitude of partial correlations depends on both the number of items in the scale, the sample size, and the number of response options, it is difficult to set fixed criteria as to what constitutes a too high partial correlation when considered in isolation. Therefore, the largest observed partial correlation was subtracted from the average partial correlation to obtain a relative estimate of local independence similar to the Q3* measure used within Rasch and IRT models (Christensen et al., 2017).

The average and maximum partial correlations as well the maximum differences for each of the scales are displayed in Table 5. Across scales, the difference between the maximum and average partial correlation ranged from .14 to .25 showing sufficient local independence for each of the scales as suggested by the .20 or .30 cut-off, which was used as the guiding principle (Christensen et al., 2017). This ensures that the previously reported alphas are realistic estimates of the scales' reliabilities as they have not been inflated by redundancy of items in the scales.

Table 5. Local independence of MATCH-V scales.

Scale	Avg. partial cor.	Max. partial cor.	Max. difference
Status	-.15	.04	.20
Achievement	-.11	.08	.19
Pleasure	-.12	.13	.25
Change	-.14	.11	.25
Curiosity	-.15	.03	.18
Idealism	-.11	.04	.15
Connection	-.12	.06	.18
Conformity	-.10	.05	.14
Security	-.14	.10	.24
Mean	-.13	.07	.20
Median	-.12	.06	.19

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, age, test purpose (recruitment vs. development), or job level (managers vs. employees). 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).

Several methods to investigate DIF have been proposed, framed within Classical Test Theory (CTT) as well as Item Response Theory (IRT) (Rouquette et al., 2019; Woods, 2009). The use of partial correlations has the advantage of being easy to compute, requiring modest sample sizes for subgroups, and having a straightforward interpretation in terms of effect sizes (Stricker, 1982; Conoly, 2003).

A range of criteria to detect DIF has been suggested, none of which are universally accepted. For instance, the use of significance tests suffers the drawback that p-values become smaller as the sample size increases, in which case negligible DIF can be flagged for significance, or substantial DIF can be overlooked in small samples. For this reason, a combination of significance and correlation magnitude was used to exclude items with substantial DIF. In addition, any demographic variable should explain less than 5 % of the variance in any item when accounting for the scale score (as proposed by Scott et al., 2010).

Table 6 shows the absolute average partial correlation of items with each of the demographic variables (gender, age, purpose, job level). For all scales and demographic variables, absolute partial correlations were low (< .10), and maximum partial correlations ranged from .05 to .19 equivalent to an explained variance of just 3.6 %, which is well below the suggested criterion of 5 %. In sum, these analyses show that the MATCH-V provides an unbiased assessment of values applicable across job levels in both a selection and development setting.

Table 6. Item invariance of MATCH-V scales.

Scale	Gender	Age	Purpose	Job level
Status	.06	.06	.04	.06
Achievement	.05	.06	.04	.09
Pleasure	.05	.06	.03	.03
Change	.05	.09	.04	.07
Curiosity	.06	.05	.04	.04
Idealism	.10	.05	.04	.06
Connection	.07	.08	.03	.06
Conformity	.04	.06	.07	.08
Security	.06	.07	.04	.03
Mean	.06	.06	.04	.06
Median	.06	.06	.04	.06

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.

The results of these analyses are listed in Table 7 alongside the sample size (N) used to validate the final scales. Across scales, the mean and median scale score were almost identical reflecting that scales had negligible amounts of skewness. In addition, the absolute values for both skewness and kurtosis were well below the typically used critical values of 2 (or even 1) and 4, respectively (Kim, 2013; Chissom, 1970). When using these scores to calculate the final normed scores (C scores), no observed proportions were allowed to deviate more than 5 % from the expected percentages.

Table 7. Scale characteristics of MATCH-V scales.

Scale	N	Mean	Median	SD	Skew.	Kurt.
Status	717	16.7	17	3.43	-0.08	-0.02
Achievement	544	21.9	22	3.63	0.43	0.69
Pleasure	698	24.8	25	3.74	0.15	0.14
Change	512	23.4	23	3.27	0.22	0.24
Curiosity	570	21.4	21	3.30	0.47	0.26
Idealism	554	28.0	28	3.81	-0.07	0.67
Connection	598	20.8	21	3.30	0.21	0.64
Conformity	587	23.9	24	3.41	-0.04	0.55
Security	1072	19.7	20	3.43	-0.12	0.50

Criterion validity

Criterion validity refers to the relationship between test results and information about test subjects derived from other sources (i.e., external criteria).

Convergent & Discriminant validity

Evidence of convergent validity is obtained by demonstrating relationships (positive correlations) with other measures of similar or related constructs. Conversely, discriminant validity is indicated by a lack of relationships (near-zero correlations) with non-similar or unrelated constructs.

The convergent and discriminant validity was studied extensively for the first version of MATCH-V. For an elaboration and detailed analyses of these results, please refer to the first version of the technical manual. A subset of the analyses is presented below in Table 8 (with relevant correlations highlighted in bold) comparing each of the seven original MATCH-V scales (Status, Achievement, Pleasure, Change, Curiosity, Connection, Integrity) to each of the Schwartz values by means of the 45-item Schwartz Basic Values Inventory (Schwartz, 1992).

Although most scales showed acceptable or even excellent levels of convergent validity (e.g., Status, Achievement, Pleasure, Change, Curiosity, and Connection), some scales had unexpected strong correlations with non-similar values (e.g., Power/Achievement and Stimulation/Pleasure),

thus showing a lack of discriminant validity. Furthermore, all of the pairwise correlations were positive, thus violating theoretical notions of conflicting or opposing values (e.g., Status/Connection and Change/Tradition).

To tackle these and other issues related to validity, a major revision of MATCH-V has been carried out with the following main purposes:

- Adding Conformity and Security as separate scales, as neither of these are very well captured by any of the existing scales (neither in isolation nor in combination).
- Renaming Integrity to Idealism to capture the value of Universalism more clearly and separate it from Curiosity.
- Revising item content to distinguish Idealism (Universalism) from Connection (Benevolence), Pleasure (Hedonism) from Change (Stimulation), and Status (Power) from Achievement.

New validity studies are currently being prepared. Once conducted, the technical manual will be updated accordingly.

Table 8. Correlations between Schwartz basic values and MATCH-V scales.

Schwartz value	STA	ACH	PLE	CHA	CUR	CON	INT
Power	.80	.66	.67	.64	.35	.58	.49
Achievement	.62	.64	.63	.54	.46	.62	.58
Hedonism	.38	.42	.50	.45	.44	.43	.48
Stimulation	.60	.55	.71	.67	.55	.55	.52
Self-direction	.04	.23	.27	.26	.57	.24	.45
Universalism	.14	.31	.33	.34	.62	.41	.64
Benevolence	.19	.37	.37	.28	.50	.51	.69
Tradition	.40	.54	.41	.39	.37	.58	.68
Conformity	.36	.44	.43	.35	.37	.51	.60
Security	.37	.47	.48	.39	.39	.47	.55

Note. N = 217. STA = Status, ACH = Achievement, PLE = Pleasure, CHA = Change, CUR = Curiosity, CON = Connection, INT = Integrity

Group differences

The validity of MATCH-V is further supported by group differences that replicate previous findings in the literature on basic human values. Group differences on MATCH-V scales for gender (male/female), age (above/below 40), purpose (development/selection), and job level (employee/manager) are presented below in Tables 8-11.

For gender and age, standardized mean differences (Cohen's d) were generally small (< .50) or negligible (< .20). On average, there was small differences in favor of males on Status, Achievement, and Change and in favor of females on Idealism, Connection, and Conformity (and trends towards slightly higher scores for females on Pleasure, Curiosity, and Security). Regarding age, the most notable differences were that younger individuals scored slightly higher than people above 40 on the values of Achievement, Pleasure, Curiosity, and Security.

These differences are well in line with previous research showing consistent, modest differences on gender and age across a wide range of culturally diverse groups. Furthermore, gender and age tend to have lower effects sizes than other demographics such as education and country (culture), especially. Previous research has also shown that women generally rate self-transcendence (Universalism, Benevolence) and self-restraint values (Tradition, Conformity, Security) the highest and vice versa for men (Struch et al., 2014; Schwartz & Rubel, 2005). However, please note that these findings are based on studies of value priorities and the original ten values defined by Schwartz, which might explain a few discrepancies between these and previous results (i.e., adaptations made to Pleasure and Curiosity and the omission of Tradition). Similarly, research has shown that older individuals tend to prioritize Self-transcendence and Self-restraint values more than younger individuals (and vice versa for values within the domains of Self-enhancement and Self-direction) (Sagiv & Schwartz, 2022).

Regarding purpose (selection vs. development), the most notable differences were slightly higher mean scores in selection compared to development for Curiosity, Idealism, and Conformity (and slightly lower scores on Status and Achievement), which might point to the socially desirable characteristics of these values. However, these differences are not sufficiently large as to warrant separate norms for selection and development, as long as the norm group has a sufficient proportion of tests completed for selection purposes.

As for job level, managers had significantly higher scores on Status and lower scores on Conformity and Security. These differences are unsurprising in light of the fact that work conditions have been shown to affect value priorities (Sagiv & Schwartz, 20212). In addition, people who value Status (i.e., having control and being in charge) is probably more likely to pursue management positions than others.

Table 9. Group differences on gender (male/female) for MATCH-V scales.

Scale	Male			Female			Comparison		
	N	M	SD	N	M	SD	Dif.	t	d
Status	422	17.1	3.36	295	16.1	3.46	1.0	3.81*	0.29
Achievement	276	22.2	3.88	268	21.6	3.33	0.6	2.04	0.17
Pleasure	379	24.8	3.73	319	24.9	3.76	-0.1	-0.46	0.04
Change	286	5.2	1.96	226	4.8	1.95	0.4	2.32	0.21
Curiosity	326	21.2	3.32	244	21.7	3.26	-0.5	-1.85	0.16
Idealism	304	27.3	3.78	250	29.0	3.64	-1.7	-5.29*	0.45
Connection	330	20.4	3.34	268	21.1	3.23	-0.7	-2.52	0.21
Conformity	344	23.7	3.54	243	24.3	3.19	-0.6	-2.07	0.17
Security	580	19.5	3.32	492	19.9	3.55	-0.3	-1.62	0.10

*p < .001

Table 10. Group differences on age (above/below 40) for MATCH-V scales.

Scale	40+			< 40			Comparison		
	N	M	SD	N	M	SD	Dif.	t	d
Status	285	16.4	3.45	432	16.9	3.40	-0.6	-2.16	0.16
Achievement	229	21.3	3.22	315	22.3	3.85	-1.0	-3.16	0.27
Pleasure	314	24.3	3.55	384	25.3	3.83	-1.0	-3.43	0.26
Change	207	4.9	1.99	305	5.1	1.94	-0.2	-0.96	0.09
Curiosity	253	21.0	2.93	317	21.7	3.54	-0.7	-2.58	0.22
Idealism	243	28.1	3.45	311	28.0	4.06	0.1	0.30	0.03
Connection	242	20.4	3.16	356	21.0	3.38	-0.6	-2.17	0.18
Conformity	241	23.9	3.46	346	23.9	3.38	0.0	0.17	0.01
Security	477	18.9	3.43	595	20.3	3.32	-1.3	-6.39*	0.39

*p < .001

Table 11. Group differences on purpose (development/selection) for MATCH-V scales.

Scale	Development			Selection			Comparison		
	N	M	SD	N	M	SD	Dif.	t	d
Status	69	17.8	3.59	648	16.6	3.39	1.3	2.94	0.37
Achievement	27	23.3	4.82	517	21.8	3.55	1.5	2.12	0.42
Pleasure	51	24.6	3.40	647	24.8	3.77	-0.2	-0.36	0.05
Change	32	5.3	2.34	480	5.0	1.94	0.4	1.01	0.18
Curiosity	42	20.2	3.33	528	21.5	3.28	-1.4	-2.57	0.41
Idealism	38	26.5	3.71	516	28.1	3.79	-1.7	-2.63	0.44
Connection	55	21.1	3.77	543	20.7	3.25	0.4	0.77	0.11
Conformity	86	22.3	3.52	501	24.2	3.31	-1.9	-4.98*	0.58
Security	86	19.3	3.54	986	19.7	3.42	-0.4	-1.04	0.12

*p < .001

Table 12. Group differences on job level (employees/managers) for MATCH-V scales.

Scale	Employees			Managers			Comparison		
	N	M	SD	N	M	SD	Dif.	t	d
Status	460	16.0	3.42	101	19.0	2.67	-3.0	-8.37*	0.92
Achievement	333	22.1	3.68	61	22.4	3.80	-0.4	-1.47	0.10
Pleasure	446	25.0	3.86	90	24.7	3.40	0.3	0.65	0.08
Change	324	4.8	1.92	69	5.3	1.95	-0.5	-1.86	0.25
Curiosity	395	21.5	3.24	88	20.6	3.21	0.9	-1.77	0.27
Idealism	338	27.9	3.95	76	27.4	3.26	0.5	1.00	0.13
Connection	395	20.7	3.40	65	21.5	2.97	-0.8	-1.77	0.24
Conformity	331	24.3	3.24	117	22.9	3.80	1.4	3.81*	0.41
Security	617	20.2	3.40	203	18.4	3.29	1.8	6.66*	0.54

*p < .001

6. 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 MATCH-V scales in Table 12. The final column of the table contains the Standard Error of Measurement (SEM) defined as:

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

Where SD represents the standard deviation and r refers to the reliability of the scale in question. As shown in Table 13, all scales have acceptable or excellent levels of reliability with alphas ranging from .70 to .81 and an average of .74. Given the dispersion and reliabilities of the scales, the average Standard Error of Measurement is only 1.77. If this is used to construct a 95 % confidence interval, the true score (T) would most likely fall within a range of no more than ±4 points from the observed score (O). When normed and converted to C-scores (with an SD of 2), the average SEM is only 1 C-score point as shown in the right most column of Table 13.

In sum, these estimates show great consistency and measurement accuracy across MATCH-V scales.

Table 13. Reliability of MATCH-V scales.

Scale	Items	Alpha	SD	SEM	SEM (C score)
Status	7	.81	3.43	1.50	0.87
Achievement	9	.70	3.63	1.99	1.10
Pleasure	9	.78	3.74	1.75	0.94
Change	9	.72	3.27	1.73	1.06
Curiosity	8	.70	3.30	1.81	1.10
Idealism	10	.72	3.81	2.02	1.06
Connection	9	.73	3.30	1.71	1.04
Conformity	9	.73	3.41	1.77	1.04
Security	8	.78	3.43	1.61	0.94
Mean	8.7	.74	3.48	1.77	1.01
Median	9.0	.73	3.43	1.75	1.04

7. 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. MATCH-V 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. The normative test not only provides answers to what is characteristic of the person as an individual, but also what is characteristic of the test person in relation to others. Normative tests thus measure interpersonal differences (i.e., 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 (in this case; culture fit) 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 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, which is displayed as the results. The interpretations, z-score ranges, percentages, and percentiles for C scores are shown in Table 14.

Table 14. Interpretation of C scores.

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 events and crises have an impact on the way people in general behave and new generations may also challenge the existing standards. Therefore, with respect to assessments, it is highly important to update norm groups at a regular basis and make sure that all candidates and people assessed are evaluated with a norm group representing the current state and what is currently considered normal behavior, since that will provide the most valid assessment. In addition, updating the norm group keeps scores balanced and avoid too many candidates getting either high or 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 are committed to checking if updates are needed at least every 2 years and update our norm groups frequently.
As the current manual reflects a major revision of the first version of the product, the initial norm group for MATCH-V is a research-based norm group based on data collected in a high-stake setting (selection and development) from October 2023 to March 2024. Once more data is collected, the norm group will be updated accordingly.

Sample size

A good norm group consists of many people, as a high number provides greater representation and statistical certainty. The prevailing 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).

As data for the revised MATCH-V scales were collected in succession, the sample size varies slightly between different scales as shown below in Table 14, ranging from 512 to 1,072 with a mean sample size of 650. For all scales, this represents a good sample size, even in high-stake decisions, according to both EFPA and COTAN guidelines.

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 weighted and balanced, especially those that can lead to potential score differences between subgroups.

The initial, research-based norm group for MATCH-V consists of people aged 18-70 who completed the assessment in a high-stake setting (selection and development) and in their native language (Danish). As statistical analyses showed mostly negligible or small group differences for gender, age, test purpose, and job level, the norm group was not further stratified for any of these demographic variables, as this would only reduce the sample size without impacting overall scores across groups. Importantly, the majority of cases in the norm group represents the test purpose of selection (on average 92 %) and the job level of employee (on average 81 %), which reflects expected proportions in real-life applications of MATCH-V and the fact that these demographic variables seem to have a larger impact on scores than gender and age.

The demographic composition of the norm group for each is listed below in Table 15.

Table 15. Demographic composition of the norm group for MATCH-V.

Scale	N	Male (%)	< 40 (%)	Selection (%)	Employee (%)
Status	717	58.9	60.3	90.4	82.0
Achievement	544	50.7	57.9	95.0	84.5
Pleasure	698	54.3	55.0	92.7	83.2
Change	512	55.9	59.6	93.8	82.4
Curiosity	570	57.2	55.6	92.6	81.8
Idealism	554	54.9	56.1	93.1	81.6
Connection	598	55.2	59.5	90.8	85.9
Conformity	587	58.6	58.9	85.3	73.9
Security	1,072	54.1	55.5	92.0	75.2
Mean	650	55.5	57.6	91.8	81.2
Median	587	55.2	57.9	92.6	82.0

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 (AI), 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 AI. Usually, a selection rate for any demographic group less than four-fifths (or 80 percent) of the selection rate for the group with the highest rate (majority group) is considered evidence of AI. The level of AI 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. A first step in preventing Adverse Impact is to ensure that differences between demographic groups reflect true differences (item impact) as opposed to item bias (or Differential Item Functioning or DIF). In other words, 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, why this is still relevant and important to study further.

Second, simulations of expected AI are conducted at different selection rates for gender (males/females) and age (above/below 40). Please note, however, that these calculations are based on the assumptions that 1) candidates are selected based on only a single score, 2) the assessment is used as the sole basis for selection and 3) a fixed selection rate is applied (i.e., hiring everyone that meets 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 16 and 17 list the standardized mean difference (Cohen's d) between groups alongside the simulated AI ratio (selection rate of the least represented group compared to the most represented group) for gender (male/female), and age (above/below 40), 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.

The results show that when applying strict selection ratios for any scale, the AI ratio for many scales is below the .80 cut-off suggested by the four-fifths rule and so should be avoided altogether. However, for lenient selection rates, the selection rate of the least represented group is no less than 80 % of the selection rate of the most represented group, except for Idealism which has an AI ratio of .75 for gender (favoring females). When using this scale for selection, careful consideration should be given to avoid fixed selection ratios (even if lenient) and combine scores on this scale with other criteria to balance out the level of Adverse Impact, thus preventing any discrimination on gender (in this case discriminating men and favoring women).

To prevent any adverse impact on age, strict selection ratios should definitely be avoided, and moderate selection ratios applied with caution for Achievement and Security, especially if these scales are used in combination (as group differences on both favor individuals below the age of 40). However, for lenient selection ratios, the AI ratio meets the .80 cut-off for all of the scales when it comes to age differences.

Although these simulations cannot be meaningfully carried out for group differences related to test purpose and job level, one should be mindful of selecting or promoting managers based on Status, as people who already occupy or are applying for management positions tend to have higher scores (as reported in Table 11). Hence, selecting or promoting managers based on Status runs the risk of "reproducing" existing management ideals, i.e. selecting people who value hierarchies and are drawn to positions of power, which is not a prerequisite needed for acquiring, let alone performing in management positions (especially not in company cultures emphasizing equality and flat hierarchies).

In conclusion, when applying proper selections ratios and decision rules (i.e., combining (multiple) scores with information derived from other sources), MATCH-V provides a fair and unbiased assessment that does not cause any Adverse Impact for protected groups when used for making employment decisions.

Table 16. Adverse Impact (AI) ratios for gender at different selection ratios (SR) for MATCH-V scales.

Scale	d	Strict SR	Moderate SR	Lenient SR
Status	0.29	0.74	0.76	0.82
Achievement	0.17	0.70	0.83	0.93
Pleasure	0.04	0.99	0.95	1.00
Change	0.21	0.82	0.83	0.79
Curiosity	0.16	0.77	0.88	0.82
Idealism	0.45	0.59	0.65	0.75
Connection	0.21	0.74	0.83	0.89
Conformity	0.17	0.78	0.88	0.86
Security	0.10	0.77	0.82	0.97

Table 17. Adverse Impact (AI) ratios for age at different selection ratios (SR) for MATCH-V scales.

Scale	d	Strict SR	Moderate SR	Lenient SR
Status	0.16	0.88	0.91	0.88
Achievement	0.27	0.55	0.71	0.85
Pleasure	0.26	0.68	0.74	0.86
Change	0.09	0.95	0.94	1.00
Curiosity	0.22	0.80	0.74	0.90
Idealism	0.03	0.92	1.00	0.99
Connection	0.18	0.70	0.79	0.90
Conformity	0.01	0.97	0.96	0.90
Security	0.39	0.58	0.69	0.80

8. References

Achor, S., Reece, A., Kellerman, G., & Robichaux, A. (2018). 9 Out of 10 People Are Willing to Earn Less Money to Do More-Meaningful Work. Retrieved October 5, 2019, from Harvard Business Review website: https://hbr.org/2018/11/9-out-of-10-people-are-willing-to-earn-less-money-to-do-more-meaningful-work?utm_source=facebook&utm_medium=social&utm_campaign=hbr.

Age Discrimination in Employment Act of 1967. Pub. L. No. Pub. L. No. 90-202, et seq (1967).

Akhtar, R., Humphreys, C., & Furnham, A. (2015). Exploring the relationships among personality, values, and business intelligence. Consulting Psychology Journal, 67(3), 258–276.

Ashforth, B. E., & Mael, F. (1989). Social Identity Theory and the Organization. Academy of Management Review, 14(1), 20–39.

Bao, Y., Dolan, S. L., & Tzafrir, S. S. (2012). Value Congruence in Organizations: Literature Review, Theoretical Perspectives, and Future Directions. SSRN Electronic Journal, September (239).

Belbin, R. M. (2017). Team roles at work. Taylor & Francis Ltd. Second Edition.

Bersin, J., Geller, J., Wakefield, N., & Walsh, B. (2016). Human Capital Trends 2016. Retrieved from https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2016/human-capital-trends-introduction.html.

Chamberlain, A. (2015). Does Company Culture Pay Off? Analyzing Stock Performance of "Best Places to Work" Companies - Glassdoor Economic Research. Retrieved from https://www.glassdoor.com/research/studies/does-company-culture-pay-off-analyzing-stock-performance-of-best-places-to-work-companies/.

Chatman, J. A. (1989). Improving Interactional Organizational Research: A Model of Person-Organization Fit. Academy of Management Review, 14(3), 333–349.

Chissom, B. S. (1970). Interpretation of the Kurtosis Statistic. The American Statistician, 24(4), 19-22.

Christensen, K. B., Makransky, G. and Horton, M. C. (2017). Critical Values for Yen's Q3:Identification of Local Dependence in the Rasch model using Residual Correlations. Applied Psychological Measurement, 41(3), 178-194.

Conoly, C. A. (2003). Differential Item Functioning in the Peabody Picture Vocabulary Test – Third Edition: Partial Correlation versus Expert Judgment. Doctoral Dissertation: Texas A&M University.

Cornelissen, J. P., Haslam, S. A., & Balmer, J. M. T. (2007). Social Identity, Organizational Identity and Corporate Identity: Towards an Integrated Understanding of Processes, Patternings and Products. British Journal of Management, 18(s1), S1–S16.

De Cooman, R., Gieter, S. De, Pepermans, R., Hermans, S., Bois, C. Du, Caers, R., & Jegers, M. (2009). Person–organization fit: Testing socialization and attraction–selection–attrition hypotheses. Journal of Vocational Behavior, 74(1), 102–107.

Deci, E. L., & Ryan, R. M. (2010). Intrinsic Motivation. In The Corsini Encyclopedia of Psychology (pp. 1–2).

Deci, E. L., & Ryan, R. M. (2012). Self-Determination Theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), Handbook of Theories of Social Psychology: Volume 1 (pp. 416–437).

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.

Equal Employment Opportunity Commission, Civil Service Commission, U.S. Department of Labor, & U. S. D. of J. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43, 38290–38309.

Furnham, A. (2001). Vocational Preference and P-O Fit: Reflections on Holland's Theory of Vocational Choice. Applied Psychology, 50(1), 5-29.

Gallup. (2019). Employee Engagement on the Rise in the U.S. Retrieved October 5, 2019, from Gallup 32 website: https://news.gallup.com/poll/241649/employee-engagement-rise.aspx?utm_source=link_wwwv9&utm_campaign=item_245786&utm_medium=copy

Grant, A. (2022). The 4 Deadly Sins of Work Culture. TED talks (transcript). Retrieved June 5, 2024 from website: https://www.ted.com/podcasts/worklife/the-4-deadly-sins-of-work-culture-transcript.

George, J. M., & Jones, G. R. (1996). The experience of work and turnover intentions: Interactive effects of value attainment, job satisfaction, and positive mood. Journal of Applied Psychology, 81(3), 318–325.

Guletkin, E. (2011). What's the Value of Your Employment Brand? Retrieved October 6, 2019, from LinkedIn Talent Blog website: https://business.linkedin.com/talent-solutions/blog/2011/12/whats-the-value-of-your-employment-brand.

Haslam, S. A., Eggins, R. a, & Reynolds, K. J. (2003). The ASPIRe model: Actualizing Social and Personal Identity Resources to enhance organizational outcomes. Journal of Occupational and Organizational Psychology, 76(1), 83–113.

Haslam, S. A., Postmes, T., & Ellemers, N. (2003). More than a Metaphor: Organizational Identity Makes Organizational Life Possible. British Journal of Management, 14(4), 357–369.

Hatch, J. M., & Schultz, M. (1997). Relations between organizational culture, identity and image. European Journal of Marketing, 31(6), 356–365.

He, S. L. & Wang, J. H. (2012). Development of the Chinese version of the Dentine Hypersensitivity Experience Questionnaire (DHEQ). European Journal of Oral Science, 22, 218–23.

Hogan, S. J., & Coote, L. V. (2014). Organizational culture, innovation, and performance: A test of Schein's model. Journal of Business Research, 67(8), 1609–1621.

Holland, J. L. (1959). A theory of vocational choice. Journal of Counseling Psychology, 6(1), 35-45.

Howard, M. C. (2016). A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? International Journal of Human-Computer Interaction, 32(1), 51-62.

Kennedy, K., & Hill, V. (2016). The New Talent Landscape: Recruiting Difficulty and Skills Shortages Media Contact China Gateway Plaza Chaoyang District United Arab Emirates. Retrieved from https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/SHRM New Talent Landscape Recruiting Difficulty Skills.pdf.

Kim, H. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Open lecture on statistics.

Kristof, A. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. Personnel Psychology, 49(1), 1–49.

Malinowski, J., Keim, T., Wendt, O., Weitzel, T. (2006). Matching People and Jobs: A Bilateral Recommendation Approach. Proceedings of the 39th Hawaii International Conference on System Sciences, 1-9.

Marais, I. (2012). Local Dependence. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), Rasch Models in Health. Wiley-ISTE.

Mc Clelland, D. (1985). Human Motivation. Glenview, Illinois: Scott Foresman & Co.

Ogbonna, E., & Harris, L. C. (2000). Leadership style, organizational culture and performance: empirical evidence from UK companies. The International Journal of Human Resource Management, 11(4), 766–788.

Parks-Leduc, L., Feldman, G., & Bardi, A. (2014). Personality traits and personal values: A meta-analysis. Personality and Social Psychology Review, 19(1), 1–27.

Pew Research Center. (2016). The State of American Jobs | Pew Research Center. Retrieved October 5, 2019, from Pew Research Center website: https://www.pewsocialtrends.org/2016/10/06/the-state-of-american-jobs/

Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job Engagement: Antecedents and Effects on Job Performance. Academy of Management Journal, 53(3), 617–635.

Rouquette, A., Hardouin, J. B., Vanhaesebrouck, A., Sébille, V., & Coste, J. (2019). Differential Item Functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. PLoS ONE 14(4), 1-16.

Ryan, R., & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology, 25(1), 54–67.

Sagiv, L. & Schwartz, S. H. (2022). Personal Values Across Cultures. Annual Review of Psychology, 73, 517-546.

Saks, A. M. (2006). Antecedents and consequences of employee engagement. Journal of Managerial Psychology, 21(7), 600–619.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. Annual Review of Psychology, 64, 361–388.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology, 25, 1-65.

Schwartz, S. H. (1994). Are There Universal Aspects in the Structure and Contents of Human Values? Journal of Social Issues, 50(4), 19–45.

Schwartz, S. H. (2006). Basic Human Values: An Overview. Jerusalem Hebrew University, 49–71.

Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. Online Readings in Psychology and Culture, 2(1).

Schwartz, S. H. & Rubel, T. (2005). Sex Differences in Value Priorities: Cross-Cultural and Multimethod Studies. Journal of Personality and Social Psychology, 89(6), 1010-1028.

Shen, L., Zeng, H. Zeng, H., Jin, X. Yang, J., Shang, S., & Zhang, Y. (2018). An Innovative Evaluation in Fundamental Nursing Curriculum for Novice Nursing Students: An Observational Research. Journal of Professional Nursing, 34(5), 412-416.

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.

Stricker, L. J. (1982). Identifying Test Items That Perform Differentially in Population Subgroups: A Partial Correlation Index. Applied Psychological Measurement, 6(3), 261-273.

Struch, N., Schwartz, S. H., & van der Kloot, W. A. (2002). Meanings of Basic Values for Women and Men: A Cross-Cultural Analysis. Personality and Social Psychology Bulletin, 28, 16-28.

Sull, D., Sull, C., & Zweig, B. (2022). Toxic Culture Is Driving the Great Resignation. MIT Sloan Management Review. Retrieved June 5, 2024 from website: from https://shop.sloanreview.mit.edu/store/toxic-culture-is-driving-the-great-resignation.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education, 48, 1273-1296.

Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53-55.

Van Bork, R., Grasman, R. P. P. P., Waldorp, L. J. (2018). Unidimensional factor models imply weaker partial correlations than zero-order correlations. Psychometrika, 83, 443-452.

Verplanken, B., & Holland, R. W. (2002). Motivated decision making: effects of activation and self-centrality of values on choices and behavior. Journal of Personality and Social Psychology, 82(3), 434–447.

Westerman, J. W., & Cyr, L. A. (2004). An Integrative Analysis of Person-Organization Fit Theories. International Journal of Selection and Assessment, 12(3), 252–261.

Woods, C. M. (2009). Testing for Differential Item Functioning with Measures of Partial Association. Applied Psychological Measurement, 33(7), 538-554.

Ziljmans. E. A. O., Tijmstra, J., van der Ark, L. A., & Sijtsma, K. (2019). Item-Score Reliability as a Selection Tool in Test Construction. Frontiers in Psychology, 9, 1-12.