Sofia Sjöberg

Cicek Svensson

Anders Sjöberg

Graphic design: Lina Strand

2019 revision: Johan Lindh

ISBN: 978-91-7418-364-1

Version 2.2

Item number: 010-301

Table of contents

CHAPTER 1

Introduction, theoretical background and the Five-Factor Model

Introduction

The driving force behind the development of MAP – Measuring and Assessing Individual Potential – was the need for a modern personality test with a scientific foundation, one that offers a good description of an individual's personality and with the capability to be used in the prediction of job performance with documented precision. Several methods that are available on the market are either lacking a scientific foundation, documentation of their alleged properties and efficiency, or the obvious modern occupational link. Moreover, most tests only generate descriptive information. This information may be relevant in certain contexts and for such use but contains no empirical information regarding an individual's suitability or plausible performance for a certain role or in a specific context. Therefore, combining the possibility of obtaining detailed descriptions about an individual's personality *and* empirically documented predictions of performance in specific roles from one tool, is something unique. The purpose of MAP, to provide standardized empirically documented predictions of performance to be used for decision-making, is in accordance with the international ISO 10667 standard for psychological assessments that came into effect in the fall of 2011.

The fact that human beings have different personalities and that these differences are significant for the way in which we act, is something that few people are doubting, and the scientific evidence supporting this has increased exponentially in the last decade. Personality drives human behavior, not least in the workplace. It sets the framework for an individual's strengths and weaknesses, potentials and challenges. This means that, regardless of the workplace and the tasks or the position that a person may have or apply for, personality is important for the way in which individuals view themselves and others, for how others perceive them and how they will function, thrive and perform.

The initial phase in a test development process of this kind should always be to define and specify the theoretical model that the instrument (measurement model) will be based upon. In the commercial test market, instruments are based on models and theories of varying scientific standards. Making a critical evaluation of the theoretical model that an instrument is based upon may be a complex and complicated matter, but today there is a broad consensus within the research community that the so-called Five-Factor Model (FFM) is the most robust and empirical model for measuring personality. This means that no matter what the purpose of identifying an individual's personality may be, e.g. selection, development or promoting self-knowledge, the empirically measurable structure is the same. The FFM is a taxonomy which postulates five broad personality dimensions. The dimensions are: Conscientiousness (abbreviated as CO), Emotional Stability (ES), Agreeableness (AG), Extraversion (EX) and Openness for experience (OP).

The FFM has a self-evident role in describing personality and is also the model with the strongest support among researchers in terms of its ability to predict behavior in the workplace. The FFM started being developed as early as in the 1930s with the aim of investigating how many personality dimensions would be necessary to describe an "average personality" in a comprehensive manner. The researchers Gordon Allport and Harold Odbert (1936) identified about 18 000 adjectives in the English vocabulary which are describing personality. To cope with this enormous amount of words, a reduction in the number of adjectives to a smaller number of related descriptions was suggested. By applying the, at that point, new approach of exploratory factor analysis, 4 500 adjectives were finally identified and summarized to approximately 30 factors – the so-called "traits". There were no particular theory underlying this approach for how personality should be constructed; the factors were based entirely on everyday personality descriptions that were grouped statistically into a number of clusters.

In the years that followed, research revolved around determining the number of factors and their composition, and it was not until the 1960s that the present-day FFM was formulated by Tupes and Christal (1961; 1992). This work was mainly based on extensive factor analysis of large amounts of data from the U.S. Air Force. However, in the late 1960s and the 1970s, personality research and the perception of individual differences was facing strong criticism (Mischel, 1968) leading to the FFM being somewhat forgotten. Research regarding individual differences, personality and the FFM was not taken up again until the 1980s, through longitudinal studies of personality development (Costa & McCrae, 1982). Today, the FFM is the dominating approach for measuring personality in the context of work psychology. The reason why the FFM has a special position in work psychology research is the stable empirical support showing that these factors, to varying degrees, are significantly contributing to the prediction of job performance and most other behaviors in the workplace.

The major breakthrough for personality research within work and organizational psychology came with the new methodological and statistical approach of meta-analysis. In meta-analysis, results from large number of studies are assembled and re-analyzed. This way, one obtains a more accurate estimate of the correlation between test scores and a criterion, or in this case, between measures of FFM and general job performance.

The meta-analysis by Schmidt, Shaffer & Oh published in 2008 provide meta-analytic estimates of the correlations between the five personality factors and the criteria job performance. The results show that the highest correlation (p) is between the personality factor Conscientiousness (CO) and job performance. Emotional Stability (ES) has the second strongest correlation with job performance, followed by Extraversion (EX) and Agreeableness (AG), while Openness (OP) display the weakest link to job performance. Note that although some correlations may seem low, they may still have a significant financial impact on corporations and organizations in for example personnel selection (Mabon, 2005). Barrick, Mount and Judge (2001) have also shown that the importance of Conscientiousness and Emotional Stability is generalizable across occupational groups, jobs, roles, and performance criteria, while other factors may only be relevant in terms of specific performance criteria.

In recent years, there has been ground-breaking research on the significance of personality in professional life and there has been great support for the possibility of describing personality in hierarchical terms, wherein the FFM represents one level. The proposed hierarchy illustrated in Figure 1 contains four levels of so-called latent constructs – abstract,

psychologically meaningful entities that can describe, explain and sometimes predict behaviors. The FFM-factors constitutes the third level as viewed from above, the level above consists of two overall factors, to which FFM-factors contribute to a varying degree and in different ways. These two so-called meta-traits are labeled Alpha and Beta respectively, and according to Digman (1997), they can be interpreted as representations of the socialization process itself, or as personal growth. The level represented in the figure as "GP" represents the General Personality factor. This factor represents an overall psychological construct, which, according to research, would be equivalent to the so-called g-factor – the general factor of mental ability – within intelligence research. Neither the GP factor or Alpha and Beta are operationalized within MAP.

The level below the FFM-factors is constituted by so-called facets. Together they constitute the various FFM-factors. In MAP, each facet is operationalized and measured in a subscale. The subscale is composed of several indicators. In psychological testing indicators manifest themselves as items traditionally designed as questions, statements or adjectives, which the respondent is asked to respond to. Although all four levels in the personality hierarchy are not operationalized in MAP, all of them have not only proved to be statistically identifiable but are also meaningful in describing and measuring personality.

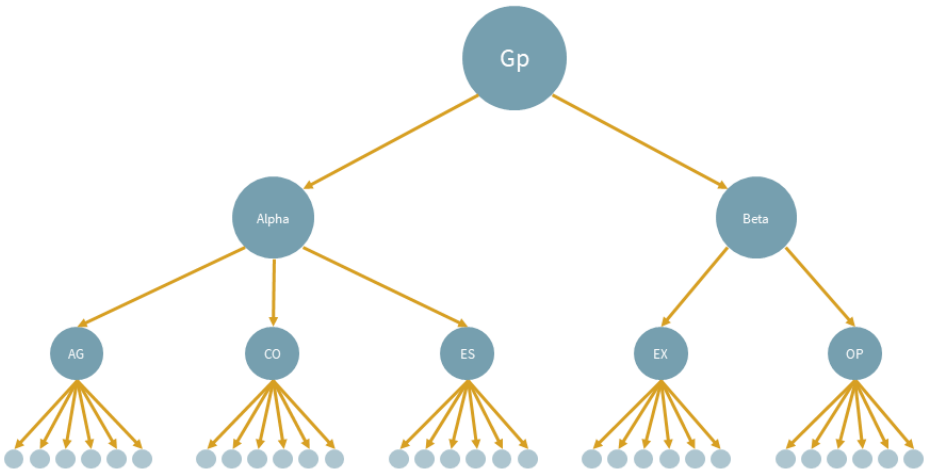

Figure 1. The hierarchical structure of personality

Description

Working with this type of hierarchical model has its advantages, both in terms of describing personality traits and in predicting behavior. Working on the facet level is a way of securing the construct validity, in the sense that the items used to measure a factor cover as much as possible of the domain of relevant thoughts, emotions and behaviors for the specific construct. A serious threat to the validity of personality measurement is that the psychological construct, which is supposed to be measured, is not fully covered – thus, is underrepresented. Working with a number of facets, increase the likelihood of covering the content of the overall factor, since each facet is a separate indicator of the individuals' characteristics, as being for example conscientious. Using facets also creates the opportunity to discover individual differences within a factor. Information on the facet level may thus offer a more detailed description of an individual or a group of individuals. Two people could have the same score

on one of the factors but opposite results on one or more facets. One individual might, for instance, have high scores on facet EX1 (the first facet/subscale of the factor Extraversion) and EX3, low scores on EX2 and EX4 and average scores on EX5, while another individual might have the opposite results – i.e. low scores on EX1 and EX3, high scores on EX2 and EX4, and an average score on EX5. This example will produce a similar result on the factor level, but by taking the facets into account, probable differences on the facet level could be highlighted. However, the starting point, both theoretically and empirically, is that the facets within one factor covary, so that a person with high scores on one facet will most likely (but not necessarily) also have a high score on the other facets within the same factor.

Considering the hierarchical model, the levels above the facets may seem redundant, since the content of the facets is more specific and detailed and may thus be perceived as being more useful, detailed and thus informative. Although this might be true regarding more detailed descriptions of the individuals' personality, the further one descends into the hierarchy, the lower will its reliability be – i.e. the certainty of the actual test scores. The higher levels in a hierarchy always provide a higher reliability and therefore indicate the test results more precisely. This psychometric fact is not always known to the end-users of psychological tests, but it is of utmost importance to clarify the reliability values of the components of a personality instrument, and to take these into account when designing the assessment process and interpreting test results.

Another important aspect is the possibility of generalizing, i.e. being able to draw conclusions about an individual's traits in a broader psychological sense. Since it is impractical to ask a person questions covering all possible behaviors that indicate a broad personality trait, one needs to choose certain core areas that the questions aim to cover. Hence, the questions function as indicators, from which conclusions are drawn about the personality trait as a whole. The higher one ascends in a hierarchy; the broader generalizations are possible to make. This is simply because the constructs higher up in the hierarchy are broader and thus entails broader psychological meaning.

A common reaction from test users is that tests, or rather the model that they are based on, is too narrow, and they oppose themselves to a reduction of personality to for example five personality factors. Traditionally developed and commercially available personality tests are often designed to provide broad descriptions of personality, and in order to escape the above mentioned "critique" on behalf of practitioners, they are often extensive and contain many scales, subscales and measures of personality traits that have not yet been identified in research. This makes them appear comprehensive and applicable to most contexts, and yet at the same time they risk leaving the impression of being able to generate large amounts of useful and reliable information about individuals in different contexts, which is simply not possible. The FFM model with its proposed hierarchical structure provides a clear framework for factors and facets with a meaningful and measurable psychological content that is relevant for behaviors in the workplace and in general. Hence, the FFM and the levels below and above are in no way claiming to capture or reflect all aspects of personality, merely the measurable parts. Note that the measurability in this case is not only referring to psychological testing as a measurement tool, but also to other methods used in psychological assessments, e.g., interviews, assessment center exercises.

CHAPTER 2

The initial development of MAP

Development and standardization

The development of MAP began with the definition of the theoretical model and thereby by deciding which individual differences, traits, to measure. These traits are based on so-called psychological constructs. The work with defining each construct was carried out thoroughly, both with respect to the constructs underlying the scales (factors) and those that form the foundation of the subscales (facets).

In the following description of the test development process there is a distinction made between what constitutes the theoretical model and the measurement model. The theoretical model refers to the proposed structure, in this case a hierarchical structure with two levels, on which the five factors constitute the upper level and the so-called facets make up the level below. The theoretical model also includes the definitions that determine and identify the psychological trait that makes up the construct, regardless of whether it takes place on a factor or facet level. The measurement model consists of the model that represents the scales and subscales that are used to measure the constructs according to the proposed structure. The terms scale and subscale are thus used when referring to the measurement model. A major part of the test development phase focuses on ensuring the congruence or correspondence between the theoretical model and the measurement model, as it forms the basis for a reliable and valid interpretation of the individual test scores.

The number of factors (five) and their general content is relatively uncontroversial from a research perspective, and there is plenty of research that links the FFM to different types of criteria, such as general job performance, leadership, health, and counterproductive work behaviors and so forth. However, the level below the five factors, thus the facet level, which entails the constructs that the subscales are based upon, is not as well researched. The research literature does not offer the same clarity into the number of partial constructs below each of the five factors, their content, or how they may be measured most efficiently. Because of this, working with the facets and subscales has been of a more exploratory character (investigative) than confirmatory (affirmative), and there has been an awareness that greater revisions might be required on this level.

The subscales are summed up and together they constitute the score on the scale that represents the overall factor. This ”bottom-up approach” is a way of ensuring the construct validity in each part of the instrument. The final construct definitions, both for the factors and facets, their meaning, the way they are measured by scales and subscales, and examples of behaviors that can be linked to high and low scores, are described in Chapter 3.

Based on the constructs, the work with operationalization (development of items, thus formulating statements, designed to measure the properties that represent each construct) was initiated. The selection of items was carried out based on both theory and empiricism.

Each item should be considered as an indicator of a particular aspect of the underlying construct, that is, each item should theoretically reflect the overall subscale. Several indicators are required for each subscale to ensure a high quality in the final measure (test score). However, according to classical test theory, it is not assumed that a certain construct is measured fully, but rather that by creating multiple indicators, an individual's true score might be conveyed more precisely. True scores only exist in theory and may never be attained in practice since all measurements to a certain degree are affected by measurement errors, with error defined as either how well an item measures the same construct or by the stability of the measurement over time.

To provide a broader measure of the scales, the original theoretical model for MAP contained six subscales for each of the five overall scales. The number of subscales, and partially the content of these corresponded well with the model that the personality test NEO PI-R (Costa & McCrae, 1985) was developed after (see Figure 2). With this theoretical structure as a starting point, the work with evaluating and revising the theoretical model and developing the measurement instrument MAP began.

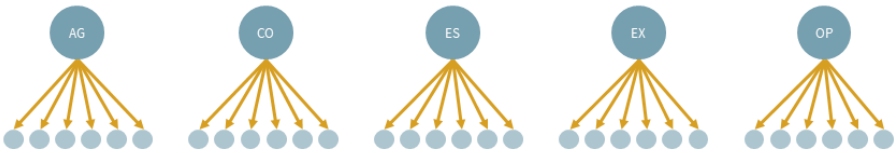

Figure 2. Theoretical model - Starting point

Based on the construct definitions of all subscales, in total 416 items with the response options Disagree, Somewhat disagree, Somewhat agree, Agree, were formulated. A team of experts: psychologists, psychometricians and test users with extensive experience in personality testing, participated in the project working with formulating and reviewing the theoretical and psychological content of each item in relation to each construct. Initially, the number of items ranged between 12 and 17 per subscale and the variation in number of items per subscale may be attributed to the fact that certain psychological characteristics, constructs, are harder to grasp than others and therefore require a tryout of more items. Some parts of the FFM constructs are also less explored and therefore less clear in terms of what they entail and what behaviors they are likely to drive. In the end, the items underwent a linguistic control focusing on the level of readability, linguistic formulations (e.g., sentence structure, grammar), and possible difficulties of being translated and adapted into other languages.

In the spring of 2009, a questionnaire consisting of all 416 items was sent to 5 000 randomly selected people (source: SPAR register) between 18 and 65 years of age and representative of the Swedish population in terms of gender and age. In total, 569 people responded to the questionnaire after one reminder.

Of these 569 respondents, 64% were women; the average age of the sample was 42 years (SD = 13). The educational level ranged from elementary school to postgraduate education. The vast majority, 212 persons (37%), had completed at least three years of high school and had some form of higher education. This group is described in relation to the population in Table

A1, Appendix A. Once the data had been collected from the average population, the analysis was initiated. The three overall objectives entailed:

1. Ensuring that the measurement model follows the theoretical model on the FFM level
2. Investigating the facet level by taking into account the number and distribution of subscales in the measurement model across the FFM factors
3. Given that the first and second objective have been achieved, creating reliable subscales (and thus indirect scales) with a maximum of 10 items per subscale

Several statistical and qualitative analyses of items were conducted to accomplish these objectives. Before the analyses process started, it was decided not to set uniform psychometric demands for all scales and subscales. This decision was based upon the quest of developing an instrument in which each part, scales and subscales, is as well developed as possible and of the best possible quality, considering its own preconditions. Setting up general psychometric requirements may impede this objective. One example is the acceptance of a lower alpha value, such as .65, in a subscale measuring a less homogeneous construct, and the rejection of the same value in a different subscale whose psychological content is more homogeneous and may thus be expected to demonstrate a higher measure of internal consistency. The fact that the scales and subscales are part of a complex structure with different levels of abstraction, that they overlap to various degrees and are of diverse psychological character, means that they are individually affected by different kinds of problems and challenges viewed from a technical measurement perspective. Taking these differences, by not applying general rules when evaluating and selecting items, into account is a way of increasing the overall quality of the final version of the instrument.

The overall model - Factors and scales

The first and perhaps most significant phase in a test development process is to ensure that the overall theoretical model, in this case the level of the five factors, is well grounded in the measurement model and that data (empiricism) supports the proposed model. It is important to reach an acceptable quality for the overall model since it is indicative of the remaining development process and sets the prerequisites for working with scales and subscales. Initiating the psychometric work with a separate part of the model may be a waste of time and result in an instrument relying on a psychological model that has lost its focus. Setting up general psychometric requirements in such an extensive model might, as mentioned earlier, be counterproductive, since the different constructs of the model may require different types of adjustments due to their unique nature. Ultimately, such a procedure might have significant and serious consequences for the meaning and interpretation of individual test scores. The five factors underlying the scales, the ways in which these are measured, and their mutual relations were thus the focus in the first phase of the test development process.

To test the overall model, a confirmatory factor analysis (CFA) was carried out in the program AMOS 17.0. A confirmatory factor analysis should always be applied when there is a theoretical model to be tested because it imposes more rigorous requirements on data. The alternative approach is to apply an exploratory factor analysis (EFA) postulating no theoretical model at all to compare against. As expected, the first analysis did not meet the requirements

for an acceptable adjustment and the work of revising the measuring model was initiated.

The overall model – Facets and subscales

As discussed above, the need for revision varied in extent and character between the five scales. Since weaknesses in the broader and overall model may be concealed behind reliability deficiencies in the subscales, the work started with reviewing each item in all the 30 subscales in relation to their "own" scale and the other four scales. The starting point of this work was the factor loadings of the subscales. The goal was to create subscales with as high loadings as possible on their "own" scale and to ensure that the correlation between the scales was not too high while at the same time maintaining acceptable reliability levels despite the exclusion of several items.

To obtain a better fit between the measurement model and the theoretical model, the subscales associated with CO (Conscientiousness), EX (Extraversion) and ES (Emotional Stability) required marginal adjustments. Items with the lowest factor loading in the respective subscale were excluded. The number of excluded items varied from subscale to subscale depending on the "balance" in the loadings. In most subscales, there was a clear clustering of items with a lower loading, which therefore were excluded. In the subscales however, this pattern was not as clear. The items with the lowest loadings were excluded until at least 10 items remained in the subscale. All the items that were eligible for exclusion on these empirical grounds, were examined regarding their psychological content and linked to the subscale's remaining items to ensure representativeness (construct validity) in the subscale. Without such a qualitative assessment, there would be a risk that the scale (or in this case the subscale) might suffer from a so-called "construct underrepresentation". This means that the items measure a too narrow part of the construct and thereby fail to provide indications of all the parts in the underlying psychological trait. These scales are often very homogeneous, with a good internal consistency (alpha), but thus have shortcomings in terms of construct validity.

The scales AG (Agreeableness) and OP (Openness) and their respective subscales were identified as in need of more extensive changes, and with empirical data as a starting point, the subscales were reviewed and reconstructed, taking the underlying constructs into account.

Scales and subscales

Openness (OP)

The review of OP resulted in a set of items, whose theoretical content was estimated to be unclear or too close to another construct, to be excluded. It is a common misconception that the psychological construct underlying OP and scales designed to measure this trait, are related to creativity, energy and momentum, and that it is automatically manifested as productive and constructive activity.

However, as described in Chapter 3, this is not an accurate understanding of this underlying personality trait. This misconception has not only led to unrealistic expectations regarding the

relevance of this trait in the context of work psychology but has also led to over-interpretations of what different results on this scale are likely to mean regarding specific individuals' behavior. The need for emotional and intellectual stimulation on the construct level is disconnected from physical activities, energy, productivity and benefit to others. The fact that this trait is impossible to measure free from elements of energy or activity, makes it particularly important not to over-interpret the results on the construct level.

To reduce the correlation with other scales and subscales aimed at measuring areas that reflect energy and efficient activity – EX and CO – items with overlapping content were excluded. Items from OP1 Imagination, which imply that the imaginative ability automatically leads to something concrete and constructive, were excluded, since the application or the result of having an active imagination is not a part of the construct definition of OP. Items from OP3 Emotions indicating the extent to which feelings are expressed towards the environment, were also excluded, since the subscale mainly includes a person's inner emotional world. The perceived feelings may or may not be reflected in the individual's behavior. Items whose content refers to level of activity or the amount of work were excluded from the scale OP4 Experiences. The construct Openness rather includes the will or the inclination to try out new activities in the search for inner sensations and experiences. The amount and intensity of a person's explicit level of activity as a means of seeking stimulation through people in their environment are aspects of the construct that constitute the basis for the scale measuring Extraversion.

Agreeableness (AG)

In the scale AG, adjustments were mainly made on the subscale level, rather than in individual items. The analysis indicated that the factor loadings in the two subscales Compliance and Modesty were not compatible with the remaining subscales in the AG scale and should therefore not belong to this scale.

Most scales designed to measure AG, that are available for commercial use and have been developed in the recent decades (thus constituting the basis for a major part of the research literature) are to a great extent measuring the degree of "kindness" and "friendliness" in an individual. This definition makes it easy to formulate items and an optimal homogeneity is often attained. Homogeneity, or internal consistency, is one of several reliability measures and it is often assessed in the form of Cronbach's alpha. This reliability measure is manifest and comparable by way of scales and methods, which sometimes leads to its importance being overestimated in the assessment of the general quality of different methods and scales. This measure of quality must always be put into context and only becomes meaningful when assessed with regard to other reliability measures, validity, theoretical models and interpretation of individual scores.

The degree of kindness is part of the underlying construct, but it includes much more than that. A narrow and single-track operationalization of AG causes deficiencies in the construct validity and limit the scale's predictive power. AG has proven itself problematic in terms of predicting behavior in the workplace in a stable and relevant manner, which leads to the conclusion that the degree of kindness is not always meaningful in terms of how a person will perform in the workplace.

A thorough review of the literature on FFM and its development showed that the theoretical starting point for the AG-scale is very broad and includes the image that others have of oneself in different relationships, as viewed from various perspectives. “Image” refers not only to one’s tendency towards kindness, and towards a warm and friendly appearance, but also to the ability to be frank, sincere and clear, when the situation demands it. Hence, the definition of AG in MAP is broader and deeper than what is common in traditional measurements and it can be said to emphasize integrity to a higher degree. It is also important to point out that the desirable levels in traits such as kindness are often picked up in selection methods that are of social character, e.g. interviews, while aspects such as frankness are much more difficult to measure with such methods.

The degree of Compliance and Modesty is reflected to some extent in Extraversion, above all in the subscales Self-assertion and Sociability, where different degrees of these traits offer higher predictive capacity in a work- related context than the subscales Compliance and Modesty. Based on this theoretical reasoning and the empirical basis, the subscales Compliance and Modesty were excluded from AG.

The changes made in the other scales are briefly described below. In addition to the afore mentioned quality aspects, revisions were made in ES, EX and CO for the purpose of creating a balanced model, i.e. the same number of items in each subscale and the same number of subscales in each scale. This balance is a precondition for ensuring that the instrument rests on a sound, psychometric foundation.

Emotional Stability (ES)

The subscale Depression in the scale ES has a clearly clinical content, as it measures the degree of sadness or depression and was considered irrelevant in a work-related context. To the extent that a permanent state of mind is relevant for the average personality in a non-clinical context, it is captured in the subscale EX Cheerfulness on the EX scale. There is also an ethical dimension to measuring aspects of an individual’s personality in the context of work psychology, which can be linked to clinical diagnoses and mental illness.

All measurements used in a psychological assessment should be justifiable by the context in which the instrument is used. In an instrument such as MAP, developed for an international market and for psychological assessments in the context of work psychology, it is neither ethically nor legally justifiable to measure such clinical traits. In the U.S., where, compared to Sweden, there is a more developed legislation for psychological assessment, it is against the law of discrimination.

Extraversion (EX)

At the beginning of the test development process, a hypothesis was established, stating that the subscale AG5 Affection does not belong to the scale EX, as specified by Costa & McCrae’s model. The scale EX is designed to primarily measure one’s need of the presence of others and the degree of energy that one directs towards the social environment. These behaviors can certainly characterize the social interaction, but this is a secondary consequence of the psychological momentum and subsequent behaviors created by the direction of an

individual's energy. The subscale AG5 Affection describes the character which usually characterizes the relationships between an individual and others. On a construct level, this is reflected in the AG scale. The empirical analysis confirmed the hypothesis and the subscale was moved from EX to AG.

Conscientiousness (CO)

The subscales Diligence and Orderliness in the scale CO were merged into one subscale, since these two subscales presented very high factor loadings on the overall CO scale and showed substantial correlations with one another. The underlying constructs are thus difficult to separate by way of measurement. Both subscales presented very high reliability, which also made it possible to measure the now broader construct (Diligence/Orderliness) with the same number of indicators that were previously required for each of the two separate subscales. The subsequent examination of the final model's adjustment indicated that the model was also favored by the merging of the two subscales.

Determining the final measurement model

At this phase in the test development process, 306 items unevenly distributed across the five scales and now 25 subscales remained.

The work of establishing the overall model was now considered to be completed and the remaining work focused on reducing the number of items to achieve a balanced model of 8 items per subscale, thus 40 items per scale and 200 items in total. The reliability levels of the scales (which are partly a function of the number of items - the more items, the higher the reliability) also indicated that 8 items per subscale would be sufficient for all subscales.

To achieve 8 items in each subscale, items from every subscale were analyzed separately using Item Response Theory, IRT. This so-called one-parameter model was applied to the item level (to each single item) using the program RUMM 2020. In this way, the items that differentiate between individuals on the same difficulty level or have great measurement errors (residuals) were identified. Items with these properties could thus be excluded after a qualitative review of the content. In those cases, where none of the above occurred, a qualitative review of the item content was undertaken and if the content of two items overlapped, the one with the highest residual (highest error and thus lowest reliability) was excluded.

The revised model of 200 items and five subscales for each of the five scales (see Figure 3) was then retested and showed a significantly better fit than the original model. The results of this analysis are presented in Chapter 5. This version with 200 items thus constitutes the standardized version of MAP.

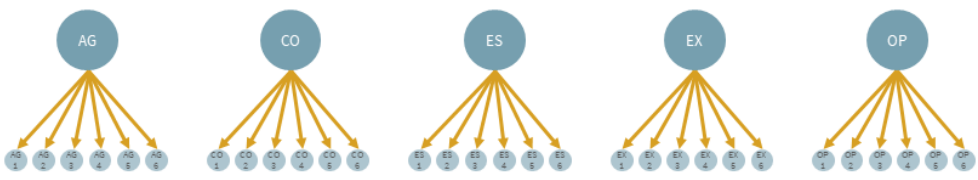

Figure 3. Theoretical model - Standardized version

CHAPTER 3

Construct definitions

Achieving the goal of developing a test for two purposes: *describing* measurable aspects of personality in a detailed and reliable way, and combining these measures in order to *predict* relevant work-related behaviors (MAP applied within the Targeted Prediction© framework), requires a solid theoretical work of defining what is to be measured, i.e. defining the underlying psychological constructs. In practical test development, it is not uncommon that this work is carried out somewhat routinely and sometimes the definitions are adapted to the easiness of formulating questions and thereby the measurement itself or the operationalization of the construct. The focus is thereby to achieve as high internal consistency (reliability) as possible, an aspect of test quality known to most test users. Reliability is something concrete, since its size may be expressed in a standalone and comparable value. As a test user however, it is important to make sure that the achieved reliability levels are not attained at the expense of the test score validity – a high accuracy in the test scores will not be interpretable unless the test score has (psychological) meaning, have impact on behavior and can be generalized. If there is too much focus on reliability, there is a risk that the construct validity will suffer, i.e. that the definitions become too narrow or that the formulated questions will not cover all the parts of the defined construct. This problem, known as ”construct underrepresentation”, inevitably results in a lack of external validity since few generalizations to other meaningful phenomenon can be made. It reduces the probability of meaningful connections to relevant criteria, such as job performance. However, it should be mentioned that in certain situations, the intention could be to measure specific parts of a construct. One such situation could be when the purpose of test scores is to predict, not to generate exhaustive descriptions, the measure (test score) does then function as an indicator of the current trait. Such reasoning should be described and stated explicitly and beforehand in the instrument's technical manual. See for example the development of Predicting Job Performance, PJP, (Sjöberg, Sjöberg & Forssén, 2006).

The basic preconditions for the balance between reliability and validity are manifested early on in a test development process. The instrument's application and its underlying theory is decisive in terms of how the construct definitions should be made and how they should be measured. The future areas of application for MAP (description *and* prediction) demand comprehensive and detailed construct descriptions, followed by detailed and relatively broad scales for measuring each construct. The definition and the subsequent description of the psychological content of each factor and facet are central to the instrument's final quality and areas of use. In-depth knowledge of the psychological construct and its definitions, which form the basis of scales and subscales, is thus a requirement to achieve a valid and reliable interpretation and use of test results.

The following section provides information about each constructs' position in the hierarchy of the FFM, the way it is commonly defined and measured, as well as a description of

probable characteristics and behaviors for individuals with low or high scores on the specific scale. Low and high scores respectively on the scales and subscales represent the opposites on the underlying dimension. This makes low and high scores more manifest and distinct on a theoretical level and more likely to represent a trait of character for an individual compared to when having an average score on a scale or subscale. Thus, high and low scores reflect personality traits that are likely to manifest themselves more clearly in terms of their behavioral expression, compared to average scores.

The following section also describes which part of the overall construct each subscale is intended to measure, and which behaviors that are likely to be associated with low and high scores respectively.

Agreeableness (AG)

History and common definitions

The construct underlying the scale Agreeableness in MAP concerns how an individual interacts with other people. Although the underlying psychological construct is very broad, most measures of this trait are operationalized narrowly. Hence, scales postulating to measure Agreeableness traditionally generate a score which only indicate the degree of kindness a person express towards others. This aspect of an individual's interpersonal style may be relevant, but the construct is broader and thereby also needs to be operationalized with a broader set of indicators. This is a prerequisite for the test scores being valid when generalized to a broader and more significant psychological content. One consequence of this narrow operationalization is that links to relevant work-related criteria are often limited although it has shown moderate links to more specific criteria, such as service and leadership. The reasonable conclusion of the lack of correlation with external criteria should rather be that the degree of kindness seemingly has a limited impact on an individual's performance at work.

Another explanation to the lack of correlation may be that the scaling is often distorted in the sense that the scale's measurement characteristics enable it to distinguish between individuals who are quite friendly and those who are extremely friendly (compliant and afraid of conflict) rather than between individuals who tend to be unfriendly and those who are friendly. The latter could be a more appropriate range to measure in this context and could thus possibly result in a higher correlation to external criteria and better underlying distributions. The aim with Agreeableness in MAP is to operationalize the entire construct, which includes more and wider areas than what other measures traditionally cover, and to develop scales with indicators that reflect all levels of the construct.

Theoretical definition

The Agreeableness scale offers an insight into which style an individual tends to apply in their interpersonal relationships, rather than the extent of, or the focus one attributes to the social environment. The latter is captured by the Extraversion scale. An individual's social style is characterized by the extent to which the person feels trust in human nature, assuming that humans are good in general. This fundamental trust affects the interaction with others through verbal communication and body language and creates the foundation

for the extent to which one radiates consideration, affection and warmth toward others. Being kind, warm and attentive are often positive traits, but the flipside of such a behavior is to show compliancy, fear of conflict, the inability to stand up for one's own opinions in relation to others, saying no and setting boundaries. Hence, these characteristics need to be balanced with the ability of being sincere, honest and direct in one's communication - even if one falls into disrepute, or is subjected to the anger, frustration or sadness of others. A sympathetic appearance, kindness and the capacity of being caring and compassionate should therefore be related to how and what one communicates. Most operationalizations of Agreeableness are missing several of these aspects, by focusing solely on kindness and friendliness. The operationalization of the construct underlying Agreeableness in MAP provide an indication of what could be described as "genuine" agreeableness, which may be viewed in contrast to being kind, accommodating, and pleasing to others (compliant) in order to avoid negative reactions from others.

Behaviors

People with high scores on the Agreeableness scale have a basic trust in other people. They are altruistic, caring, attentive and care about what others think and feel. They adjust their own behavior to take others' feelings into account, and they are easily affected and engaged in other people's problems and emotional states. They want and like to help and offer their support to others. They often focus on collaboration, consensus and would like to please everybody, which means that they usually are perceived as tolerant and humble. If their kindness and desire to please others is not combined with a straightforward communication and an ability to stand up for their own views, they could be perceived as naïve and weak turncoats who are lacking principles. These individuals are often very comfortable and pleasant to be around, and they usually make good progress in organizations since they are often adaptable and do not question the views of others or enforce their own. In a work context however, the relevance of this feature has proved itself to be questionable.

People with low scores on the Agreeableness scale are more reserved, cautious and skeptical against their environment and are bound to adopt a more critical approach. They don´t attach great importance to feelings, wishes or the views of others, and rarely feel the need to adapt their own behavior to please others. They are more focused on themselves and their own person. Their approach towards others is more often characterized by competition, rather than cooperation and support. They relate to authority and are respected by others for being skeptical, tough, safeguarding, competitive and sometimes aggressive. These individuals might be perceived as hostile, rude, self-centered, arrogant or combative but also thick-skinned, independent and objective.

In MAP, the Agreeableness scale consists of five subscales (this part in the subsequent paragraphs will also be illustrated by figures, in which the current part of the figure is highlighted)

AG1 Trust

AG2 Communication

AG3 Altruism

AG4 Compassion

AG5 Affection

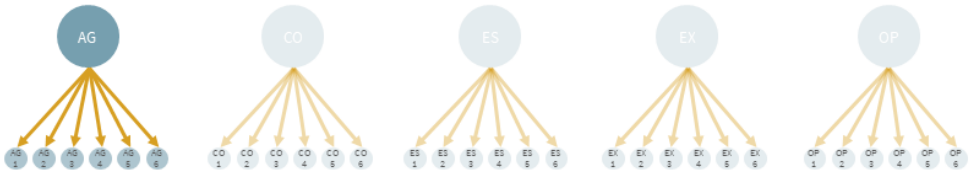

Figure 4. Theoretical model – Standardized version (AG)

AG1 Trust

The subscale Trust reflects one's credence in others and the easiness one has for trust in others. People with high scores on the Trust subscale have a general belief in the goodness of humanity and have confidence in human nature. They tend to believe that most people have honest and good intentions. Individuals with low scores tend to be more restrained and reserved towards others. They might be perceived as cynical, skeptical but also realistic and thick-skinned.

Example of an item in the subscale Trust: I know that most of those who do me a favor are honest and reliable

AG2 Communication

The Communication subscale indicates both the content of and the way in which a person tends to communicate. People with high scores in this subscale tend to be sincere, straight, honest and undisguised. They are frank, unaffected and sincere, they are not prone to suppress or shape information to their own benefit. People with low scores are often more likely to be compliant, to please others, shape their way of communicating or the content in their communication to get their way (thus, they can benefit from it) and to avoid negative reactions from others. People with low scores rarely have problems bending the truth and are often more careful in expressing their real emotions and opinions. People with low scores often view these strategies as necessary and desirable elements of social skills, since it makes them more socially smooth and adaptable in the immediate situation. People with low scores may think that outspoken people are naïve or excessively sincere.

Example of an item in the subscale Communication: You don't always need to make your intentions so very clear (revised score)

AG3 Altruism

The subscale Altruism measure how importantly one feels about others' well-being and how this influence oneself. High scorers are altruistic, attentive towards the needs of others and care a lot about other people. They are generous, caring, sacrificial and always willing to assist those in need. These individuals wish to be there for others and feel good when they

can support others and offer them relief. People with low scores do not have the same need of being there for others and are not as attentive and aware of others' needs. They are more self-centered and might be reluctant towards becoming involved in others' problems.
Example of an item in the subscale Altruism: I am there for other people, even if it's at my own expense

AG4 Compassion

The Compassion subscale measures the level of tenderness, sympathy and concern for others. High scorers on this scale are considered kind and compassionate and are easily affected by the problems and needs of others. They are caring and empathetic. People with a lower degree of sympathy may perceive them as squeamish and overemotional. People with low scores are more practically-minded and not as affected by pleading and pity, neither towards nor from others. These individuals could easily be perceived as tough and insensitive, while they rather see themselves as realists, who think rationally and are unaffected by emotions.
Example of an item in the subscale Compassion: Other people think I'm thick- skinned

AG5 Affection

The results on the Affection scale reflects the extent to which a person experience and display affection, love and tenderness towards and in front of others. These traits are relevant to the ability of forming friendly and intimate relationships. People with high scores on this scale are very affectionate, warm and loving people who value their relationships with others. They have a warm, friendly and personal approach towards others, while those with low scores are more moderate and reserved. People with low scores are often perceived as more formal and impersonal, since they maintain more emotional distance in their relationships.
Example of an item in the subscale Affection: I attach great value to nurturing my relationships

Conscientiousness (CO)

History and common definitions

The factor Conscientiousness constitutes a less problematic psychological construct compared to for example Openness. Conscientiousness is well-defined and may be measured in a sound and relevant way. The width and meaning of the construct have given this factor a special position in work-related contexts in general, and in the context of psychological testing in particular. This is due to more interactive forms of assessment methods, such as interviews, often having difficulties capturing this very important personality trait. In fact, such assessment methods have the tendency of downplaying the importance of Conscientiousness, due to people having a high score appearing too goal oriented, ambitious, and driven - characteristics not always valued in the socially interactive assessment setting. The difficulty of interpreting scores measuring this trait often springs from the fact that the indicators – the questions meant to measure this trait – often focus on the level of detail orientation and the aspect of orderliness. In these cases, the construct validity is suffering and generalizations from scores to the broader construct become

problematic, and in addition, correlations with external criteria may fail to occur. The fact that the criterion validity of this factor, despite a narrow operationalization (low construct validity), is still so highly correlated with the criteria, may depend upon the fact that its so-called nomological net is more complex than for instance that of Agreeableness. The theoretical parts included in the Agreeableness construct are more loosely connected, which may lead to a narrow operationalization affecting the criterion validity to a greater extent, compared to if the theoretical parts would overlap or if they were more closely connected.

Theoretical definition

Conscientiousness is the main personality trait of interest in all type of situations where performance is in focus; e.g., work, learning, education. This characteristic is constantly linked to criteria that relate to performance at work and mostly it proves itself to be the most significant personality factor. This personality trait represents the urge for achievement and contains the necessary traits to comply with his driving force. This includes the ability to be organized, systematic, dutiful, efficient, neat and persistent. Conscientious individuals have clear objectives; have a strong will, are determined and are easily able to motivate themselves. The positive aspects associated with high scores are academic and personal achievements. The negative aspects are that this may lead to an annoying degree of precision, an excessive need for control, perfectionism or workaholism. It should be mentioned however, that whether this type of conclusion may be drawn from the scale's test scores depends on how the scale has been operationalized. The Conscientiousness scale in MAP describes how one prefers to work and how one relates to obligations and commitments.

Behaviors

Individuals with high scores are conscientious, they have a strong will and are determined. They often possess a basic motivation for working hard, performing and achieving their set targets. They have high demands on themselves and their environment. 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, conscientious and loyal members of the organization, who think before making decisions. They put a lot of time and effort into preparing, organizing and scheduling, often to maintain a high and constant level of efficiency and to be able to maintain control. They may experience ambiguity and inefficiency as something troublesome and thrive in situations that are somewhat predictable. The accuracy and attention to detail can make others perceive them as controlling, overambitious or demanding, and their perseverance can lead others to the impression that they are stubborn in their way of working. At the same time, they are conformist in the sense that they adapt to the prevailing rules, customs, practices, norms and to the expectations of the group. Their own awareness of rules and high moral principles makes them perceive less rule-conscious people as unreliable and provocative.

Low scores indicate that a person tends to approach their, often unclear, goals and commitments in a more relaxed and easygoing way. They often conceive of planning and preparation as restrictive and sometimes inhibitory, and rather follow the spur of the moment, taking each day as it comes.

It is easy for low scorers to delay things, since they rarely value efficiency or hard work in itself. Even when they are involved in many different things at once, their driving force is often spontaneity and desire. This also characterizes their way of making decisions: quick and sometimes hasty, based on feelings and impulses, rather than on logic and thoughtfulness.

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. The latter may lead them to being perceived as irresponsible, rash and unproductive, but also as flexible, spontaneous and adaptable. Low scorers are not necessarily lacking in moral principles, but are less strict in applying them. They are insensitive to rules, expectations, norms and authorities, and may lack the ability to understand the effects of their own behavior on others. It is important to note that people with low scores are not necessarily lacking ambition, but are distinguishing themselves from high scorers in terms of their lack of focus and the tendency to work towards their goals in a more strenuous, disciplined and independent manner.

In MAP, the Conscientiousness scale consists of five subscales:

CO1 Intensity

CO2 Diligence

CO3 Ambition

CO4 Self-Discipline

CO5 Decision Making

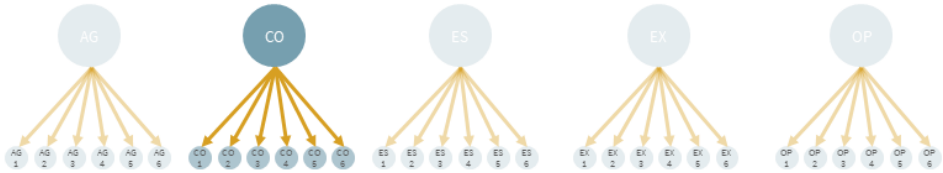

Figure 5. Theoretical model - Standardized version (CO)

CO1 Intensity

The subscale Intensity indicates the intensity and persistence with which an individual carries out work. A person with high scores often have the ability to work hard and values hard work and efficiency and also values this for its own sake. The sense of inefficiency may make them uncomfortable, restless and dissatisfied with their own performance.
High scorers are often persistent and do not give up easily. They often view themselves as suitable, capable and qualified, and they are often perceived by others as competent, wise and sensible. This perception of oneself is important and central to their self-image, events that disrupt this may therefore be experienced as troublesome. Individuals with low scores have a more relaxed approach to efficiency and hard work; these phenomena have no intrinsic value to them. Low scorers have no problem cutting down on the work pace and they might come across as a bit lazy and slow. To them, giving up and postponing is closer at

hand than to high scorers and they rarely have the same stamina. Since low scorers do not have the same attitude towards efficiency and hard work, they are rarely bothered by the fact that their self-image and the image of others may have of them often is that they are being unprepared, incapable and less suitable overall. Note that this scale does not reflect whether or not a person actually is capable or suitable to perform a certain task, only whether or not this person perceive him or herself or are perceived by others as suitable and capable.

Examples of an item in the subscale Intensity: Being efficient is very important to me

CO2 Diligence

The Diligence subscale consists of two equally important parts – on one hand the level of orderliness that a person expresses, on the other hand the attitude of a person towards their commitments. A person with high scores on this subscale is orderly, well-groomed, diligent, well-behaved and organized. They have a great sense of order, are careful and like to keep track of things. To maintain this, they work in a methodical and structured manner, taking one thing at a time. They might be stressed by disorder or if they need to deal with too many tasks at once. They might also have a hard time letting go of things, especially if they consider them as imperfect or incomplete.

Individuals with high scores stick to the rules and take their commitments and obligations very seriously. They are often conscientious, responsible and loyal people, who follow their ethical principles and moral beliefs. Low scorers are not as organized and may come across as unstructured, sloppy, careless and negligent, while they might also find it easier to relate to unclear situations and instructions. They do not have as great need for structure and do not spend as much time organizing and structuring. Low scorers don't take obligations and commitments as seriously and they rarely have trouble letting go of things, even if they are not perfect.

Example of an item in the subscale Diligence: I devote a lot of time to preparing myself for things

CO3 Ambition

This subscale reflects someone's desire to perform, how far one is willing to go and how hard one is willing to work to achieve their goals. Individuals with high scores on this scale are hard-working, energetic and put a lot of effort into achieving their goals. They are ambitious, tenacious, conscientious and often have a clear direction in life. They might be perceived as overly performance- and goal-oriented, and thereby lacking focus on what is taking place here and now. Low scorers do not have the same incentive to perform, they take each day as it comes and have a relaxed attitude towards goals, performance and achievements. They are governed by spontaneity – efficiency and hard work have no intrinsic value. They are often perceived as lustful, lazy and aimless, as well as flexible and prone to change. These people are usually very satisfied with their lives and their level of ambition.

Example of an item in the subscale Ambition: Once I have made up my mind, I work hard to get there

CO4 Self-Discipline

Self-Discipline is about initiating tasks or taking on commitments and carrying these

through, regardless of whether one is unwilling or finding it difficult to complete them, due to boredom or distractions. High scores on this scale indicate that the individuals have the capacity to motivate themselves and get the job done, even if it's a monotonous, routine assignment. These individuals often succeed in maintaining their focus despite distractions and disturbing elements. To them, the result – achieving the goal – is what is most important. They often feel their best when they can take on tasks and finalize them. People with high scores may have a difficulty postponing things. Low scorers may have a hard time motivating themselves to finish what they started on, especially if it is a monotonous or routine assignment that requires persistence. People with low scores are easily distracted and have difficulties staying focused on one thing at a time.

Example of an item in the subscale Self-Discipline: I make sure I complete boring jobs straight away

CO5 Decision Making

The Decision-Making subscale indicates how a person tends to make decisions, thus their strategy when collecting, aggregating information and ending up with a decision. High scorers are prudent and take their time to think and analyze before making decisions. They are careful, thorough, reserved and diligent in their search for relevant information. This process often results in sensible and informed decisions, and they usually appear wise and trustworthy because they can usually explain how and why they made a particular decision. Sometimes, their reflections make them appear suspicious, doubtful and irresolute, and they might find it difficult to make quick decisions. Low scores tend to make quick and hasty decisions, often without thinking through the consequences. Low scorers might also have difficulties gaining acceptance and respect for the decisions they make, regardless of their accuracy, since the logical rationale behind them is often missing or difficult to express.

Example of an item in the subscale Decision Making: Sometimes I'm a little too swift in reaching a decision

Emotional Stability (ES)

History and common definitions

In an academic context, the psychological construct underlying the Emotional stability scale in MAP is called Emotional Stability, or Neuroticism in case it has an inverse scoring system. Emotional Stability is known as the 'g-factor of personality' due to its unique position in the Five-Factor Model hierarchy. This scale is a key asset, as it often affects how the other four scales are likely to manifest themselves. It is also the construct most clearly connected to the clinical context, e.g. depression, anxiety disorder and psychopathology. However, it is important to stress that the scale's introduction in the field of work and organizational psychology, as well as the definition and measurement of this trait, are by no means claiming to capture the abnormal aspects of an individual's emotional stability. The work of measuring this trait has been going on for a long time. The first inventories of measuring neurotic tendencies were developed during World War I with the aim of assessing the soldiers' ability to handle stress. Being emotionally stable has proven to be an important feature to withstand stress and strains in most professions.

Theoretical definition

The construct Emotional Stability refers both to the stability and the character of an individual's overall emotional state. The construct includes both poles of the phenomenon – that is, on one hand being emotionally stable, well-adjusted and balanced, as opposed to anxiety-driven, unpredictable and unsure of oneself. In particular, a low level of Emotional Stability reflects the tendency to experience negative emotions such as fear, depression, embarrassment, anger, guilt and disgust. Characteristics, such as the ability to adapt, how well one is able to resist impulses, the extent to which they perceive and handle stress and an individual's general mood, are of central importance to the construct. Note that the latter deals with the stability of moods (not being moody), rather than a type of mood (positive/negative).

Behaviors

Individuals with low scores on this factor tend to be anxious, worried and easily get depressed or discouraged. They are vulnerable to external stress and their uneven disposition makes them come across as moody and irritable in pressured situations. Difficulties in setting priorities, distinguishing what is important and taking one thing at a time is a probable behavior for individuals with low emotional stability. The tendency to experience negative emotions interferes with the ability to adapt, and irrational thinking patterns are sometimes a consequence of this. One example is that problems or awkward situations are often taken personally, despite this not being perceived as such by the others. The negative emotional state comes to expression even more often in individuals with low scores, since they are not as much in control of their impulses. In pressured situations, people with extremely low scores might be eager, tense, restless, temperamental or easily discouraged.

Individuals with high scores will mostly act calmly in stressful situations and are able to make rational and safe decisions, even under strained circumstances. They rarely feel defeated by adversity; they are confident, independent, believe in their own ability and are able to take care of themselves. One of the typical features of emotionally stable individuals is their consistent mood; they are rarely angry or provoked. Moreover, they are rarely troubled by feelings of guilt or regret over things they did in the past and are able to resist impulses and temptations, meaning that they rarely end up in unexpected situations. Individuals with extremely high scores might also be perceived as detached, insensitive or unaware of the seriousness of a situation.

In MAP, the Emotional stability scale is determined by five subscales:

ES1 Emotions

ES2 Temper

ES3 Confidence

ES4 Self-Control

ES5 Stress

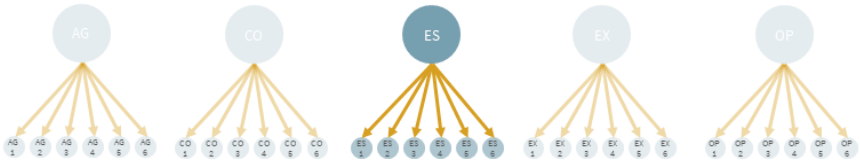

Figure 6. Theoretical model - Standardized version (ES)

ES1 Emotions

The subscale Emotions may be regarded as the core of the overall scale Emotional Stability. It contains statements aiming at measuring the level of anxiety or the emotional stability of a person, and whether or not a person is usually feeling calm and relaxed or is rather nervous and frequently anxious. The degree of confidence an individual has for the future, the ability to deal with uncertainty, not dwelling on or regretting things that one has done, represent high scores on this subscale. Anxiety prone individuals – low scorers – are per definition emotionally unstable and are often anxious and worried, afraid of trying out new approaches and can easily get anxious because of trivial matters and agonize over minor decisions. They are nervous, tense and worry about the future. It is important to bear in mind that anxiety also works as a driving force: being (moderately) watchful, on one's guard and worrying about making mistakes, forgetting or omitting something, may lead to a thoroughly performed work and give the impression that a person is taking things seriously.
Example of an item in the subscale Emotions: I can find it hard to stop dwelling on incidents in the past

ES2 Temper

The Temper subscale measures a person's tendency to be irritable. This tendency affects an individual's behavior, especially from an interpersonal perspective. People with a low score tend to feel and express anger, frustration, disappointment, bitterness and other states related to general dissatisfaction more often than others and is likely to be perceived as edgy and moody. High scorers have a higher tolerance level and do not lose their temper as easily as low scorers. They take things much more lightly and it takes a lot to get them to feel anger and irritation. Note that a high score on this subscale does not automatically imply that a person has a positive or optimistic appearance (this is reflected in the EX5 subscale Cheerfulness), only that their proneness to irritability is low.
Example of an item in the subscale Temper: I sometimes feel irritated for no particular reason

ES3 Confidence

The subscale Confidence reflects how confident individuals are of themselves and the confidence they have in their own ability, aptitude and capability. Individuals with low scores have more doubts about their own abilities and may feel inferior to others. They often worry about what others may think which makes them feel insecure, they easily get embarrassed

and they are often perceived as shy. They might have difficulties making decisions and to feel comfortable with the decisions they have already made. They often feel insecure and might give a willy-nilly and uncertain impression of themselves. High scorers are more balanced, worry less and often give a more confident impression in social contexts. They are self-confident, have great faith in their own abilities and seldom worry about other people's opinions. They can easily make decisions, are easy to get along with, go for what they want and show what they can achieve.

Example of an item in the subscale Confidence: I see myself as a very capable person

ES4 Self-Control

The Self-Control subscale measures the extent to which a person is driven by impulsivity. Impulsivity is often associated with spontaneity, risk-taking and quick decision-making. In MAP, the subscale Self-Control is designed to indicate what resources one has to resist temptation, sudden impulses or desires, that might be directed towards food, cigarettes, alcohol or possessions. The ability one has to control these internal drives, which to a certain extent are present in everyone, will probably also affect the way in which other personality traits are expressed. For individuals with low scores, desires or temptations might be perceived so forcefully that a person's self-control is not strong enough to resist them, although the person might be aware of the fact that the consequences of this behavior in the long run are not always beneficial to them or to others. Acting out one's impulses provides instant gratification. Individuals with high scores can often resist temptations more easily and have enough self-control, so as not to act on their impulses. These individuals often have enough self-control and are able to suppress their impulses.

Example of an item in the subscale Self-Control: I give way to my impulses pretty easily

ES5 Stress

The Stress subscale indicates an individual's stress and strain tolerance. Low scorers are more sensitive to stress and easily get overexcited and anxious by various forms of strain. They may react to stress and strain with feelings of panic and hopelessness and might feel incapable or unable to deal with difficult or stressful situations. Since they seek help from others in those circumstances, they might find it difficult to build confidence in their own ability to handle stressful situations on their own. However, high scorers have a greater resistance to stress and strain. They often appear calm and focused even in stressful situations, and others often perceive them as relaxed and laid-back. They tend to maintain a sense of control even in difficult situations. They maintain belief in their own ability, even when there is great strain in their environment.

Example of an item in the subscale Stress: I always appear calm and focused under stress.

Extraversion (EX)

History and common definitions

The underlying construct Extraversion is found in most personality theories and is measured in nearly all available personality inventories. The actual construct, which in fact is quite broad, featuring both the need for social interaction and the level of energy, is often regarded as a not-too-difficult construct to operationalize. However, many

operationalizations of Extraversion have come to include elements of character in social activities, that is, whether a person is nice, friendly and caring. But according to the Five-Factor Model, these psychological aspects belong to the construct Agreeableness, which underlies the scale with the corresponding name in MAP. Extraversion is a construct that has proved to be important in some roles, occupations or behaviors in the workplace. What these have in common is their strong social component (such as leadership and service) and that they require energy and momentum (e.g. sales).

Theoretical definition

The construct forming the basis for the Extraversion scale in MAP is mainly characterized by the degree of sociability and energy directed towards the external world. Sociability includes both coping with and being interested in social interaction, as well as the need for continuous and extensive social contact with others. Both the absolute level of energy and the extent to which it is directed towards the external world are considered in this construct.

Behaviors

Individuals high in Extraversion often have a need to and enjoy being surrounded by other people. They thrive in situations which, in similarity to themselves, are characterized by a high pace, and they enjoy being the center of others' attention and having the role of leaders in different groups. Some of their characteristics are; being talkative, enthusiastic, lively, optimistic, social, light-hearted and happy. Sometimes they might also be perceived as outspoken, troublesome, aggressive, bold, arrogant or superficial. Because of their tendency to dominate a conversation, they might also be perceived as poor listeners. Extroverted people are often self-confident, give a confident impression and take up a lot of space in the social sphere. Outgoing people often enjoy excitement and seek activities and new environments that satisfy this need. Depending on the outcome of their activity, risk-taking and the search for excitement is sometimes interpreted as bold and innovative, but is just as often perceived by others as foolhardy and irresponsible.

Low scores on the Extraversion scale imply a more introverted attitude towards other people and the environment, and a greater interest in one's own psyche. Focus is more directed toward one's own ideas and thoughts; stimuli from the external world, such as another person's presence or impressions, is not required. Because of this, they sometimes come across as independent and detached from what is happening in the external world. Introverted people are often perceived as less social and reserved, since they often prefer and/or need seclusion or solitude. Social engagement tires them out, while quiet environments give them strength and energy. An introvert person often prefers to work alone or in smaller intimate groups, and often looks for projects that require collaboration with only a few other people and on a temporary basis. Small talk and superficial social interaction rarely interest them which leads to them being perceived as serious, quiet, reserved and somewhat withdrawn. Introverts usually keep a low profile socially and rarely feel the need to be the focus of others´ attention. It is important to note that these individuals are not unhappy, sad or pessimistic by default, they only have a less exuberant expression in relation to their environment.

In MAP, the Extraversion scale consists of five subscales:

EX1 Social Need

EX2 Social Image

EX3 Pace of Life

EX4 Excitement-Seeking

EX5 Cheerfulness

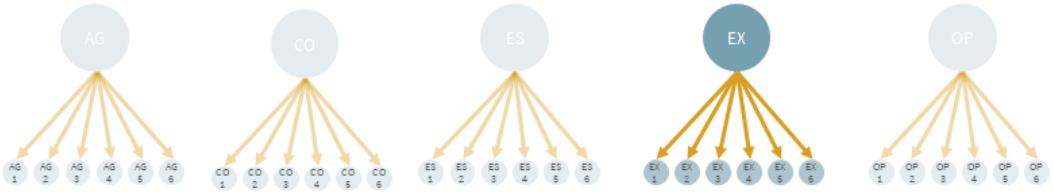

Figure 7. Theoretical model - Standardized version (EX)

EX1 Social Need

The subscale Social Need deals with how sociable a person is, what desire and need a person has of being in the company of others, usually more than a single person. Sociable people are stimulated and energized by spending time with others and are actively searching for social situations or to initiate group activities. People with low scores are more reclusive and more comfortable being alone or with fewer people. They prefer and require solitude and calm.

Example of an item in the subscale Social Need: I like going to parties where there are lots of people

EX2 Social Image

The Social Image subscale reflects a person's need for self-assertion. High scorers have a greater need for self-assertion and tend to come across as dominant and powerful enjoying having a prominent role or being the focus of others' attention. Individuals with a higher need for self-assertion often have an easier time than others gaining a social advantage and take a lot of space socially. High scorers can easily express themselves and enjoy taking a leading position in different groups. Low scorers on this scale have a smaller need for self-assertion and often appear dimmed, being reclusive or withdrawn. They rarely have the need of being at the center of others´ attention, rarely dominate the social space, and rarely enjoy activities that attract the attention of others. Low scorers are more comfortable of being in the background, creating and producing rather than being seen and being explicitly acknowledged. Sometimes people with low scores may be perceived as lacking in opinion or being indifferent; however, this is usually because of their limited need of expressing their opinion to others.

Example of an item in the subscale Social Image: I am a person who keeps a low profile in a group

EX3 Pace of Life

The subscale Pace of Life includes aspects such as power and energy. High scorers are often energetic, powerful, and keep a high pace in all their endeavors. They may become impatient and restless if the pace is too low. These individuals often leave a vibrant impression and exude vitality and life force. People with low scores take it more easily and have a more relaxed pace. They are not necessarily slow, tired or lazy, but simply less energetic in their expression and do not have an equally high demand of an active and fast pace of life.

Example of an item in the subscale Pace of Life: I like to have many things going on at the same time

EX4 Excitement-Seeking

The subscale Excitement-Seeking concerns the need for external stimulation, excitement, drama and whether one enjoys bright colors and noisy environments. High scores on this scale indicate that the individual is in need of excitement in their life, that they expose themselves to risks, exercise risky activities in their leisure time and would be willing to work in occupations with a high-risk factor, such as bodyguards or in the military. These individuals easily get restless and are not as stimulated by working with routine assignments, but are always on the lookout for new challenges and to experience new environments. They come across as fun, happy, interesting and captivating. However, there is a risk that these individuals are perceived as unserious and somewhat irresponsible. People with low scores do not feel that they need as much excitement and prefer a life that is more secure and calm, preferably in a familiar environment, something that would seem boring to a high scorer. Individuals who are more cautious might be perceived as serious, reliable, intelligent and sometimes a bit indecisive and overcautious.

Example of an item in the subscale Excitement-Seeking: I sometimes do things just because it's exciting

EX5 Cheerfulness

The subscale Cheerfulness measures the tendency to experience positive emotions, e.g., joy, happiness, love, satisfaction and to experience the feelings of cheerfulness and being content. Individuals with high scores on this scale are generally happy and positive and perceived as easy to talk to. They are usually perceived as cheerful, pleasant, sociable and fun. Individuals with low scores are less exuberant and less lively. They may come across as a bit dull and downhearted, but are just not as cheerful and high-spirited, as high scorers.

Example of an item in the subscale Cheerfulness: It's easy to make me laugh

Openness (OP)

History and common definitions

The dimension Openness is the least obvious factor in the FFM hierarchy, and consequently; maybe also the most misunderstood. The debate and research on the content of the construct has been going on for as long as the FFM has been applied as a framework for

measuring personality. Because of this discord, the definitions as well as the measurement have varied across different instruments and today the definition of the construct is still unclear. It might seem incomprehensible to many, especially to practitioners, that this trait often lacks relevance to the performance in the workplace. In research literature, Openness rarely presents any significant links to criteria related to job performance and can generally be said to be the least important trait in terms of predicting an individual's level of job performance. If one starts from the very common notion that this trait measures an individual's creativity, innovation and openness to new constructive ideas this is likely to seem incomprehensible. And in addition, since the meaning of this factor is often interpreted as the opposite of being reserved, reticent and unwilling to embrace the new, it would seem more likely that a high level of Openness would benefit an individual's performance at work.

It is true that this trait is dealing with the features mentioned above, but only in part. The underlying construct of Openness, originally labeled Openness to Experience, was soon reduced to merely Openness in most measurement contexts. Since then, this name has been interpreted and partially filled with a different psychological content, often linked to various forms of performance. But in its original definition, there is not part of the construct which is automatically linked to performance in the workplace.

Theoretical definition

The definition of the construct Openness in MAP is based on the original basic construct, in which Openness is an intrapsychic factor. This means that it deals with processes that occur and operate within the individual, and for the sake of the individual and for his or her need of inner emotional or intellectual stimulation. Meeting this need requires an openness that allows stimuli to flow in and out. The flow that is being referred to does not necessarily take place between the individual and the environment, but in and out of the individual's inner emotional world. Stimulus can be sought from the environment, but might just as well be sought for or created within the individual, without any intervention from the environment. This means that imagination, creativity and innovation may be a result of openness and might be accessible and useful to the environment and other people, it may however also only be accessible and useful to the individual him- or herself. Openness is thus primarily aimed at responsiveness to inner emotional experiences and openness to these new experiences. Hence, this intrapsychic trait does not automatically lead to something productive, constructive or something that will be put to use in an external context such as at the workplace.

Behaviors

The scale Openness in MAP involves an active imagination, aesthetic sensitivity, attentiveness to inner emotional life, love for variety, intellectual curiosity and independence from other people's judgements. High scorers on Openness are curious of their own internal world and of the external world, and their lives are rich in terms of inner experiences. These individuals often have the ability to come up with new ideas and have unconventional and independent values. Their emotional world is often more intense and nuanced than people with low scores on this scale. They are often more attentive to the inner world of emotions, and make more room for it in their lives, for instance in decision- making, when allowing themselves to be guided by their intuition. They have a wide-ranging imagination, and in

many cases, they have an artistic orientation and an intellectual curiosity. They often have a strong desire to try out what's new, they are open and constantly on the lookout for new experiences. These individuals get bored easily and seek or create situations that satisfy the need for emotional and intellectual stimulation. They often see themselves as original and artistic, while others may perceive them as eccentric and complex people always looking for new experiences to reflect upon. Generally, they prefer the complex and often dislike traditional approaches and conservative values.

People with low scores on the Openness scale are practical and down-to-earth, focusing on what is happening here and now. They neither have a great need of, nor do they search for, intellectual or emotional experiences for the pure sake of the experience. They offer their own emotional world limited attention and space, both for their own sake and in relation to the external world. Their emotional reactions can, in contrast to those of high scorers, be perceived as somewhat subdued or blunt, but also as straightforward and simple. These people make logical analyses and objectively weigh pros and cons against each other. Intuition and emotional experiences are given a limited space. People with low scores usually prefer what's already known and tested, in front of the new, unknown and uncertain. They might therefore be perceived as less flexible and reflecting. These individuals are more comfortable when engaging in repetitive activities and often feel satisfied when they know what to do and how to do it. Individuals with low scores on Openness tend to have a more limited number of areas that interest them. This often results in the possession of expert knowledge on one or two specific areas or topics. Individuals with low scores can easily adapt to the prevailing values and to people in authoritarian positions.

In MAP, the Openness scale consists of five subscales:

OP1 Imagination

OP2 Aesthetics

OP3 Emotional Sensitivity

OP4 Experiences

OP5 Mindset

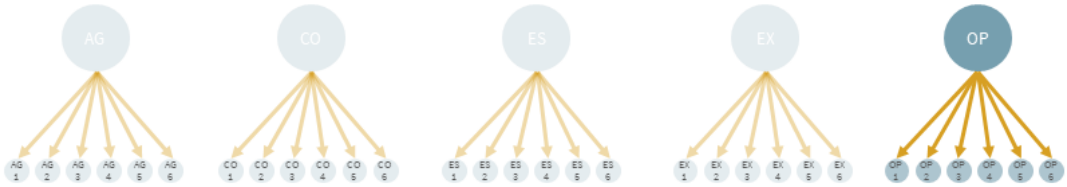

Figure 8. Theoretical model - Standardized version (OP)

OP1 Imagination

The subscale Imagination measures whether or not a person has a vivid and active imagination. High scorers have a dynamic imagination and create an interesting inner world

by daydreaming. Through these dreams they develop their fantasies, something they feel is contributing to a rich and creative life. Individuals with low scores are more down-to-earth and realistic. They prefer to focus on the task at hand and on what is happening here and now.

Examples of an item in the subscale Imagination: I am often told that I'm creative

OP2 Aesthetics

The subscale Aesthetics reflects the interest in aesthetic and artistic expression, such as poetry, literature, music, etc. High scorers on Aesthetics subscale often have a deep appreciation for different forms of aesthetic expression and are easily absorbed and captivated by how different phenomena are expressed in color and form. The key to these people is not being artistic, talented or even tasteful. The interest in aesthetics is driven by the trait of being affected by such expressions. People with low scores on the Aesthetics subscale are not as easily affected and are thus not as interested in aesthetic expression.

Example of an item in the subscale Aesthetics: Poetry and music often move me profoundly

OP3 Emotional Sensitivity

The Emotional Sensitivity subscale includes how receptive a person is to one's own emotional state and to the emotional atmosphere in the environment. High scorers tend to experience deeper and more specific moods, and often perceive their emotional life as an important part of their life. Individuals with high scores are more sensitive and they tend to experience both joy and sadness more intensely than those with low scores. They find it easy to express feelings and do not mind exposing their sensitive side to others. Individuals with low scores are more toned down. They do not act out their feelings to the same extent, and do not give their feelings as much space or ascribe them as much importance. Low scorers do not show their emotions to the same extent or with the same intensity and might perceive people with high scores as slightly exaggerated.

Example of an item in the subscale Emotional Sensitivity: I rarely let my heart rule my head

OP4 Experiences

This subscale measures the need of a person to make new experiences. Experiences are a means of reaching emotional or intellectual stimulation, and can be sought both from within the individual or in different types of external activities. The novelty aspect is a key element, as it generates new and more intense emotional or intellectual stimulation. High scorers prefer variety, they are keen on trying out new activities, visit new places and make changes for the sake of change. They are curious and frequently engage in a variety of hobbies. Individuals with low scores prefer the familiar and routines, and are happy to visit places that they already know of or have already visited. Change needs to be well motivated since it is often exhausting to low scorers and they prefer not to expose themselves to the strain. They choose what is familiar and already tested.

Examples of an item in the subscale Experiences: I try to learn something new as soon as I get a chance

OP5 Mindset

The Mindset subscale reflects a person's interest in and need for different forms of

intellectual stimulation. High scorers may be described as intellectually curious, with a great need for and interest in activities stimulating their own intellect. These individuals are actively seeking intellectual interests and have a sensitivity and willingness towards new ways of thinking and unconventional ideas. They appreciate engaging in philosophical discussions, abstract reasoning and other kinds of brain teasers, such as crossword puzzles and sudoku. People with low scores usually have a more limited interest in intellectual activities and are not as active in seeking out contexts which individually stimulate this specific need. They are not necessarily disinterested in intellectual reasoning, but they are less likely to engage in such activities for the pure sake of it. Note that scores on this scale do not indicate an individual's level of intelligence, the scores merely represent an intellectual interest or approach.

Examples of an item in the subscale Mindset: I like to ponder tricky questions

CHAPTER 4

Instructions for use and interpretation – decision making and feedback

This chapter provides guidance and advice to those who administer, interpret and provide feedback on test results generated from MAP, thus to the test administrator. In the following chapter, 'MAP' is referring to both full MAP and Essence if not anything else is explicitly stated or when the discussion concerns the subscales which are not included in Essence. The purpose of the following instructions, recommendations and guidelines is to create optimal conditions for the test administrator to provide the *respondent* with the opportunity to complete MAP in a standardized way, and thereby ensuring fair and comparable results. Another aim is to create optimal conditions for the *test administrator* to use the results of MAP in a standardized and professional manner, for the intended purposes, and in the appropriate contexts. Overall, this aim to provide the best possible conditions for a fair, professional and accurate personality assessment.

Areas of use

MAP has been developed for the purpose of measuring personality according to the five-factor model of personality, excluding the lower facet level. This foundation makes the areas of application for MAP extensive. Although the delivery of MAP, the way test scores are presented and how the standardized reports are structured, imply and somewhat guide the areas of use, the test scores themselves generated by MAP may be useful in all areas where human personality is relevant. From research, we know that the range of areas where personality is relevant is extremely wide. Thus, MAP is developed to measure personality according to the FFM of personality and aim to provide two types of interpretations: descriptions of individuals based on their personality, and to be used as a component in mechanical interpretation for the purpose of predicting future performance at work (applied within the Targeted Prediction© framework).

The descriptive information which MAP provides is primarily developed to be used for feedback (e.g., in a selection context) and development, career guidance, teambuilding and coaching. MAP applied within the Targeted Prediction© framework is to be used in personnel selection.

Administration and scoring

MAP is available via the cloud-based platform Ascend by Assessio and via partner systems using the Ascend API[1], MAP is not available in a paper-and-pencil format. In this technical

[1] The API is an abbreviation of Application Programming Interface and helps companies to share data in a controlled manner.

manual only the delivery of MAP via Ascend by Assessio is described. A test administrator may choose to start an administration on site or to send an invitation to the respondent via email and thus administer MAP remotely. Administering MAP remotely makes it possible for a respondent to take the test at home or at a different location, at the same time it entails that the mode of administration is unsupervised which always means lack of control for the test administrator. This lack of control affects the test administrators' possibility to ensure a standardized test administration. Therefore, to ensure reliable and valid test scores, a supervised mode of administration is always recommended.

Regardless of administration mode, the actual test session as delivered by the Ascend UI will be identical for a respondent. The test session starts with the respondent being presented with instructions on screen and asked to provide responses according to the requirements. After completion, Ascend will score, convert raw scores to standard scores, and generate feedback reports. The feedback reports are several and intended for different stakeholders (test administrator, respondent), contexts (please see the section Standardized reports in this chapter for more information regarding the structure, content and areas of use for each feedback report), and are delivered in different formats (on screen, pdf). The test administrator may choose what feedback report to retrieve, how to use it, whom to provide it for, in what format (e.g., orally – face-to-face or over the phone and/or in the standardized text-based format) and if it is to be sent to a respondent automatically or not. Note that any reports that are not listed below, are not an official product delivered by Ascend by Assessio and no responsibility will be taken for such content on behalf of Assessio Psychometrics AB. In addition, the set-up of the administration, interpretation, use of test scores and the possible feedback, is the sole responsibility of the test administrator.

Before the test session

Requirements and conditions of testing

There are several aspects requiring the test administrators' attention and awareness before administrating MAP to a respondent. Conditions of testing and requirements for administration are therefore listed in the following along with the information which should be provided to the respondent.

Expenditure of time

MAP is not a performance-based assessment and it does not require any preparation on behalf of the respondent. There are however conditions which needs to be met on behalf of the respondent for the assessment to be applicable.

The administration time, i.e. the time that a respondent has at their disposal to answer all statements, is not limited. The instructions recommend an even and steady pace when answering the statements. It is also recommended that all items are responded to on the same occasion (in a coherent session), implying that the testing session should not take place on several, separate occasions. It is, however, technically possible for the respondent to resume an interrupted or postponed session at a later point in time. Responding to the 200 items in full MAP takes on average approximately 20 minutes and the instructions states

that the respondent should set aside 40 minutes to complete the assessment. Responding to the 75 items in Essence takes on average approximately 10 minutes and the instructions states that the respondent should set aside 15 minutes to complete the assessment.

As mentioned, when the test administration is completed, Ascend will automatically score, calculate and transform test scores from raw scores to standardized test scores and may also generate the different feedback reports. Depending on the purpose of testing, the time for both interpretation and feedback may vary.

Environment

A non-distracting testing environment is needed. Public environments, e.g., internet cafés, and public transportation, are not suitable for taking MAP.

A personal computer is recommended since MAP has been visually adapted and developed for administration on a full-sized computer screen. Test-taking via tablet, smartphone or similar device is possible but may affect the test results.

A stable internet connection is needed for the full duration of the testing to ensure a valid result.

Overall and regardless of the mode of administration, the test administrator is responsible for creating an accurate and friendly atmosphere; the respondent should feel safe and comfortable in the situation and given the opportunity to provide a correct picture of themselves.

The test administrator should be well acquainted with MAP – both theoretically, psychometrically and practically - and be able to convey a calm, competent and secure environment for the respondent. It is important to consider that a test situation may well be an entirely new experience for a respondent, who will, at times, be in a vulnerable position, being aware of the fact that future opportunities (such as e.g. job offers) may be affected by the test results. The test administrator should spend enough time on giving the respondent a thorough introduction. He or she should have the opportunity to ask questions prior to the testing, which must be answered truthfully and accurately.

It is the test administrators' responsibility to ensure that the above requirements are fulfilled.

The respondent

MAP is intended for adults, thus respondents being 18 years or older. Data for standardization, norms and psychometric evaluation is therefore always collected from individuals being of 18 years or older. Note that testing of minors (under the age of 18), in most geographies and legal areas, require consent from the legal guardian. A test administrator may of course administer the process of consent, but it is the responsibility of the test administrator to ensure that it is done correctly and in accordance with legislation and good practice.

MAP is developed for use within the work and organizational setting, thus for selection, development and coaching etc. MAP is not intended to be used in a clinical setting or administered to individuals belonging to a clinical population.

Computer skills

Since MAP is a web-based assessment, a certain amount of computer skills and experience of working online is required. The respondent must be able to handle the necessary technical equipment such as a mouse and/or a keyboard. Prior to the testing, it is the responsibility of the test administrator to make sure that the technical aspects do not cause any difficulties for the respondent, as this may have a negative effect on the results.

It is thus the responsibility of the test administrator to inform themselves of the respondent's prerequisites in these matters. If there is even the slightest doubt, the respondent should be given the opportunity to, under the supervision of the test administrator, demonstrate that they possess the necessary skills to complete MAP. If a respondent is completely unfamiliar with the technical equipment (computer, tablet or smartphone) which is required to complete MAP or expresses a strong reluctance towards taking a computer-based assessment, MAP should not be administered.

Readability

Before starting the development of MAP, and throughout the entire development process, it has been a true ambition to keep instructions and items as simple, clear, straightforward and as short as possible. There is nevertheless a certain demand on the respondents´ level of reading comprehension. In Sweden, the readability index called LIX (Björnsson, 1968) is often used to estimate the level of complexity in a text. LIX is based on the average number of words per sentence and the number of long words (long words defined as words containing more than six letters). There are levels of difficulty that are generally agreed upon and which relate to different LIX-values, from very easy to read (children's books) to very difficult to read (Swedish bureaucratic language). The LIX-value for the items in the Swedish version of full MAP is 23, which is categorized as "very easy to read" and thus on the same level as children's books. The instructions in MAP have a LIX-value of 36, which is classified as "easy to read" and on the same level as fictional texts and popular magazines. Note however, that even if a text is classified as simple, factors such as severe reading and/or writing difficulties (dyslexia) or having another linguistic background (another mother tongue) may affect the understanding of content and thus the results. At present, there is no specific information regarding if, to what extent, and how MAP scores may be affected by this type of factors. It is therefore important that the test administrator ensures that the respondent has enough reading comprehension and appropriate linguistic background to understand the instructions and items so that the respondents may respond accurately.

Other impairments

Other impairments, including but not limited to perceptual, visual and cognitive impairments, that may have a negative effect on the test results should be identified, addressed and remedied by the test administrator before administration of MAP and any other assessment. The consequences and potential negative effects of a respondents´ specific impairment when being administrated MAP requires investigation by the test administrator, as well as deciding upon what actions are needed to most successfully accommodate and adapt the (overall) assessment process according to his or her specific needs.

Information to respondent before testing

The areas of application for MAP are both extensive and diverse and it is always recommended that the respondents are well informed before testing. Regardless of context, purpose of testing (development or selection), technical platform (via Ascend UI or the API) and administration mode (supervised or unsupervised) there is information and instructions that are necessary and crucial. Therefore, this information and instructions are provided to each respondent in a standardized fashion to ensure transparency, fairness and equal treatment among respondents. This information is not due to change or deletion.

Generic and standardized information and instructions

Respondents are provided with the generic and standardized information and instructions on screen and prior to testing. The information and instructions provided on screen and prior to starting the session include the following information:

- That MAP measures personality characteristics using statements describing different situations, behaviors, and preferences.
- That the respondents' task is to read each statement and consider to what extent it applies to oneself.
- That there are no "right" or "wrong" answers.
- That there will be four response options (Eng., Disagree, Somewhat disagree, Somewhat agree, Agree).
- That there will be two practice statements, and that the responses to them will not affect the results.
- That the respondent is to make sure that he/she can complete all statements without any distractions or interruptions.
- That there is no time limit, but it is recommended to not spend too much time on any specific statement.
- That it is possible to go back and change responses at any time during the session.
- That the respondent will be administered 200 items in total if taking full MAP and 75 if taking Essence.
- That the completion time is approximately 40 or 15 minutes respectively.

Note that the instructions are designed to be self-instructive; all instructions needed for completion of MAP are shown at the beginning of the session.

It is the test administrators' responsibility to ensure that each respondent understands how MAP is structured, how the items should be responded to, and how to complete the assessment.

Context specific information

In addition to the generic and standardized instructions (aimed at providing the most basic information and instructions on how to complete MAP) there may be other relevant

information a test administrator should provide to a respondent before testing, this includes but is not limited to:

- The purpose of testing.
- What type of test MAP is in relation to why it is being used in the present context.
- How MAP will be administered and what is required for completing the test (see the previous section Requirements and conditions of testing).
- If and how the test results will be used and stored, by whom, for how long and why. Note, this does not only include the platform Ascend by Assessio, but also how the test administrator/organization will use and store data.
- That the respondent has the right to choose whether the test score is to be included as part of the information about oneself.
- Whether feedback will be provided to the respondent and, if so, when it will be provided, in what format (e.g., standardized on screen, personal feedback, face-to-face meeting, over the phone), and what the feedback will contain.
- That the respondent should contact the test administrator in case of unexpected problems or questions arising during the test session.
- Contact details to the test administrator.

Thus, the above listed information is context specific information that the test administrator should inform all respondents about before testing according to best practice. A test administrator may also have additional information he/she wants to share with the respondents (for example details regarding the next step in a specific personnel selection process). Regardless, this type of context specific information is not generic and may thus not be delivered by Ascend in a standardized format. Context specific information needs to be adapted to each specific purpose and context of testing and may therefore be included in the editable mail-invitation available in the Ascend user interface. If MAP is delivered via any other UI (via the API), the test administrator is responsible for providing the information to the respondent in a suitable way.

More information about the rights and obligations of test distributors, test administrators, and respondents are to be found in international guidelines for testing (e.g., www.intestcom.org, www.efpa.eu/professional-development, www.iso.org/standard/56436.html) and is often provided by national psychologists' associations.

During the test session

As mentioned above, the testing should take place in a calm and quiet environment, in which the respondent is not disturbed. Phones and other disturbing elements should not be present in the room.

It is recommended that respondents respond to the statements in an even pace, not spend

too much time on a single item, and complete MAP in a coherent session. Technically however, there is no time limit for completing MAP and it is possible to postpone a test session and restart at a later point in time an infinite number of times.

Evidently, it is easier to make sure that these conditions are met in a supervised administration setting. Therefore, this mode of administration is always recommended. If MAP is administered in an unsupervised setting, it is the responsibility of the test administrator to make sure that these conditions are met during the test session and to take into account any uncertainties that might have affected the results.

After the test session

MAP presupposes that a trained test administrator is responsible for the administration of MAP, the decision of whether to give feedback or not, and for providing the actual feedback to the respondent. If feedback is provided, there are several standardized written options from which the test administrator may choose from. Note however that regardless of format, the test administrator is responsible for both the decision to provide feedback or not, the choice of format for the feedback, and the content of the feedback.

It is the test administrators' responsibility to ensure that all respondents leave a test session with the feeling of having been treated fairly, having understood the purpose of the testing, and feeling that they got the opportunity to take MAP under optimal conditions. If chosen by the test administrator, they should also have received feedback on their results, in a way that is perceived as fair, respectful and non-intrusive.

One of the most important aspects in this type of individual assessment is that the respondents know where to turn with questions. This information should thus be unambiguous and repeated clearly to the respondent, this includes before testing, during testing, after testing and after feedback has been provided. Listed below is some general information and guidelines for the interpretation and feedback on MAP results.

Theoretical model, interpretation and feedback

The theoretical model underlying MAP – the Five-Factor Model – is selected, defined and developed based on the accumulated research available on measurable personality traits and the identification of personality traits that are important for predicting different types of behavior in the workplace. The meaning of the scores generated by MAP is defined by this model, by the research conducted on this model, and by the empirical support provided by the test development work. The test development work and the empiricism underpinning the report and feedback structure is presented in Chapters 2, 4 and 5. Regardless if full MAP or Essence is administered and which context MAP is applied to, the results are based on the five scales. In addition, full MAP provides results on the 25 subscales.

C-scale and norm construction

The distribution of raw scores on each scale and subscale in MAP has been transformed to the standard C-scale, with a mean value of 5 and a standard deviation of 2. The limit values

for each C-score, expressed in z-scores, as well as the percentile limits for each C-score, are shown in Table A2. The percentage of the norm group falling within each C-score level is shown in the last column of the table.

The choice of the C-scale as a standard scale for the MAP test scores is based on a comprehensible and easily communicated range (0-10) and its natural connection to the properties of normal distribution. If the test scores are normally distributed, then the scale will represent the norm group accurately. The C-scale has a rougher classification of points than, for instance, the T-scale and is likely to convey the extent to which psychological test scores may differentiate among individuals. More finely divided scales encourage and easily lead to over-interpretations of minor differences within individuals (between different scale scores) and between individuals on the same scale. Raw scores transformed to C-scores are rounded downwards to integers to adequately represent the percentage in the population within each C-score interval.

Different types of information

Information generated by psychological tests may be divided into two distinct categories. These two categories are crucial to how the results should be used, what conclusions can be drawn and what kind of decision that may be taken based on the test results. The first category of information is of *descriptive* character. Based on such information, an individual's traits, such as their personality, may be *described* according to one or several test scores that are often compared (or put in relation) to a reference group, a so-called norm group. The second category relates the test scores to a specific set of behaviors – a so-called criteria, such as overall job performance or a manager's achievements. This information has thus been specifically designed to forecast, to *predict*.

The two categories of information – the descriptive and the predictive – are often confused and it is easy to make a subjective jump from a description of an individual's personality to a conclusion about an individual's future performance in a specific role and thus suitability for a specific job. However, describing an individual's strengths and weaknesses based on a personality test, or any other personality assessment tool, does not automatically mean that the individual is suitable or unsuitable for a specific task or position. Descriptive information is of course useful for other purposes, such as self-awareness and individual development.

Descriptive information

The vast majority of tests and other assessment methods generate data of descriptive character. An individual's traits are expressed in a test score, whose level is determined in relation to a relevant comparison group, a so-called norm group. A norm group may be described as a group of individuals, who have also taken the assessment under equivalent conditions, against which an individuals' score is compared or put in relation to.

It is important that the norm group is relevant for comparison so that the normative score (for MAP, the C-score) becomes meaningful and comprehensible for the test administrator *and* the respondents. The debate of appropriate size and composition of norm groups is constantly undergoing. Today commercially available assessments use samples of

convenience as their norm groups and so does MAP.

To the respondent, the comparison with the corresponding description will be comprehensible when it's done in relation to what they may think of as ”most others”. An individuals' relative image of his or her personality is usually based on comparisons with most other people and not on comparisons with, for example, a specific occupational group or a specific organization.

To describe a person's basic traits in relation to a specific and perhaps extreme group in terms of personality (for example a specific occupational group) may create confusion among respondents, as they are unable to recognize themselves.

The actual text-based descriptions provided in standardized feedback reports of how the personality traits manifests itself and how an individual come across to others are, in nearly all commercially available methods, put together by having a normal population as a reference point. Having a norm group being too far from a normal population (for example a specific occupational group) or changing the norm group without adapting the descriptions may thus create an imbalance in the relationship between interpretation, description, feedback and decision-making. The descriptions will thus lack validity.

Individuals who have already been tested or undergone psychological assessments on previous occasions may experience conflicting results as a consequence of various norm groups being applied in different assessments. Even if the norm group is explicit, which is not always the case, it is difficult for the respondents to keep their traits in mind throughout the assessment process, which might include multiple methods and comparison groups.

The relevance of comparing the respondent against a specific group may often be questioned. It is important that the respondent, regardless of the decision or the measure, perceives the testing and feedback as being understandable, relevant and meaningful, not only in the current situation but also in the future. For example, it is questionable to give a respondent the feedback that they are introverted, based on the comparison against a norm group representing a specific group (e.g., an occupational group) having high scores (and often a skewed distribution) on Extraversion, if he or she is actually outgoing, compared to a more general population.

For the description of results in MAP, the aim is to have a relevant and psychometrically sound norm group as possible for every language version, geography, culture etc. The goal is to provide test administrators and respondents with standardized scores that may be interpreted as relative to “most others”. The recommendation to use a norm group not representing a narrow and homogeneous norm group does not imply that comparisons against such groups could never be relevant and informative, only that the fundamental description of a respondent should be made in relation to a norm group being more general or average in character. In this way, the description of the respondent's remains intact and both the respondent and the test user will be able to avoid confusing discussions such as ”but am I extroverted or not?”. The latter is a common consequence of switching norm groups between reports, test sessions, or of using different tests and methods with different types of norm groups (e.g. norm groups based upon normal population, occupational

groups, gender or age). After describing an individual's personality, it is usually easier to proceed by adding and relating the individual's characteristics to the characteristics of other more specific groups (e.g., managers) that might be relevant. Note however, that such comparisons are still only descriptive in character. They do not automatically reveal anything regarding future performance.

The above reasoning is based on the fact that a distribution of scores in any group (name it norm group, reference group or any other group used for relative comparison) lack the link to any external criteria, such as performance. Such scores and results describe the average level on a single variable on the group level. In practice, this means that an if individual for example gain high test scores (compared to any norm group), it does not automatically imply that the individual will perform well. It will only reveal that the individual is more Extroverted, Emotionally stable, Agreeable, Conscientious or Open to new experiences than the average in the norm group. Norm groups, regardless of composition, may thus only serve as a way of describing an individual's personality but it does not provide information about future suitability or performance in any way.

The above is quite often unknown to practitioners, making them susceptible to sales arguments based solely on face validity related to "the supplier offers a norm group relevant to my type of candidates"-reasoning. This reasoning makes practitioners prone to pay extensively for the use of occupational specific (or even organization specific) norm groups for comparison. It also makes practitioners (customers) prone to dismiss suppliers not offering such, although their offer might be significantly more psychologically and psychometrically sound and cost efficient.

To connect traits with performance, and thereby estimate an individuals' suitability and/or probable level of performance for a specific task, role or function, a predictive approach is required. For more information regarding a predictive approach, please see the Targeted Prediction© Technical Manual (Sjöberg & Sjöberg, 2017).

Measurement error and the interpretation of differences between test scores

All psychological measurement, testing included, contains a certain degree of measurement error. These measurement errors affect not only the precision of single test scores but also the interpretation of differences between test scores.

Most methods, including tests, available on the commercial market traditionally withhold this type of deficiency or are less explicit in the actual impact of measurement error on actual test scores by for example only presenting reliability coefficients. This gives an impression of precision which does not exist in any method or measurement and increases the risk of over-interpretation of differences between scores.

To facilitate for the test administrator to take the measurement error into account when interpreting single test scores, the Standard Error of Measurement, SEM, has been calculated for all scales and subscales in MAP. The calculation of SEM takes the reliability and standard deviation of the scale/subscale into account and the following formula has been used to calculate SEM.

$$SEM = SD\sqrt{1 - r}$$

This formula generates an interval that covers the “true” value with 68% probability. A 68% probability level, however, is relatively low, even when the intervals are narrow. In practice, this value is often corrected to generate an interval which cover the "true" test score with the higher probability level of 95% expressed in C-scores, the interval becomes broader, but also safer. In MAP, all SEM are indicated with a 95% probability, and all intervals are rounded off to whole C-scores.

On average, SEM on the scale level for full MAP are approximately ±2 C-scores and ±1 C-score for the subscales. This means for example that a respondent gaining a C-score of 5 on the Agreeableness scale, the interval stated in the SEM column will be 3-7 C-scores (5 ± 2 C-scores). The SEM interval should be taken into account in all interpretations of test scores, including the interpretation of differences between test scores. Differences between test scores may either be between two individuals on a scale or a subscale, or between two different scales or subscales from one individual.

It is common that test administrators interpret differences between test scores although there are no statistically guaranteed differences. This over-interpretation of differences may result in an incorrect feedback to respondent(s) and may lead to incorrect decisions and interventions. SEM, however, may be used to get an indication of whether there is a probable difference between two test scores. Intervals that are not overlapping indicate a true difference, while overlapping intervals indicate a lower probability of there being a true difference.

Between scores for one individual

The above procedure is quite blunt and to be able to establish differences between test scores in a more reliable manner, some calculations are necessary. The measurement error for the difference between the test scores is expressed in the Standard Error of difference, SE_{diff}, and indicates how great the difference between two test scores must be for the difference to genuinely reflect different levels of the trait. Differential scores are always affected by greater measurement error compared to interpretation of scores from single scales. This is due to the measurement error coming from two directions (either from two individuals or two scales). To estimate an individual’s SE_{diff} in two uncorrelated scales (scales that have no intrinsic correlation and the content does not overlap), the following formula is used (Anastasi & Urbina, 1997):

$$SE_{diff} = \mathrm{SD}\sqrt{2 - \mathrm{r}^1 - \mathrm{r}^2}$$

The standard deviation for the C-scale is 2, and r^1 indicates the reliability of the first scale (or subscale), while r^2 indicates the reliability of the second scale (or subscale). To calculate, for example, an individual’s SE_{diff} between his or hers scores on the scales Conscientiousness (CO) and Extraversion (EX) which have a reliability (alpha) of .90 and .92 respectively, the following calculation is made:

$$2\sqrt{2 - .90 - .92} = 0.85$$

To determine how large a difference score can be obtained by chance on the .05 level (5% level), SE_{diff} is multiplied by 1.96. The result is 1.66, which may be rounded to 1.5 C-scores. This means that the difference between an individual's C-score on Conscientiousness and Extraversion should be at least 1.5 scores to be significant at the .05 level.

Between individuals

To calculate the SE_{diff} between two individuals' scores on the same scale (or subscale), SEM from the scale may be multiplied by 1.414, which is equivalent to 2 and indicates the contribution of two error sources (one for each scale) to the difference score.

SE_{diff} = 1.414 x SEM

An example that may be generalized to all scales in full MAP, since the SEM for each of these is estimated to 1 C-point, is the following:

Example: 1.414 x 1 = 1.414

As with the SEM, this formula indicates SE_{diff} with a 68% probability, and therefore needs to be corrected, by multiplying it by 1.96 (1.414 x 1.96), which is 2.77.

Hence, on the scale level the difference between two individual's scores should be close to 3 C-scores to be sure that a true difference is identified. As mentioned above, SE_{diff} may also be used to calculate the interval around of the test scores, and should be interpreted as the SEM of individual measurements. Note that even if two intervals do not overlap, a difference may only be interpreted as probable; it cannot be interpreted as a "true" difference without any uncertainty. Also note that, as shown by the formulas, the SEM and SE_{diff} are a direct function of a scale's or subscale's reliability. This automatically means that the SEM and SE_{diff} are generally larger for subscales than for scales, since subscales, being a function of a smaller number of items, have lower reliability.

Reliability in difference scores

If calculations of SE_{diff} are carried out in practice, the above formula is the most common approach. These formulas are often appropriate, at least when the measures are not correlated. Sometimes the formulas are considered sufficient for providing an indication of the validity in a difference score. For example, a measure of general mental ability and a measure of the personality trait Extraversion usually have zero correlation. By applying the aforementioned formula, reliable estimates of "real" differences between these measurements can be calculated. However, uncorrelated measurements of psychological traits are quite rare. The vast majority of psychological phenomena correlate to some degree and personality traits are no exception, although the consequences of this covariance are rarely attended to or taken into consideration.

The strength of the correlations between the factors in the Five-Factor Model varies, as do the correlations between the subscales, both within and across scales. The fact that there are correlations between this type of psychological phenomena is natural and even if the

correlations were zero, it would not be possible to use indicators as statements or interview questions to measure a separate and ”pure” psychological trait, free from measurement bias or correlations with other properties.

The strength of a correlation is somewhat affected by how constructs are defined, operationalized and measured. The way in which this is managed in a test development process varies between instruments and methods. The fact that the scales and subscales correlate is of importance, not only for the validity of the overall model, but also for the interpretation of scores and the differences between them. The reliability in a difference score is thus central, due to interpretations of differences between different scores being very common and because the difference between test scores and profile interpretations are actually made on correlated scales. By calculating the reliability, one may take this correlation into account and thereby reduce the risk of over-interpreting the differences.

To correct for the correlation, one needs to know the relationships between scales and between subscales. These are presented in Table A7.

The formula (Harvill, 1991) for estimating the reliability in a difference score is:

$r_{DD} = (r_{xx} + r_{yy} - 2r_{xy})/(2(1-r_{xy}))$

r_{xx} and r_{yy} correspond to the reliability of the scales and r_{xy} corresponds to the correlation between the two scales. It is obvious that the difference scores will be reliable if the single scales have high reliabilities themselves and there is a low correlation between the two scales. Difference scores will therefore always have a lower reliability and a greater SEM compared to single test scores (note that this might not always be reflected in standardized and rounded scores), which imply that difference scores always should be interpreted with great caution. For MAP (and all other methods assessment methods), this means once again that differences in subscale scores, being intercorrelated and less reliable, require relatively large C-score differences C to determine a probable true difference. Scales that are more reliable and generally less overlapping require smaller differences.

General principles for interpretation and feedback

MAP may be applied in many contexts and the context to some extent determines the way in which the results should and can be used adequately. For example, in a selection context the empirical correlation between scales and subscales respectively, and job performance should be the main focus. In a development context on the other hand, the individual face-to-face feedback session is an integrated part of the assessment process is likely to be the most relevant aspect. There might also be several and equivalent purposes of using a tool like MAP. Regardless of the purpose of use, the results are interpreted and the respondent is likely to receive some form of feedback. The following recommendations, which concern interpretation and feedback, are general in character and therefore all the stages might not apply to all contexts. There might be situations in which the interpretation or feedback benefit from being performed in a different order.

The meaning of test scores

The development of full MAP and Essence and the empirical relationship between the individual's responses to items and probable behaviors in situations outside the test situation are not the only parameters that affect the quality of the interpretation of results. The test administrator's understanding of the theoretical model, the underlying psychological constructs and the knowledge about the data forming the basis of the test scores is also relevant for the understanding of test results and thus the feedback. Prior to giving feedback to a respondent, the test administrator should engage in:

1. Getting to know the core of every psychological construct as defined in this manual (i.e. the construct underlying each scale, and subscale if using full MAP) and to avoid filling these with new meaning or making associations of one's own to behavior outside of what is described as the core of the construct. The possibility of making generalizations based on the test score is framed by the construct definitions.

2. Evaluating the results of the scales. The order of the scales in the feedback reports (AG, CO, ES, EX and OP) have no meaning in itself. The scales may be interpreted in whatever order suits the test administrator or according to what suits the purpose of use. In general, the following order and grouping is recommended for the interpretation and feedback of the scales:
   EX and AG reveal the way in which an individual acts and functions in the interpersonal sphere (*between individuals*) in terms of outgoing energy and what style one tends to have in social interactions. OP and CO are related to the way in which the individual's intrapersonal sphere (*within the individual*) operates in terms of new emotional experiences, and whether or not he or she is conscientious and focused on performance. ES is suitable to interpret lastly because it indicates an individual's general emotional adaptability and how he or she handles insecurity, adversity, stress and strain from an emotional perspective.

3. If using full MAP, evaluating the results on the subscales. Examine questions such as: Which subscales seem to be the driving forces for the results on the scale? Are average scores on scales reflecting a uniform or a varied profile of results in the subscale? Regarding high and low scores on this scale: Are there any subscales whose scores are going in the opposite direction, and so on. Keep in mind that subscales belonging to the same scale correlate, so there is a statistical tendency that results from the subscales will follow the same direction as the result on the scale, even if deviation from this trend is not unusual.

Unusual patterns of scores on the subscales might be difficult to understand, especially for new users, if too much emphasis is put on the adjectives chosen to label the scales and subscales. Keep in mind that the labels can never reflect a complete and perfect description of a psychological construct or trait, and they cannot describe how the construct distinguishes itself from other constructs and results on the scales or the subscales. Understanding what is actually measured is as important as knowing what is not measured in each scale or subscale. One should also keep in mind not to over-interpret differences connected to point 2 and 3 above and to consider the argument about SEM and reliability

throughout this process.

Average scores

The standardized written reports available for MAP (see later in this chapter) divide scores into four levels: low, average low, average high and high. Dividing average scores into average low (scores in the interval between the mean value of the norm group and 1 standard deviation below the mean) and average high (scores in the interval between the mean value of the norm group and 1 standard deviation above the mean) respectively makes it possible to differentiate to a slightly more detailed extent without risking over interpretation. The discussion below, however, refers to both average low and average high when discussing the challenges with and meaning of average results.

Sometimes average scores are perceived as more difficult to interpret and to give feedback on, compared to high or low scores. This might especially be the case when a respondent has many average results compared to the number of high or low results. In the case of most scales and subscales, one often feels greater certainty interpreting combinations of more extreme results and significantly different levels of test scores. When the score falls within the average range, we tend to be less certain of how to interpret and thus give feedback. Most respondents, more specifically 68% of all the subjects tested on a particular scale, will achieve results that correspond to ± 1 standard deviation from the mean value of the norm group. In MAP, this means that nearly 70% of the respondents will fall within the score interval of 3-6 C-score (5 ± 2 C-score), i.e. achieve average results. This is a statistical property by which a scale is constructed and defined, and which is a prerequisite for the scale's validity. A respondent with an average result on a scale or subscale are like most others in the norm group regarding this trait; this by itself provides information. Average results reveal that the particular trait is unlikely to be a strong characteristic for the respondent, thus, it will most likely not be perceived as one of his/her prominent traits. The expression of an "average trait" is likely to be more moderate than the expression of a trait on which the respondent has gained high or low scores (the total lack of a trait may be strongly characterizing). It is also important to note that the width of underlying constructs makes average results on scales represent a broader psychological meaning than average results on subscales.

Feedback

In line with current best practices and guidelines the overall recommendation is to provide respondents with feedback regarding their test results on MAP (note the difference between *test results* and the *overall basis for decision* such as a Targeted Prediction© score and an actual *decision* such as hired/not hired). Respondents should also be provided with the opportunity to ask questions regarding their test results.

The feedback, however, may be delivered in different ways. The appropriate form of feedback in a specific context is dependent upon several factors, for example the type of assessment (thus the actual content), the number of respondents which is nested in a financial aspect (face-to-face feedback is time consuming and thus expensive while written

standardized feedback is more cost effective), availability of the respondent (respondents may be at a remote location).

One of the most common ways to provide feedback on test results in practice is to provide the respondent with a standardized written feedback report generated by the system. This approach should be accompanied with a clear offer the respondent to return with questions about their results (via for example phone or email). The second most common approach to provide individual feedback is in an oral session between the test administrator and the respondent. This may be carried out via phone or in a face-to-face meeting. Traditionally, the results are reviewed, discussed and the respondent is provided with the opportunity to comment on the results. Often the test administrator will use the written standardized reports generated by the system as the basis in this feedback format. In general, the standardized feedback that are generated by the system and intended for respondents do not require any additional personal feedback by the test administrator.

If feedback on test results is carried out in an individual oral session it is recommended to start with an initial conversation regarding how the respondent experienced the test session, how it was perceived, whether the respondent feels that he/she was given the opportunity to provide an accurate picture of him-/herself, if he/she understood the instructions, if he or she was able to answer all statements with comfort and if the respondent would like to comment on any circumstances that he/she thinks may have affected the results. This should take place before the feedback of the results begins. After this phase, it is appropriate to describe the structure, the purpose and the length of the feedback session, and to encourage questions and reflections during the entire feedback session. Stress that no test (assessment) results are exact, they are always affected by measurement errors, and that adjectives used for the labeling of scales and subscales may never fully and justly reflect the underlying psychological construct.

Encourage the respondents to ask questions throughout the feedback conversation and as a test administrator; try to reject, rather than confirm, the hypotheses regarding the respondent's personality, that are suggested by the results.

Start the feedback by explaining why it is relevant to measure personality in the given context. Describe the overall structure of scales and subscales if using full MAP and give an overview of how the results will be reviewed. Then, start with EX and AG. Describe their position in the interpersonal sphere, their main content, and describe the subscales of the respective scale if using full MAP. Proceed to the respondent's score on the scale and the subscales that belong to it. Proceed in the same manner with OP and CO, by describing their intrapersonal characteristics and the main content of the scales and subscales, followed by the respondent's result. Conclude with the ES scale.

Standardized reports

The presentation of results on MAP are provided as test scores (C-scores) and as levels (low, average low, average high and high) with associated narrative descriptions. Results in the format of levels with corresponding narrative descriptions constitutes the different standardized reports.

Before describing the available standardized report types, please note that the Data Overview, available to the test administrator in the Ascend UI, is a compilation of the C-scores for the five scales (and 25 subscales if using full MAP) for all respondents in a project. The Data Overview is intended and available for the test administrator.

The standardized report types for MAP are described regarding content, structure and intended recipient in the following (for an overview of all reports, please see Appendix D). Please note that all reports available for MAP are individual reports, thus contain information about a specific respondent - there are no group reports available for results on MAP. All reports contain basic information regarding personality, the structure of the report, guidelines for interpretation, and important aspects to consider when interpreting results from an assessment such as MAP.

If applying full MAP, the following reports are available:

- The On screen-feedback Report, which is automatically made available to the respondent after completing MAP if the test administrator has set this up in the project management. The report provides results on the scale and subscale level by a graphic illustration of circles corresponding to the four result levels; low, average low, average high, and high. The respondent's results are illustrated by one of the circles being marked. Each pole, thus low and high, are anchored by a number of adjectives being representative of the content measured by the scale or subscale.

- The Descriptive Report presents the result level (low, average low, average high or high scores) on the scales and subscales in a figure. The result level on the scales and subscales are accompanied by a descriptive text outlining likely behaviors related to the result level.
  The content in the Descriptive report reflects the content in the Interpretive Report (see below), the tone and complexity of the descriptions however is somewhat adapted to be more suitable for a respondent (who is not a trained MAP user). This is important, due to the Descriptive Report being intended for the respondent. The Descriptive Report is only available to the test administrator but may be sent via email to the respondent as a pdf. Note however that there are no obstacles for using this report in a personal feedback-session as well.
  The last page in the Descriptive Report contains the Profile Overview. The profile is a graphic presentation of the result levels on all scales and subscales. The Profile Overview is well suited for the respondent when for example receiving personal feedback.

- The Interpretive Report presents the results on the scales and subscales in a figure which graphically illustrate the actual C-score and the result level (low, average low, average high or high scores). The results on the scales and subscales are accompanied by a descriptive text outlining likely behaviors related to the result level. The descriptive text on the scales are followed by characteristics that are likely to be areas of strengths and resources for the respondent. The next part brings up areas where the respondent may experience challenges and suggestions for behavioral development are presented. The Interpretive Report is intended for the

test administrator.
As mentioned above, the content of the Interpretive Report is reflected in the Descriptive Report. The tone, reasoning, use of psychological concepts and complexity of the descriptions are adapted for use by a trained test administrator. This is important, due to the Interpretive Report being intended for the test administrator. The Interpretive Report is generated by the system and available to the test administrator as a pdf. Note that this report may be used in a personal feedback-session as well but is not intended to be given to the respondent.
As in the Descriptive Report, the last page in the Interpretive Report contains the Profile Overview. The Profile Overview in the Interpretive Report however is a graphic presentation of C-scores and the results levels on all scales and subscales. The Profile Overview is well suited as a summary when providing feedback to a respondent and as support for skilled users well acquainted with personality measurement in general and MAP in particular.

- The Managerial Report presents the results graphically in the same way as in the Interpretive Report, it thus illustrates both the actual C-score and the result level (low, average low, average high or high scores) on scales and subscales. In similarity to the Interpretive Report, the results on the scales and subscales are accompanied by a descriptive text outlining likely behaviors related to the result level. This is followed by characteristics that are likely to be areas of strengths and resources, and then areas where the respondent may experience challenges. The challenges are accompanied by suggestions for behavioral development. The difference is that all of the above is framed for the respondent within the role of a manager. This includes the scale and subscale descriptions.
  Like the Interpretive Report, the content of the Managerial Report is reflected in the Descriptive Report. The tone, reasoning, use of psychological concepts and complexity of the descriptions are adapted for use by a trained test administrator. This is important, due to the Interpretive Report being intended for the test administrator. The Managerial Report is generated by the system and available to the test administrator as a pdf. Note that this report may be used in a personal feedback-session as well but is not intended to be given to the respondent. For this latter purpose and use, the Descriptive Report is recommended.
  As in the Descriptive and interpretive Reports, the last page in the Managerial Report contains the Profile Overview. The Profile Overview in the Managerial Report is identical to the Profile Overview in the Interpretive Report with its graphic presentation of C-scores *and* the results levels on all scales and subscales. The Profile Overview is well suited as a summary when providing feedback to a respondent and as support for skilled users well acquainted with personality measurement in general and MAP in particular.

If applying Essence, the following reports are available:

- The On screen-feedback Report, which is identical to the corresponding report for full MAP with the exception of not providing results on the subscale level.

- The Descriptive Report , which is identical to the corresponding report for full MAP

with the exception of not providing results on the subscale level.

- The Interpretive Report which is identical to the corresponding report for full MAP with the exception of not providing results on the subscale level.

CHAPTER 5

MAP - Psychometric properties and standardization sample

Standardization sample

Detailed information regarding the process of collecting data and the sample (n=569) used for standardization and thus representing the initial norm group (applied from publication until February 2018), is described in Chapter 2. In Appendix A Table A1, the sample is described regarding age, gender and educational level and compared to the Swedish normal population.

Descriptive statistics

Mean values, standard deviations, minimum and maximum values, and reliability (Cronbach alpha) for scales and subscales are presented in Table A3. The corresponding values for females and males are presented in Table A4 and A5. In appendix table A9 the correlations between the scales, as well as the correlations with age, gender and educational level are presented. In table A6 the correlations between all scales and subscales for the sample of n=569 is outlined.

Factor Structure

Testing the overall theoretical model is a critical part of the test development process. The theoretical model and the instrument's ability to measure the constructs according to the hypothesized theoretical model determines whether the results generated by the instrument are valid and thus may be interpreted according to what is postulated by the theoretical model. Hence, this part of the test development process is critical to obtain an indication of the instrument's overall validity. The result of not having a solid theoretical model as a starting point is often that the empirical model does not measure up, or that there is a significant gap between the theoretical and the empirical model. This situation indicates that what is actually measured does not reflect the structure and content of the theoretical model. In practice, this implies that there is no support for interpreting and generalizing results according to what the theoretical model postulates, thus there is no insurance that the interpretations are possible, theoretically meaningful, and accurate.

The theoretical model postulates all assumptions for the model; which constructs are included, their structure and relationships between them. Altogether this makes up the complete theoretical model. For MAP, the theoretical model is based upon five factors, each measured with a separate scale, which in turn have five so called facets, each measured with a separate subscale. Psychological tests, and other assessment tools as well, often lack technical documentation, and for those who report technical documentation, results of more rigorous testing of the postulated factor structure tend to be missing. Usually, so-

called principal component analyses are carried out and the results presented.

Such an analysis however may, in its basic assumption, not be compared to any predetermined theoretical model. This is a so-called exploratory approach and makes no real demands on the empirical model. The prevalent view is that this is not an appropriate course of action when there is a theoretical model to start from (e.g., the five-factor model of personality). This means that when there is a theoretical model (which is the most common situation) a so-called confirmatory factor analysis (CFA) is the appropriate analysis to conduct. Borsboom (2006) expresses this as follows:

> ”Clearly, there is no conceivable way in which the Big Five [FFM] could cause subtest scores on personality tests (or anything else, for that matter), unless they were in fact not principal components, but belonged to a more interesting species of theoretical entities; for instance, latent variables. Testing the hypothesis that the personality traits in question are causal determinants of personality test scores thus, at a minimum, requires the specification of a reflective latent variable model (Edwards & Bagozzi, 2000). A good example would be a Confirmatory Factor Analysis (CFA) model.” (p. 426).

Thus, to test the overall model, a CFA was carried out in the program AMOS 18.0. The analysis was conducted with the correlation matrix (Table A6) as an input in the analysis. Since the theoretical model specifies five factors with five subscales for each factor, this was the first model to be tested.

The requirements for an “acceptable” fit of the model are assessed based on various statistical measures. One of these is the chi square (χ^2), which examines the difference between the theoretical and empirical model in terms of significance. The correlations in the assumed model are compared to the correlations in the empirical model; a significant value for the chi square means that there is a discrepancy between the theoretical model and the measurement model. The results of the analysis suggest a statistically significant difference between the model and the data (df = 569; χ^2 = 2924.25, p <.001). It should be considered however that measurements of personality traits seldom reveal non-significant chi square values.

Moreover, the chi square value is often regarded as an overly conservative and unrealistic measure of fit between model and data. This partly depends on the fact that the chi square is influenced by the sample size – the larger sample, the more likely it is to obtain a significant value. Analyses based on a selection of more than 200 individuals are likely to generate a significant chi square.

To create a more realistic assessment of the model’s fit other adjustment measures are recommended, e.g., the Root Mean Square Error of Approximation (RMSEA) and the Comparative Fit Index (CFI). There is, however, an ongoing scientific debate about the limits that should apply to these adjustment measures and there is no generally accepted recommendation available. Most analysts, however, choose to follow Hu & Bentlers’ (1999) recommendation that the CFI should be at least .95 to be acceptable, and that models with an RMSEA above .06 should be discarded.

The adjustment measures from the analysis (RMSEA =. 13; CFI =. 60) indicated that the

measurement is in need for improvement to reach fully acceptable levels. From this point of view, a measurement model may never be ”finalized” or perfect in any absolute sense. This kind of improvement may and should be carried out on a continuous basis for all types of psychological measurements.

To allow a more detailed evaluation of the factor structure, the factor loadings from the CFA-analysis are presented in Table A7, while the correlations between the theoretical constructs are shown in Table A8. These are examples of correlations that could be lowered to achieve a better fit of the model.

Concurrent Criterion-related validity – Managerial performance

When individuals, that are already selected, hired and employed with a company, takes a psychological assessment and at the same point in time are evaluated regarding their performance (criterion-data), it is referred to as a concurrent validation. To investigate the concurrent validity of MAP, a group of managers was administered MAP and at the same point in time performance (criterion) data was collected according to the process described below.

The validation sample (N=73) consists of a group of managers from two administrations of a large Swedish municipality. The average age in the sample was 44 years (SD = 9) and 58% were female. The level of education ranged from elementary school to postgraduate education, with a majority - 57% - having completed at least three years of high school education and/or some form of tertiary education.

Supervisory ratings of managerial performance

In the validation study, two summarized indexes constituted the overall criterion Managerial Performance. A rating questionnaire with 28 items was designed to describe the employee’s (in this case the manager) behaviors at work. The questionnaire included items such as: The manager/ leader carries out the work carefully and thoroughly, or, The manager/leader has a positive attitude. Items were answered on a 4-point Likert scale (1 = Strongly disagree, 2 = Partially disagree, 3 = Partially agree, 4 = Strongly agree). The ratings were carried out by each managers supervisor. In a factor analysis including all 28 items, one factor dominated heavily with an Eigenvalue value of 12, explaining 43% of the total variance. The second factor had an Eigenvalue of 2, and thus explained only 2% of the total variance. Based on these results, it was decided to summarize all items into an overall index representing the criteria of Supervisory Performance Rating of managerial performance. The reliability (Cronbach’s alpha) of this overall Supervisory Performance Rating was estimated to α=.94.

Subordinates ratings of managerial performance

The managers performance was also measured using subordinate’s ratings. Data underlying this measure was collected in an annual employee survey at the municipality. Items regarding their managers behaviors was administered to all employees: *My closest manager is good at planning and organizing work*; *I know what my closest manager expects of me with regard to my work*; *My immediate boss makes sure that what we decided to do really gets done*; *My closest manager communicates openly and honestly*. The questions were answered

on a 5-point Likert scale where 1 corresponded to “Strongly disagree” and 5 corresponded to “Strongly agree”. A factor analysis including all 8 items, showed that one factor dominated with an Eigenvalue of 5, explaining 67% of the total variance. The second factor had an Eigenvalue of less than 1 (.82) and thus only explained a small part of the total variance. Based on these results, it was decided to summarize all items into an overall index representing the criteria of Subordinates performance rating. The reliability (Cronbach’s alpha) of this Subordinates Performance Rating was estimated to α=.94.

Overall Managerial Performance

In a second step, the Supervisory and Subordinates Performance Ratings were summarized in an overall index representing the overall criteria of Managerial Performance (Cronbach’s alpha=.94). An assumption was made that the two indexes partly measure different constructs: different questions were asked to the subordinate and the supervisor and there were different assessors. As expected, the correlation between the two indexes was not significantly different from zero (r =.12; p <.05), implying the two measures to be supplementary to each other.

Before determining concurrent validity, it is crucial to take the amount of restriction of range in the sample into account. This is because a concurrent validation may result in an underestimation of the actual validity, since the organization is unlikely to have hired the staff in a random manner. It is very likely that there is a restriction of range in data among already employed staff. To investigate the occurrence and size of the restriction, the variation in the standardization sample of the overall managerial performance composite score (SD =. 47) was compared to the variation among the 73 managers who participated in the validation study (SD =. 35). Dividing .35 by .47 estimates the restriction of range to μ=. 73. This value is used to correct for the restriction of range in the current study making the estimated validity to reflect the actual validity of MAP when used in practice for selection.

Another reason why this type of studies may underestimate the validity is that there are reliability deficiencies in the estimation of performance. Since the Supervisory Performance Rating is produced by only one supervisor, it is not possible to estimate the reliability of the performance rating. This would require multiple raters for each subject (in this study; manager). To overcome this and still be able to take the reliability into account, the most reliable meta-analytic results were used to estimate the reliability in the performance rating. The average reliability for this type of performance ratings has been estimated to .52 (Viswesvaran, Ones & Schmidt, 1996). Table A14 shows the measured correlation (r) between the overall managerial performance composite score and the overall criterion Managerial Performance, the covariance is corrected only for restriction of range (r_1), and finally the operational validity (r_2) is corrected for the range restriction *and* reliability deficiencies in the criterion (performance ratings). The observed correlation between the overall managerial performance composite score and the overall criteria Managerial Performance was significant (r =. 31; p <.05). Overall, the conclusion is that the scale scores from MAP mechanically weighted together into an overall composite score meets the requirements for being a good basis in the decision-making process for selection of managers based on probability of future performance.

It may also be of interest to investigate how the two indexes relate to the entire personality model which MAP is based upon, including the subscales. To investigate this, the correlation (r) and the operational validity (r_2) for both the Supervisory and Subordinates ratings were calculated. The results, presented in Table A15, is to be viewed as descriptive information highlighting the difference between assessing an individual's personality traits in relation to performance from above (Supervisory) and below (Subordinate).

Descriptive information and the purpose of norm groups

As previously outlined, MAP generate two types of information: descriptive and predictive information. Evidence for the usefulness of MAP for prediction of managerial performance is provided under the section on criterion-related validity above. Practitioners who wants a standardized and mechanical process for interpretation of MAP scores with the sole purpose of predicting performance in several roles, managerial and leadership is one, may apply MAP within the Targeted Prediction© framework. Please see the Targeted Prediction© Technical Manual (Sjöberg & Sjöberg, 2017) for more information. The descriptive information constitutes C-scores, result levels (low, average low, average high and high) and narratives outlining typical and likely characteristics and behaviors for a respondent. The results and descriptions are in comparison to a norm group.

The purpose of a norm group is to provide meaning to a test score. By collecting data from a group of individuals and using the groups' distribution of test scores to compare single test scores against, meaning is assigned to the test score. The norm-referenced result is usually accompanied by a narrative description.

For the comparison to be meaningful to all stakeholders, the norm group should reflect a population which is relevant and comprehensible to test administrators and to respondents. The norm groups' test score distribution should also correspond to an approximative normal distribution (for all scales and subscales) to fulfill one of classical test theory's basic assumptions. Serious violation of this assumption may invalidate the interpretation of the test score.

In practice, it is common with customers demanding or requesting specific norm groups, e.g. for an industry, occupational group, or even a specific organization. Customers often consider it an advantage if the supplier offers this type of specific norm groups, it is sometimes even a requirement in public vending. Thus, providing the test administrator with the opportunity of choosing and "switching" between different norm groups may be regarded as a competitive strength. This however, set up demands and requirements few, if any, supplier fulfill. One problem is that when using a specific norm group, it may be hard for the test user and/ or the respondent to relate to the norm group, for example for example relating to a norm group of accountants at company X. Another example is a candidate applying for a job in the service sector being compared to a norm group consisting of employees already working in the service industry. This may result in the candidate being described as less Extrovert (thus Introvert) despite him/her being slightly above average compared to a more diverse norm group (normal population) than a group of sales people who are likely to be Extrovert on the group level. Moreover, specific groups of this kind rarely constitute approximate normal distributions. This causes, among other things, minor

differences between (raw) test scores to produce invalid large differences in the standardized norm-based scores. Another problem of applying specific norm groups, and of offering the possibility to switch between norm groups, is that the interpretive narratives in the reports (with descriptions of the traits and of the result levels) may become invalid. Usually, the narratives are formulated in relation to a more general population, either a normal population or a general population such as adults in work or applying for work (sample of convenience). Descriptive narratives are seldom (ever?) adapted to specific norm groups. In terms of interpretation and feedback, the above means that it becomes a gap between the results and the descriptive narratives. Without a deeper knowledge of psychometrics, test theory and measurement of individual characteristics, the use of specific norm groups and the possibility for test users to switch between norm groups may generate questions about why an individual's normative test scores and descriptions differ to such great extent.

One explanation to why focus among practitioners often is directed against the questions of norm groups is the misconception that the norm group by default is linked to performance. It is often the perception that the comparison between an individuals' test score and the norm group's score provide information about the individuals ability, performance or suitability of some kind. For example, if an individual in comparison to a group of managers (could be current managers in a specific organization) achieve high scores or a pattern of scores similar to the group of managers, this would be interpreted as an indicator of the individuals' high performance or suitability for the manager role. This is, however, a false assumption. A norm group does not automatically provide information about an individuals' overall suitability or future performance. Hence, it does not affect the ranking of individuals according to suitability or performance in relation to the criterion (e.g., performance as a manager). This type of information is predictive in nature and to gain this information, another approach is needed (Sjöberg & Sjöberg, 2017).

In MAP, the comparison and evaluation of an individual's trait level is always related to a norm group relevant for the language version, geography and the area of application (work and organization context). For information about the norm groups currently applied, please see the section Norm update 2018 below. In turn, the descriptive narratives generated in the standardized reports are written with an adult working population as the target recipients when describing individual behavior although the descriptions are relative in nature.

Norm update 2019

The standardization sample, described in previous chapters and in Appendix A, constituted the initial norm group upon which the first norms were based. These norms, thus based on a Swedish sample (n=569), were implemented when MAP was first published in year 2011.

Due to the extensive popularity of MAP following the year of publication, a massive amount of data has since been collected. Several translations and adaptations into additional language versions has been developed between 2011 and 2019 and MAP has also been implemented in the new digital platform Ascend by Assessio. In 2019, enough additional data had been collected and it was decided to replace the initial Swedish norm group with an updated and extended Swedish norm group. Additionally, it was decided to develop local norm groups for the language versions for which enough data was collected, and to provide

an International norm group.

By delivering local norm groups for language versions where possible, using larger sample sizes in general, and by making an international norm group available, the new norms will make up a great improvement for all MAP users. The work with updating norms continuously is also in line with the requirements from independent test review systems and is thus necessary in order to comply with the rules for approved certifications and quality assurance processes.

All data in this update is collected via Ascend by Assessio between the years 2017 and 2019. All data is thus collected as part of an assessment process, traditionally for either development or selections purposes, in working life. This imply that the new norm groups, compared to the old Swedish norm group, to a somewhat greater extent represents a working population because of a higher probability of the respondents being either in work, applying for work, being in transfer between positions, and/or being in individual or group development programs.

Initial analysis of the collected data showed that five local norm groups could be implemented (US English, Swedish, Norwegian and Finnish). It is important to note that the local norm groups consist of data only from respondents who have completed the language version of MAP corresponding to the respondents chosen native language. This criteria for inclusion was enforced to ensure that the norm groups for the US English, Swedish, Norwegian and Finnish language versions were closely connected to a geography and culture (e.g., the Swedish language version is highly likely to have been responded to by native Swedish speakers within the geography and cultural region “Sweden”).

The following language versions getting local norms in the 2019 norm update (sample size within brackets) are:

US English (N=3 919)

Swedish (N=188 504)

Norwegian (N=4 821)

Finnish (N=8 883)

In addition, the local norm groups together constitute the International norm group of N=205 217. Demographic information for the norm groups, mean values, standard deviations and reliability estimates for both scales and subscales are provided in Appendix B.

In general, the mean values in the new norm groups are higher compared to the initial Swedish norm group (N=569). The overall elevation of mean values is likely due to the differences in composition of the samples upon which the initial and the new norms respectively are calculated. The initial Swedish norm group consisted of respondents randomly sampled from the Swedish normal population while the new norms are calculated based on samples of convenience which in this case is data collected from a population of individuals active in the world of work or applying for work.

The magnitude of the elevation of mean values across the scales between the initial Swedish

norm group and the new norm groups is apparent comparing the mean values presented in Appendix A and B respectively. Note that analyses showed small differences *between* the new local norm groups – the significant differences are between the initial Swedish norm group and the new norm groups – making it relevant to focus on the International norm group for comparison. The latter also led to the conclusion that weighting by sample size was not required for the International norm group.

To facilitate mean level and distribution comparison the differences on the scale level between the initial Swedish norm group and the new International norm group, figures illustrating the distributions are included in Appendix B along with the differences estimated using Cohens *d*. As Cohens *d* estimate the number of standard deviations and the standard deviation is 2 for the applied C-scale, the differences in C-scores may be estimated. The consequences, on the standardized C-score level (which is the level of interpretation and thus relevant for test administrators), when updating from the initial Swedish norm group to the new International norm group (and the local norm groups due to them being approximately similar) is an overall lowering of C-scores for the respondents. Conscientiousness is the scale with the largest difference, approximately 3 C-scores and Emotional stability shows a difference of almost 3 C-scores. The other three scales show smaller differences, for Extraversion and Agreeableness the difference is slightly more than 1.5 C-scores and for Openness the differences are approximately 1 C-score.

For the International norm group, the correlations with age at the scale level are presented in Appendix B. The results show that the correlations with age in general are somewhat lower compared to the corresponding estimates presented for the initial Swedish norm group (N=569) presented in Appendix A. In Appendix E and F respectively, the development, psychometric properties and additional evidence for the validity of the Norwegian and Swedish language versions of MAP is outlined.

In Ascend by Assessio

In Ascend by Assessio, the 2019 norm update will be implemented according to the following logic:

Each language version will by default be connected to *one* norm group.

Application of norm group will follow the *chosen language of administration*, note that this choice is done by the respondent. It is thus not possible for a test administrator to choose norm group.

A language version will by default be connected to its local norm group (if available), and to that norm group only.

There is one exception to the above: for the language versions with local norm groups available, the test administrator will have the possibility to enforce the International norm group when setting up a project. Assessio customer support will provide the test administrator with this feature if relevant.

The remaining language versions will by default be connected to the new International norm

group.

CHAPTER 6

MAP Essence

Due to the popularity of MAP in the market, the number of assessments from publishing in year 2011 until 2017, has been extensive. MAP has been particularly used in the context of selection. Consequently, and due to the fact that selection processes seldom require or benefit from assessing personality at a lower order trait level than the FFM-level, it was decided to develop a short personality test based on MAP. This product is labeled MAP Essence (in the following labeled "Essence" to be distinguished from the full version of MAP which is labeled "MAP"). The aim of the development of Essence was thus to provide a reliable and valid assessment of personality at the FFM-level based upon MAP but with a shorter response time for the respondent.

Essence may be used as a single assessment or as part of a process with several assessments set up by the test administrator in Ascend. It may also be used as a component along with Matrigma (Mabon & Sjöberg, 2017) or Adaptive Matrigma (Mabon, Niemelä, Sjöberg & Sjöberg, 2017) within the Targeted Prediction© (Sjöberg & Sjöberg, 2017) framework. The aimed areas of application are thus mainly description of individuals according to the FFM-level of personality and predictive purposes (relevant in a selection context).

In summary, Essence is a short version of MAP. It is based on items included in MAP. Essence provides a brief but comprehensive measure of the five domains of personality (Agreeableness, Conscientiousness, Emotional Stability, Extraversion and Openness). It consists of five 15-item scales, thus 75 items in total, each scale measuring one FFM domain. Information on specific facets (subscales) within each domain (scale) is not provided by Essence.

Development

The items in Essence are a selection of items from the 200 items in MAP. This implies that scores on Essence may be interpreted as *indicators* of scores on MAP. Note that the scales scores of Essence are somewhat less reliable and valid compared to the scales on the full MAP. This is due to the lower number of indicators (statements/items) in Essence.

Sample

The development work was carried out in several steps focusing on multiple aspects of test score quality, this work is described below. The sample used for the analyses consisted of N=25 733 in total. This sample of convenience was collected via Assessio's web-based test platform Assessio Select. Demographic information regarding this sample is presented in Table C1 along with descriptive statistics and reliability of the five scales.

Process

The items included in Essence have been selected on both theoretical and empirical requirements. The requirements were formulated before the final selection of items were made. The requirements were:

- Each item content should (theoretically) reflect the factor in question
- Each item should correlate most highly with the scale to which it belongs
- Item content should not overlap theoretically

The overall goal with the development of Essence was to develop a short assessment, preferably taking half the time responding to compared to MAP. Therefore, it was decided to strive for a number of items below 100 as it would keep the time of administration to approximately 15 minutes or less.

First, with the aim of ensuring the content validity, and in the long run the construct validity, it was decided to use the subscale structure to maintain as much of the construct and content validity as possible. It would for example be possible to identify and use the most appropriate items at the scale level, not using the subscale structure as a structure. This approach may likely benefit the reliability (alpha) but could render a decline in construct validity due to item content not ensuring coverage of the whole construct. Practically, it would for example be possible that the main part of items came from the same subscale. This would result in construct under representation and thus lack of content validity.

With an equal number of items drawn from the subscales this imply a maximum of four items representing each subscale. To identify the most suitable items from the subscales the item-total correlations between each single item and the subscale was computed. This ended up with the identification of a short version consisting of three items from each subscale, thus 15 items representing each scale, and thus 75 items in total. At this point a qualitative review was conducted to ensure that item content did not overlap theoretically. Few adjustments were made in his phase. Three items were replaced by the second-best item in that subscale due to theoretical overlap with another item.

To further support the construct validity, analyses supporting both the divergent and convergent validity of the scales was conducted. In psychological testing, divergent validity is used to determine if a measure (scale/subscale) is too similar to another measure. If a measure is found to correlate too strongly (or be too similar) with another measure it suggests that the measures are capturing the same construct and are too alike to be considered different. An example would be a test used by a company for hiring purposes that measures how proficient someone is at a specific skill. If the test correlates too strongly with an IQ test then it essentially is just another IQ test instead of measuring something uniquely different.

To establish convergent validity for a measure, evidence is needed to show that a measure really is empirically related to other measures which it from a theoretical perspective should correlate with. Convergent and divergent or discriminant validity are both considered subcategories or subtypes of construct validity. Most important is to recognize that the two

aspects of validity work together - if you can demonstrate that you have evidence for both convergent and discriminant validity, you have per definition demonstrated that there is evidence for construct validity. But, neither one alone is sufficient for establishing construct validity.

In general, the strive is for the convergent correlations to be as high as possible and discriminant ones to be as low as possible, but there is no hard and fast rule. Regardless, the convergent correlations should always be higher than the discriminant ones.

In Table C3 the correlations between Essence and MAP are presented. The difference between the average discriminant correlations off-diagonal and the diagonal correlation (convergent) is relatively high, which provides support for the construct validity of Essence.

To test the overall model, a CFA was carried out in the program Lavaan (Rosseel, 2012). Since the theoretical model specifies five factors with five facets this was the model to be tested. As mentioned previously, the requirements for what is an "acceptable" fit of the model are assessed based on various statistical measures. One of these is the chi square (χ^2), which examines the difference between the theoretical and empirical model in terms of significance. The correlations in the assumed model are compared to the correlations in the empirical model; a significant value for the chi square means that there is a discrepancy between the theoretical model and the measurement model. The results of the analysis suggest a statistically significant difference between the model and the data (df = 569; χ^2 = 60289.93, p <.001). Measurements of personality traits, however, seldom reveal non-significant chi square values (which is often regarded as an overly conservative and unrealistic measure of fit between model and data) which partly depends on the fact that the chi square is influenced by sample size. The sample size in this analysis is likely to render a significant value.

Instead, recommended adjustment measures are for example the Root Mean Square Error of Approximation (RMSEA) and the Comparative Fit Index (CFI). The adjustment measures from the analysis (RMSEA =. 09; CFI =. 69) indicated that the measurement is in need for some improvement to reach fully acceptable levels.

Norm groups

The norm groups for Essence are based on the same samples as for the MAP 2019 norm update. The compositions, demographics, structure and applied logic is thus identical to that of MAP (which is outlined in Chapter 5 under the section Norm update 2019). Thus, Essence is delivered in Ascend by Assessio with local norm groups for the following language versions:

Swedish (N=188 504)

Norwegian (N=4 821)

Finnish (N=8 883)

US English (N=3 919)

For language versions without local norm groups, the International norm group of N=205 217 is available and applied. Demographic information for the norm groups are provided in Table B1, while the mean values, standard deviations and reliability estimates for the International and the local norm groups are provided in Appendix C. In table C4 correlations with age and gender are presented for the International norm group.

References

Allport, G. W. & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs*, 47 (211), 171.

Anastasi, A. & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.

Barrick, M. R. & Mount, M. K. & Judge, T. A. (2001). Personality and job performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of selection and Assessment*, 9, 9–30.

Björnsson, C. H. (1968). *Läsbarhet*. Stockholm: Liber.

Borsboom, D. (2006). The attack of the psychometricians. *Psykometrika*, 71, 425–440. Costa, P. T. & McCrae, R. R. (1982). An approach to the attribution of aging, period, and cohort effects. *Psychological Bulletin*, 92, 235–250.

Costa, P. T. & McCrae, R. R. (1985). *Revised NEO Personality Inventory* (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) *Professional manual*. Odessa, FL: Psychological Assessment Resources. Svensk version: NEO-PI-RTM. Stockholm: Psykologiförlaget AB, 2003.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist* 34, 571–582.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73, 1246–1256.

Edwards, J. R. & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.

Freyd, M. (1926). The statistical viewpoint in vocational selection. *Journal of Applied Psychology*, 4, 349–356.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E. & Nelson, C. (2000). Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, 1, 19–30.

Harvill, L. M. (1991). An NCME Instructional Module on Standard Error of Measurement. *Educational Measurement: Issues and Practice*, 10, 33-41.

Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis*. CA: Sage Publications.

Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.

Internationella riktlinjer för testanvändning (2000). Stiftelsen för Tillämpad Psykologi/Sveriges psykologförbund.

ISO10667 (2010). *Bedömningstjänster i arbetslivet – Processer och metoder för bedömning av*

människor i organisationer. Del 1: Krav på uppdragsgivare. Del 2: Krav på leverantör, 2010 (No. 10667.1.). Stockholm: Author.

Judge, T. A., Bono, J. E., Ilies, R. & Gerhardt, M. W. (2002). Personality and Leadership: A Qualitative Review. *Journal of Applied Psychology*, 87 (4), 765–780.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, 3, 296–310.

Mabon, H. (2005). *Arbetspsykologisk testing. Om urvalsmetoder i arbetslivet*. Stockholm: Assessio International AB.

Mabon, H., Niemelä, F., Sjöberg. A, & Sjöberg, S. (2017). *Matrigma Adaptive.* Technical Manual. Stockholm: Assessio International AB.

Mabon, H., & Sjöberg, A. (2017). *Matrigma.* Technical Manual. Stockholm: Assessio International AB.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. *A theoretical analysis and a review of evidence. Minneapolis*; University of Minnesota Press.

Mischel, W. (1968). *Personality and assessment*. New York: Academic.

Rosseel, Yves (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Sjöberg, S., Sjöberg, A. (2007). *Measuring Integrity*. Manual. Stockholm: Assessio.

Sjöberg, A. R., Sjöberg, S. & Forssén, K. (2006). *Predicting Job Performance*. Manual. Stockholm: Assessio AB.

Sjöberg. A, & Sjöberg, S. (2017). *Targeted Prediction*. Technical manual. Stockholm: Assessio International AB.

Schmidt, F. L., Shaffer. J. A. & Oh, I. S. (2008). Increased accuracy for range restriction corrections: implications for the role of personality and general mental ability in job and training performance. *Personell Psychology*, 61, 827–868.

Tupes, E. C. & Christal, R. E. (1961, 1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, 60, 225–251.

Viswesvaran, C., Ones, D. S. & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557–574.

Viteles, M. S. (1925). The clinical viewpoint in vocational selection. *Journal of Applied Psychology*, 2, 131–138.

Appendix A

Table A1
Comparison between the norm group (n = 569) and the population by age, sex and educational level

Age Group	**Norm group**	**Population**
18-30 years	18%	25%
31-40 years	15%	23%
41-50 years	17%	21%
51-60 years	13%	23%
61-65	9%	8%
No information available	28%	
Gender		
Women	64%	50%
Men	36%	50%
Educational level		
Elementary school	10%	25%
Elementary school and 2 years of upper secondary school	11%	26%
Elementary school and 3-4 years of upper secondary school	20%	19%
Less than 3 years of tertiary education	15%	13%
3 or more years of tertiary education	36%	16%
Postgraduate studies	1%	1%
Other	3%	
No information available	3%	

Table A2

The properties of the standard C-scale

C-score	z-Score limit	Percentile limit	% within interval
	+2.75	99.7	
10			0.9
	+2.25	98.8	
9			2.8
	+1.75	96.0	
8			6.6
	+1.25	89.4	
7			12.1
	+0.75	77.3	
6			17.4
	+0.25	59.9	
5			19.8
	–0.25	40.1	
4			17.4
	–0.75	22.7	
3			12.1
	–1.25	10.6	
2			6.6
	–1.75	4.0	
1			2.8
	–2.25	1.2	
0			0.9
	–2.75	0.3	

Table A3

Mean values, standard deviations, minimum and maximum values, and reliability (Cronbach's alpha) in the scales and subscales of the norm group (n = 569)

Scale/ subscale		M	SD	Min	Max	Alpha
AG	Agreeableness	121.50	11.31	82	152	.84
CO	Conscientiousness	119.58	13.55	76	155	.90
ES	Emotional Stability	113.27	15.10	56	150	.90
EX	Extraversion	109.15	15.85	52	152	.92
OP	Openness	105.24	14.35	70	148	.88
AG1	Trust	24.80	3.64	12	32	.72
AG2	Communication	22.62	3.36	11	31	.64
AG3	Altruism	25.13	3.03	11	32	.64
AG4	Compassion	23.96	3.32	11	32	.66
AG5	Affection	24.98	3.38	14	32	.69
CO1	Intensity	24.91	3.55	11	32	.78
CO2	Diligence	25.16	3.38	13	32	.69
CO3	Ambition	24.61	3.58	13	32	.80
CO4	Self-Discipline	23.14	4.13	10	32	.81
CO5	Decision Making	21.76	3.48	10	32	.69
ES1	Emotions	22.56	4.52	9	32	.83
ES2	Temper	23.67	4.29	10	32	.78
ES3	Confidence	22.89	4.46	9	32	.81
ES4	Self-Control	21.49	3.53	10	31	.65
ES5	Stress	22.65	3.40	8	32	.78
EX1	Social Need	23.45	4.25	8	32	.83
EX2	Social Image	20.30	5.02	8	32	.87
EX3	Pace of Life	21.89	4.07	11	32	.78
EX4	Excitement-Seeking	19.26	4.15	9	32	.76
EX5	Cheerfulness	24.25	3.95	11	32	.86
OP1	Imagination	20.87	4.25	10	32	.77
OP2	Aesthetics	19.89	5.33	8	32	.85
OP3	Emotional Sensitivity	22.41	2.76	12	30	.51
OP4	Experiences	22.40	3.90	12	32	.77
OP5	Mindset	19.67	4.99	8	32	.84

Table A4

Mean values, standard deviations, minimum and maximum values, and reliability (Cronbach's alpha) in the scales and subscales for females (n = 366)

Scale/ subscale		M	SD	Min	Max	Alpha
AG	Social Style	123.56	10.83	88	152	.83
CO	Conscientiousness	120.87	12.81	79	155	.89
ES	Emotional Stability	112.10	14.98	56	150	.90
EX	Extraversion	109.71	15.54	52	152	.92
OP	Openness	106.58	14.02	70	148	.88
AG1	Trust	24.93	3.64	12	32	.72
AG2	Communication	23.07	3.19	11	31	.60
AG3	Altruism	25.56	3.05	11	32	.67
AG4	Compassion	24.47	3.22	11	32	.64
AG5	Affection	25.53	3.29	14	32	.68
CO1	Intensity	25.09	3.35	14	32	.76
CO2	Diligence	25.61	3.16	14	32	.66
CO3	Ambition	24.75	3.41	14	32	.77
CO4	Self-Discipline	23.66	4.00	10	32	.80
CO5	Decision Making	21.75	3.49	12	32	.68
ES1	Emotions	22.10	4.54	9	32	.83
ES2	Temper	23.47	4.44	10	32	.79
ES3	Confidence	22.54	4.50	9	32	.81
ES4	Self-Control	21.53	3.65	10	31	.66
ES5	Stress	22.45	3.83	8	32	.76
EX1	Social Need	23.81	4.33	8	32	.84
EX2	Social Image	20.25	5.04	8	32	.86
EX3	Pace of Life	22.23	4.08	11	32	.78
EX4	Excitement-Seeking	18.74	4.22	9	31	.78
EX5	Cheerfulness	24.68	3.81	12	32	.85
OP1	Imagination	20.84	4.31	10	32	.78
OP2	Aesthetics	20.75	5.15	9	32	.83
OP3	Emotional Sensitivity	22.88	2.76	14	30	.52
OP4	Experiences	22.45	4.00	12	32	.79
OP5	Mindset	19.66	4.74	8	32	.83

Table A5

Mean values, standard deviations, minimum and maximum values, and reliability (Cronbach's alpha) in the scales and subscales for males (n = 203)

Scale/	subscale	M	SD	Min	Max	Alpha
AG	Social Style	117.77	11.22	82	148	.84
CO	Conscientiousness	117.25	14.53	76	155	.92
ES	Emotional Stability	115.38	15.12	73	148	.91
EX	Extraversion	108.14	16.39	62	149	.93
OP	Openness	102.83	14.66	70	143	.89
AG1	Trust	24.56	3.64	13	32	.73
AG2	Communication	21.81	3.50	12	31	.68
AG3	Altruism	24.36	2.85	16	31	.56
AG4	Compassion	23.06	3.31	15	30	.65
AG5	Affection	23.99	3.31	14	32	.69
CO1	Intensity	24.58	3.86	11	32	.81
CO2	Diligence	24.34	3.61	13	31	.72
CO3	Ambition	24.36	3.88	13	32	.83
CO4	Self-Discipline	22.22	4.20	10	32	.81
CO5	Decision Making	21.76	3.47	10	32	.71
ES1	Emotions	23.40	4.37	10	32	.82
ES2	Temper	24.04	4.00	12	32	.74
ES3	Confidence	23.53	4.32	11	32	.79
ES4	Self-Control	21.41	3.31	12	29	.63
ES5	Stress	23.00	4.19	10	32	.81
EX1	Social Need	22.78	4.03	9	32	.81
EX2	Social Image	20.39	4.99	9	32	.88
EX3	Pace of Life	21.29	3.98	12	31	.78
EX4	Excitement-Seeking	20.20	3.84	10	32	.71
EX5	Cheerfulness	23.48	4.08	11	32	.86
OP1	Imagination	20.94	4.15	12	32	.76
OP2	Aesthetics	18.34	5.32	8	32	.86
OP3	Emotional Sensitivity	21.57	2.55	12	28	.41
OP4	Experiences	22.30	3.71	12	32	.75
OP5	Mindset	19.68	5.43	8	32	.87

Table A6

The correlation matrix for the norm group (n = 569), the group of managers (n = 297) and the validation sample of managers (n = 73)

Scale	AG	CO	ES	EX	OP	AG1	AG2	AG3	AG4	AG5	CO1	CO2	CO3	CO4	CO5
AG	1.00	.36	.48	.50	.36	.75	.64	.70	.72	.79	.31	.28	.32	.27	.17
CO	.31	1.00	.54	.43	.16	.24	.33	.18	.19	.36	.78	.81	.74	.82	.52
ES	.34	.42	1.00	.41	.16	.54	.34	.21	.17	.41	.55	.40	.35	.43	.25
EX	.38	.32	.32	1.00	.49	.34	.26	.33	.22	.62	.51	.27	.52	.32	-.03
OP	.24	.18	-.01	.44	1.00	.23	.13	.23	.33	.38	.21	-.00	.28	.07	.07
AG1	.69	.16	.47	.39	.13	1.00	.35	.40	.39	.53	.25	.18	.22	.12	.14
AG2	.55	.27	.26	.02	-.00	.25	1.00	.23	.37	.37	.24	.24	.26	.28	.18
AG3	.70	.26	.13	.29	.23	.28	.15	1.00	.48	.50	.18	.14	.17	.11	.07
AG4	.69	.13	.01	.03	.12	.27	.24	.50	1.00	.42	.09	.13	.14	.12	.23
AG5	.75	.22	.27	.54	.32	.46	.20	.51	.36	1.00	.32	.30	.35	.32	.02
CO1	.29	.80	.41	.49	.24	.22	.15	.26	.11	.23	1.00	.49	.69	.53	.22
CO2	.26	.79	.28	.10	-.01	.10	.29	.21	.14	.13	.48	1.00	.50	.67	.30
CO3	.26	.75	.25	.51	.37	.19	.12	.22	.08	.27	.69	.42	1.00	.45	.15
CO4	.25	.78	.35	.19	.03	.07	.30	.19	.14	.16	.52	.63	.40	1.00	.27
CO5	.08	.61	.25	-.11	.05	.01	.15	.07	.02	.03	.29	.45	.30	.29	1.00
ES1	.32	.25	.87	.34	-.03	.46	.26	.08	-.01	.25	.29	.17	.14	.25	.08
ES2	.33	.25	.74	.05	-.04	.36	.23	.19	.15	.17	.22	.19	.08	.20	.24
ES3	.31	.28	.76	.52	.16	.43	.18	.07	-.04	.39	.33	.13	.28	.24	.06
ES4	.05	.27	.51	-.30	-.27	.15	.19	-.10	-.04	-.07	.08	.28	-.02	.19	.48
ES5	.20	.47	.70	.46	.08	.25	.09	.20	-.04	.17	.55	.29	.39	.39	.10
EX1	.46	.20	.28	.77	.29	.47	.04	.30	.12	.61	.29	.09	.32	.11	-.06
EX2	.20	.17	.24	.74	.37	.25	.05	.08	-.08	.36	.32	-.01	.37	.07	-.10
EX3	.32	.46	.29	.80	.30	.28	.08	.27	.07	.37	.62	.23	.56	.34	-.05
EX4	.02	.10	.11	.69	.39	.12	-.17	.07	-.13	.19	.27	-.10	.34	.00	-.15
EX5	.42	.27	.26	.69	.28	.34	.07	.37	.17	.46	.34	.20	.32	.18	-.05
OP1	-.02	-.09	-.29	.25	.69	-.09	-.21	.14	.01	.10	.04	-.22	.18	-.19	-.13
OP2	.08	.16	-.11	.15	.73	-.04	.02	.09	.08	.15	.13	.08	.22	.07	.12
OP3	.41	.07	.04	.27	.47	.26	.20	.25	.27	.41	.09	.04	.12	.05	-.06
OP4	.32	.28	.30	.59	.61	.27	.07	.27	.12	.33	.38	.11	.38	.19	-.02
OP5	.13	.16	.07	.29	.77	.14	-.02	.11	.02	.20	.17	-.05	.31	-.01	.17

Below the diagonal: norm group (n = 569)

Above the diagonal: group of managers (n = 297)

Table A6 cont

ES1	ES2	ES3	ES4	ES5	EX1	EX2	EX3	EX4	EX5	OP1	OP2	OP3	OP4	OP5	Scale
.42	.45	.36	.19	.31	.58	.17	.34	.11	.53	-.06	.26	.55	.37	.17	**AG**
.40	.32	.44	.37	.49	.33	.26	.43	.12	.38	-.24	.20	.16	.33	.11	**CO**
.88	.76	.74	.57	.76	.35	.21	.33	.09	.46	-.26	.14	.26	.40	.07	**ES**
.37	.16	.59	-.13	.49	.76	.68	.78	.61	.72	.12	.37	.44	.58	.18	**EX**
.14	.13	.27	-.16	.17	.32	.33	.36	.33	.41	.63	.77	.41	.54	.75	**OP**
.53	.50	.39	.26	.32	.41	.08	.24	.08	.38	-.06	.13	.41	.25	.12	**AG1**
.29	.25	.30	.20	.23	.26	.17	.19	.04	.26	-.16	.10	.30	.19	.08	**AG2**
.17	.28	.10	-.00	.20	.43	.04	.23	.08	.36	-.01	.16	.38	.28	.06	**AG3**
.12	.27	.05	.13	.03	.28	-.03	.14	.03	.33	.05	.26	.36	.19	.24	**AG4**
.37	.33	.41	.08	.30	.69	.33	.40	.16	.57	-.01	.30	.54	.42	.13	**AG5**
.43	.32	.52	.22	.55	.34	.35	.56	.23	.37	-.15	.19	.11	.43	.14	**CO1**
.30	.22	.30	.34	.37	.23	.18	.23	.05	.27	-.30	.15	.13	.13	-.07	**CO2**
.25	.17	.37	.10	.42	.32	.40	.55	.21	.40	-.04	.24	.18	.39	.16	**CO3**
.36	.25	.35	.26	.41	.27	.18	.31	.11	.28	-.28	.13	.18	.27	.01	**CO4**
.12	.25	.09	.46	.06	.07	-.14	-.03	-.13	.11	-.08	.06	-.03	.04	.19	**CO5**
1.00	.60	.67	.35	.62	.32	.20	.24	.12	.42	-.23	.15	.28	.34	.03	**ES1**
.56	1.00	.30	.45	.43	.21	-.14	.11	-.01	.40	-.13	.09	.19	.28	.06	**ES2**
.69	.34	1.00	.20	.55	.43	.55	.43	.24	.44	-.08	.20	.29	.42	.13	**ES3**
.29	.40	.16	1.00	.22	.02	-.14	-.08	-.30	-.01	-.39	-.05	-.06	.00	-.02	**ES4**
.54	.35	.49	.13	1.00	.29	.31	.51	.24	.42	-.20	.14	.23	.41	.06	**ES5**
.27	.12	.43	-.13	.27	1.00	.37	.49	.31	.49	-.00	.27	.38	.44	.06	**EX1**
.24	-.08	.54	-.24	.34	.43	1.00	.45	.23	.33	.09	.26	.30	.27	.19	**EX2**
.27	.04	.36	-.20	.52	.51	.49	1.00	.43	.49	.05	.26	.29	.50	.15	**EX3**
.16	-.02	.25	-.35	.28	.43	.39	.47	1.00	.23	.23	.16	.14	.46	.12	**EX4**
.31	.17	.29	-.19	.30	.51	.32	.50	.29	1.00	.09	.35	.45	.41	.13	**EX5**
-.27	-.24	-.11	-.39	-.09	.11	.19	.14	.29	.20	1.00	.31	-.01	.14	.45	**OP1**
-.12	-.09	-.01	-.12	-.05	.10	.14	.11	.12	.07	.39	1.00	.23	.26	.41	**OP2**
.08	.03	.15	-.14	-.01	.25	.20	.18	.10	.25	.20	.22	1.00	.28	.11	**OP3**
.27	.16	.36	-.15	.38	.41	.36	.48	.52	.40	.27	.21	.24	1.00	.21	**OP4**
.01	.05	.18	-.11	.07	.16	.34	.13	.28	.11	.41	.42	.21	.37	1.00	**OP5**

Table A6 cont

Scale	AG	CO	ES	EX	OP	AG1	AG2	AG3	AG4	AG5	CO1	CO2	CO3	CO4	CO5
AG	1.00														
CO	.49	1.00													
ES	.43	.46	1.00												
EX	.23	.22	.22	1.00											
OP	.31	.04	.10	.48	1.00										
AG1	.68	.25	.41	.13	.18	1.00									
AG2	.66	.46	.42	.02	.04	.36	1.00								
AG3	.72	.44	.27	.22	.26	.36	.25	1.00							
AG4	.64	.22	.09	.10	.23	.19	.27	.45	1.00						
AG5	.70	.28	.25	.34	.36	.44	.23	.46	.27	1.00					
CO1	.26	.74	.43	.48	.01	.12	.33	.30	.02	.12	1.00				
CO2	.33	.80	.21	-.03	-.16	.19	.37	.24	.16	.13	.47	1.00			
CO3	.26	.75	.35	.47	.18	.07	.27	.36	.06	.14	.77	.43	1.00		
CO4	.45	.77	.44	.12	-.04	.30	.46	.36	.16	.25	.47	.59	.45	1.00	
CO5	.39	.49	.21	-.20	.13	.19	.19	.27	.33	.32	.02	.36	.06	.18	1.00
ES1	.27	.36	.85	.26	.08	.32	.33	.13	-.04	.15	.37	.20	.30	.40	.02
ES2	.40	.35	.77	.02	.06	.34	.28	.31	.20	.24	.25	.16	.20	.28	.35
ES3	.36	.34	.76	.44	.22	.40	.34	.12	.08	.27	.38	.16	.31	.34	.03
ES4	.30	.29	.54	-.32	-.12	.15	.37	.17	.15	.16	.07	.23	-.04	.18	.54
ES5	.21	.33	.69	.37	.09	.25	.21	.23	-.06	.10	.47	.02	.47	.37	-.13
EX1	.54	.09	.21	.52	.36	.36	.17	.31	.33	.68	.06	-.03	.06	.04	.17
EX2	-.10	-.01	.01	.65	.28	-.15	-.07	-.07	-.16	.12	.29	-.18	.33	-.11	-.29
EX3	.12	.33	.21	.72	.28	.02	.10	.22	.05	.02	.58	.09	.54	.22	-.19
EX4	-.07	.04	.03	.70	.42	-.09	-.13	.08	-.05	.01	.26	-.10	.26	.03	-.26
EX5	.27	.30	.28	.75	.26	.29	-.02	.22	.17	.27	.45	.13	.41	.24	-.12
OP1	-.06	-.38	-.37	.28	.61	-.08	-.27	.04	.12	.05	-.26	-.48	-.08	-.33	-.20
OP2	.30	-.02	.13	.27	.77	.18	.02	.19	.35	.31	-.11	-.13	.02	-.11	.23
OP3	.29	.16	.25	.20	.39	.32	.11	.16	-.01	.41	.12	.10	.02	.11	.20
OP4	.22	.30	.34	.58	.52	.13	.18	.22	.09	.11	.32	.12	.45	.25	-.04
OP5	.24	.11	.03	.24	.82	.10	.10	.22	.09	.31	.04	-.05	.18	.01	.20

Validation sample (n = 73)

Table A6 cont

ES1	ES2	ES3	ES4	ES5	EX1	EX2	EX3	EX4	EX5	OP1	OP2	OP3	OP4	OP5	Scale
															AG
															CO
															ES
															EX
															OP
															AG1
															AG2
															AG3
															AG4
															AG5
															CO1
															CO2
															CO3
															CO4
															CO5
1.00															ES1
.51	1.00														ES2
.68	.50	1.00													ES3
.29	.41	.18	1.00												ES4
.55	.33	.41	.14	1.00											ES5
.11	.18	.37	.05	.05	1.00										EX1
.05	-.26	.32	-.26	.19	.23	1.00									EX2
.23	.03	.24	-.26	.45	.01	.36	1.00								EX3
.10	-.04	.14	-.37	.24	.08	.30	.60	1.00							EX4
.37	.16	.37	-.24	.32	.36	.24	.51	.43	1.00						EX5
-.33	-.30	-.14	-.39	-.20	.12	.36	.09	.28	.09	1.00					OP1
.05	.22	.17	-.02	.05	.34	-.05	.15	.25	.23	.39	1.00				OP2
.29	.18	.27	.10	.08	.35	.11	.01	-.07	.23	-.05	.25	1.00			OP3
.38	.15	.32	-.10	.43	.15	.28	.55	.59	.38	.12	.17	.05	1.00		OP4
-.01	-.05	.14	.04	-.01	.24	.27	.09	.22	-.02	.40	.48	.31	.34	1.00	OP5

Validation sample (n = 73)

Table A7

Factor loadings from the CFA (n = 569)

AG		CO		ES		EX		OP	
AG1	.58	CO1	.88	ES1	.89	EX1	.71	OP1	.56
AG2	.27	CO2	.59	ES2	.59	EX2	.61	OP2	.51
AG3	.61	CO3	.78	ES3	.76	EX3	.77	OP3	.58
AG4	.45	CO4	.61	ES4	.34	EX4	.54	OP4	.59

Note. Chi Square = 2924.253, (df = 265); RMSEA = .133; NFI = .578; TLI = .546; CFI = .599

Table A8

Correlations between theoretical constructs (n = 569)

	1	2	3	4	5
Agreeableness	1.00				
Conscientiousness	.37	1.00			
Emotional Stability	.41	.44	1.00		
Extraversion	.67	.61	.49	1.00	
Openness	.50	.36	.07	.62	1.00

Note. All correlations are significant p<.05

Table A9

Correlations between scales and age, gender and educational level (n = 569)

	Scale	Gender	Age	Education
1.	AG	-.25*	.04	.14*
2.	CO	-.13*	-.06	.13*
3.	ES	.11*	.12*	.10*
4.	EX	-.05	-.29*	.15*
5.	OP	-.13*	-.35*	.25*

Note. *p<.05

Table A10

Distribution by age, gender, educational level, ares of responsibility and number of subordinates in the group of managers (n = 297)

Gender	
Female	56%
Male	44%
Educational level	
Elementary school	3%
Elementary school and 2 years of upper secondary school	7%
Elementary school and 3-4 years of upper secondary school	13%
Less than three years of tertiary education	18%
3 or more years of tertiary education	57%
Post-graduate studies	2%
Areas of Responsibility	
Budget	83%
Results	84%
Operations	88%
Number of subordinates	
1–10	47%
11–49	44%
50–99	4%
More than 100	5%

Table A11

Mean values, standard deviations, minimum and maximum values and reliability (Cronbach's alpha) in the scales and subscales for the group of managers (n = 297)

Scale/ subscale		M	SD	Min	Max	Alpha
AG	Social Style	121.77	11.95	66	156	.88
CO	Conscientiousness	125.35	12.17	76	152	.90
ES	Emotional Stability	124.65	13.80	67	155	.91
EX	Extraversion	119.28	13.24	71	156	.91
OP	Openness	111.48	12.49	66	146	.87
AG1	Trust	26.00	3.58	12	32	.81
AG2	Communication	23.08	3.36	11	31	.69
AG3	Altruism	24.51	3.00	17	32	.68
AG4	Compassion	22.69	2.94	11	31	.59
AG5	Affection	25.49	3.60	14	32	.78
CO1	Intensity	27.17	2.87	15	32	.73
CO2	Diligence	25.36	3.30	13	32	.73
CO3	Ambition	27.47	3.15	16	32	.81
CO4	Self-Discipline	23.44	3.95	8	32	.83
CO5	Decision Making	21.91	3.22	12	30	.73
ES1	Emotions	25.69	3.97	10	32	.85
ES2	Temper	25.57	3.97	11	32	.81
ES3	Confidence	26.11	3.80	15	32	.80
ES4	Self-Control	21.91	3.20	13	30	.65
ES5	Stress	25.36	3.49	13	32	.80
EX1	Social Need	24.18	4.08	12	32	.86
EX2	Social Image	24.46	3.91	10	32	.84
EX3	Pace of Life	24.35	3.18	14	32	.71
EX4	Excitement-Seeking	20.96	3.54	10	31	.71
EX5	Cheerfulness	25.34	3.93	14	32	.88
OP1	Imagination	20.66	3.87	11	31	.77
OP2	Aesthetics	21.05	5.11	9	32	.87
OP3	Emotional Sensitivity	22.50	2.52	15	29	.51
OP4	Experiences	25.27	3.26	15	32	.78
OP5	Mindset	22.00	4.47	12	32	.83

Table A12

C-score mean values and standard deviations for the group of managers (n = 297)

Scale/	subscale	M	SD
AG	Social Style	5.05	2.11
CO	Conscientiousness	5.85	1.80
ES	Emotional Stability	6.51	1.83
EX	Extraversion	6.28	1.67
OP	Openness	5.87	1.74
AG1	Trust	5.59	1.97
AG2	Communication	5.01	2.11
AG3	Altruism	4.31	1.97
AG4	Compassion	3.89	1.83
AG5	Affection	4.98	2.19
CO1	Intensity	6.24	1.71
CO2	Diligence	4.84	2.09
CO3	Ambition	6.60	1.85
CO4	Self-Discipline	4.89	1.97
CO5	Decision Making	5.09	1.84
ES1	Emotions	6.58	1.75
ES2	Temper	5.95	1.79
ES3	Confidence	6.59	1.69
ES4	Self-Control	5.21	1.75
ES5	Stress	7.07	2.48
EX1	Social Need	5.17	1.89
EX2	Social Image	6.67	1.55
EX3	Pace of Life	6.04	1.56
EX4	Excitement-Seeking	6.05	1.68
EX5	Cheerfulness	5.35	2.06
OP1	Imagination	4.92	1.80
OP2	Aesthetics	5.12	1.98
OP3	Emotional Sensitivity	4.72	1.82
OP4	Experiences	6.41	1.63
OP5	Mindset	5.99	1.89

Table A13

Multiple regression analysis: FFM factors and Managerial Performance

Beta Weights	**ß**
AG Agreeableness	-.01
CO Conscientiousness	.19
ES Emotional Stability	-.04
EX Extraversion	.14
OP Openness	-.02
R	.26

Table A14

Validity of the overall managerial performance composite score MPCS (n = 73)

Predictor	r	ρ1	ρ2	CI r	CI ρ1	CI ρ2
MPCS	.31	.41	.54	.09 – .51	.11 – .62	.16 – .86

Note. r = observed correlation; ρ1 = operational validity. Corrected cropped distribution in MPCS (u = .73. based on SD in sample. .35/ SD in the population. .47) ; ρ2 = operational validity. Corrected for measurement error in the criterion (.52) based on meta-analysis by Viswesvaran, Ones & Schmidt (1996), and cropped distribution of MPCS (u = .73). CI r = 95% confidence interval for r; CI ρ1 = 95% confidence interval for ρ1; CI ρ2= 95% confidence interval for r2. The calculations of the confidence intervals for ρ are presented in Hunter & Schmidt (2004; pp. 109-110).

Table A15

The correlation, r and $\rho 2$, between MPCS and performance scores (manager and employee rating) for the validation sample of managers (n = 73)

Scale/	subscale	Manager	Employee	Manager	Employee
		r	r	$\rho 2$	$\rho 2$
AG	Social Style	0.18	0.16	0.32	0.32
CO	Conscientiousness	0.08	0.24	0.14	0.42
ES	Emotional Stability	0.07	0.06	0.14	0.12
EX	Extraversion	0.16	-0.06	0.35	-0.13
OP	Openness	0.25	-0.08	0.45	-0.11
AG1	Trust	0.03	0.17	0.08	0.44
AG2	Communication	0.16	0.09	0.25	0.14
AG3	Altruism	0.19	0.13	0.43	0.17
AG4	Compassion	0.27	0.17	0.53	0.35
AG5	Affection	-0.03	-0.01	-0.06	-0.02
CO1	Intensity	0.12	0.16	0.26	0.36
CO2	Diligence	0.04	0.25	0.09	0.51
CO3	Ambition	0.07	0.10	0.12	0.17
CO4	Self-Discipline	0.13	0.21	0.25	0.39
CO5	Decision Making	-0.06	0.13	-0.09	0.21
ES1	Emotions	0.03	0.06	0.07	0.14
ES2	Temper	0.08	0.20	0.18	0.42
ES3	Confidence	0.11	-0.07	0.33	-0.22
ES4	Self-Control	-0.13	0.02	-0.35	0.06
ES5	Stress	0.15	0.01	0.22	0.01
EX1	Social Need	-0.05	-0.06	-0.09	-0.11
EX2	Social Image	-0.05	-0.27	-0.11	-0.54
EX3	Pace of Life	0.24	0.03	0.47	0.06
EX4	Excitement-Seeking	0.15	-0.16	0.33	0.35
EX5	Cheerfulness	0.25	0.27	0.42	0.45
OP1	Imagination	0.10	-0.13	0.21	-0.26
OP2	Aesthetics	0.22	-0.11	0.38	-0.19
OP3	Emotional Sensitivity	0.08	0.18	0.34	0.58
OP4	Experiences	0.14	0.07	0.24	0.12
OP5	Mindset	0.22	-0.14	0.38	-0.28

Appendix B

Norm update 2019

Table B1. Demographic information for the 2019 norm groups

	Age (years)		Gender (%)		
	Mean	SD	Female	Male	Other
International	33.08	12.01	54	45.4	.6
Swedish	33.25	12.12	53.9	45.4	.7
Norwegian	38.56	10.38	49.9	50	.1
Finnish	26.69	8.49	71.9	27.9	.2
US English	34.31	9.53	28.8	70.2	1

Table B2. Highest level of completed education for the 2019 norm group sample

	Highest level of completed education (%)					
	Elementary school	Middle or junior high school	Less than 3 years of post-secondary education college or university	3 or more years of post-secondary education college or university	PhD	Other
International	5.6	38	15.4	35.8	.7	4.5
Swedish	6.3	38.7	15.3	34.7	.6	4.3
Norwegian	1.1	13.2	12.6	69.4	.8	2.8
Finnish	4.1	46.4	19	21.9	.2	7.6
US English	.9	9	12.3	70.1	2.4	5.3

Table B3. Mean values, standard deviations, reliability (Cronbach's alpha) for scales and subscales and correlations with age and gender for the International norm group (N=205 217) on the scale level

		Mean	SD	alpha	*r* with age	*r* with gender
Scale	AG	128.84	10.90	.86	-.03**	-.21**
	CO	136.32	11.12	.91	-.22**	-.06**
	ES	130.32	12.37	.89	.02**	.09**
	EX	119.18	12.75	.90	-.19**	-.08**
	OP	111.33	12.39	.86	-.19**	-.10**
Subscale	AG1	26.48	3.03	.73		
	AG2	24.57	3.12	.66		
	AG3	27.02	2.84	.70		
	AG4	24.38	3.01	.59		
	AG5	26.38	3.05	.72		
	CO1	28.08	2.69	.76		
	CO2	28.26	2.57	.72		
	CO3	28.79	2.76	.80		
	CO4	26.88	3.28	.79		
	CO5	24.32	2.92	.70		
	ES1	26.15	3.33	.78		
	ES2	27.90	3.16	.78		
	ES3	25.92	3.50	.78		
	ES4	23.68	3.13	.56		
	ES5	26.67	3.08	.77		
	EX1	25.88	3.56	.82		
	EX2	22.39	3.75	.82		
	EX3	24.06	3.17	.70		
	EX4	20.56	3.35	.69		
	EX5	26.29	3.38	.84		
	OP1	19.78	3.53	.72		
	OP2	20.85	4.95	.85		
	OP3	24.24	2.71	.47		
	OP4	24.77	3.24	.75		
	OP5	21.70	4.11	.81		

Note: All correlations are significant at p<.01.

Table B4. Difference in sample distribution between the Swedish development sample (N=569) and the 2019 International norm group sample (N=205 217) expressed as Cohen's *d*

	Cohen's *d*	95% confidence interval
AG	.67	.59–.75
CO	1.50	1.42–1.58
ES	1.37	1.29–1.46
EX	.78	.70–.87
OP	.49	.41–.57

Table B5. Mean values, standard deviations, reliability (Cronbach's alpha) for scales and subscales for the norm group of the Swedish language version (N=188 504)

		Mean	SD	alpha
Scale	AG	128.94	10.93	.87
	CO	136.46	11.07	.91
	ES	130.65	12.32	.89
	EX	118.93	12.78	.90
	OP	111.10	12.38	.86
Subscale	AG1	26.46	3.04	.73
	AG2	24.67	3.11	.66
	AG3	27.06	2.85	.71
	AG4	24.38	3.02	.60
	AG5	26.37	3.05	.73
	CO1	28.10	2.68	.75
	CO2	28.26	2.57	.72
	CO3	28.83	2.75	.80
	CO4	26.97	3.26	.79
	CO5	24.30	2.90	.70
	ES1	26.20	3.34	.78
	ES2	27.99	3.12	.78
	ES3	25.92	3.51	.78
	ES4	23.73	3.12	.56
	ES5	26.82	3.05	.78
	EX1	25.80	3.56	.81
	EX2	22.35	3.76	.82
	EX3	23.94	3.16	.70
	EX4	20.54	3.35	.68
	EX5	26.30	3.39	.84
	OP1	19.73	3.54	.72
	OP2	20.68	4.93	.84
	OP3	24.21	2.72	.48
	OP4	24.78	3.25	.75
	OP5	21.70	4.11	.81

Table B6. Mean values, standard deviations, reliability (Cronbach's alpha) for scales and subscales for the norm group of the Norwegian language version (N=4 821)

		Mean	SD	alpha
Scale	AG	125.70	10.38	.87
	CO	130.31	10.80	.90
	ES	126.32	11.63	.89
	EX	118.14	11.58	.90
	OP	108.32	11.54	.86
Subscale	AG1	26.05	2.87	.74
	AG2	23.99	3.00	.65
	AG3	26.06	2.79	.72
	AG4	23.40	2.76	.60
	AG5	26.21	2.98	.75
	CO1	27.17	2.57	.73
	CO2	27.28	2.58	.69
	CO3	28.41	2.84	.82
	CO4	24.24	3.37	.82
	CO5	23.21	2.75	.70
	ES1	25.44	3.12	.77
	ES2	26.63	3.37	.79
	ES3	25.50	3.36	.82
	ES4	23.12	3.02	.58
	ES5	25.63	2.84	.75
	EX1	25.65	3.23	.80
	EX2	22.43	3.32	.81
	EX3	24.45	2.94	.71
	EX4	20.08	3.07	.68
	EX5	25.53	3.20	.82
	OP1	19.60	3.56	.75
	OP2	20.09	4.49	.83
	OP3	23.78	2.49	.47
	OP4	23.60	3.05	.73
	OP5	21.25	3.75	.80

Table B7. Mean values, standard deviations, reliability (Cronbach's alpha) for scales and subscales for the norm group of the Finnish language version (N=8 883)

		Mean	SD	alpha
Scale	AG	127.80	9.71	.84
	CO	134.48	10.78	.91
	ES	124.90	11.71	.88
	EX	122.87	11.52	.90
	OP	113.71	11.52	.85
Subscale	AG1	26.77	2.75	.70
	AG2	23.29	2.76	.59
	AG3	26.56	2.53	.66
	AG4	24.56	2.60	.51
	AG5	26.62	2.83	.68
	CO1	27.52	2.72	.76
	CO2	28.50	2.33	.68
	CO3	27.67	2.70	.76
	CO4	26.60	3.04	.79
	CO5	24.19	2.87	.70
	ES1	25.58	3.22	.75
	ES2	26.67	3.29	.76
	ES3	25.58	3.18	.77
	ES4	23.10	3.06	.53
	ES5	23.98	2.49	.53
	EX1	27.47	3.19	.82
	EX2	22.34	3.54	.80
	EX3	25.84	2.83	.71
	EX4	20.51	3.28	.70
	EX5	26.71	3.04	.81
	OP1	20.86	3.22	.68
	OP2	22.74	4.54	.84
	OP3	24.87	2.53	.44
	OP4	24.56	3.00	.68
	OP5	20.69	3.95	.81

Table B8. Mean values, standard deviations, reliability (Cronbach's alpha) for scales and subscales for the norm group of the US English language version (N=3 919)

		Mean	SD	alpha
Scale	AG	130.45	11.88	.84
	CO	142.21	10.53	.87
	ES	131.84	13.39	.87
	EX	124.43	12.49	.86
	OP	120.73	11.04	.80
Subscale	AG1	27.11	3.46	.72
	AG2	23.70	3.86	.67
	AG3	27.66	2.78	.60
	AG4	25.27	3.19	.47
	AG5	26.71	3.28	.66
	CO1	29.68	2.27	.63
	CO2	28.69	2.69	.63
	CO3	30.14	2.30	.74
	CO4	26.82	3.38	.63
	CO5	26.88	3.02	.62
	ES1	25.86	3.61	.67
	ES2	27.74	3.48	.67
	ES3	27.17	3.63	.75
	ES4	23.48	3.84	.57
	ES5	27.59	3.03	.58
	EX1	26.60	3.80	.80
	EX2	24.22	3.69	.71
	EX3	25.08	2.81	.47
	EX4	22.44	3.28	.57
	EX5	26.10	3.68	.77
	OP1	20.04	3.26	.62
	OP2	25.31	4.45	.79
	OP3	24.97	2.80	.36
	OP4	26.19	3.04	.59
	OP5	24.23	3.70	.69

Appendix C

MAP Essence

Table C1. Demographic information, development sample

Age		Gender (%)		
Mean	**SD**	**Female**	**Male**	**Missing**
37.34	10.93	48	52	0

Table C2. Mean values, standard deviations and reliability (Cronbach alpha) (N=25 733)

		M	**SD**	**alpha**
Scale	AG	47.94	4.52	.70
	CO	49.76	4.66	.77
	ES	49.43	5.60	.83
	EX	45.59	4.93	.77
	OP	43.60	5.13	.73

Table C3. Intercorrelations between scales for Essence and MAP (N=25 733)

		AG	**CO**	**ES**	**EX**	**OP**
Scale	AG	**.93**	.41	.38	.30	.28
	CO	.57	**.90**	.55	.33	.21
	ES	.45	.54	**.95**	.32	.04
	EX	.37	.39	.37	**.92**	.35
	OP	.32	.26	.12	.46	**.93**

Note: All correlations are siginificant at p<.01.

Table C4. Mean values, standard deviations, reliability (Cronbach's alpha) and correlations with age and gender for the International norm group (N=205 217)

		Mean	SD	alpha	*r* with age	*r* with gender
Scale	AG	48.19	4.41	.70	.00	-.20**
	CO	51.91	4.43	.81	-.23**	-.07**
	ES	49.53	5.21	.82	.00	.08**
	EX	45.64	4.99	.78	-.15**	-.09**
	OP	43.78	5.21	.74	-.16**	-.08**

Note: All correlations are significant at p<.01.

Table C5. Mean values, standard deviations, reliability (Cronbach's alpha) for the norm group of the Swedish language version (N=188 504)

		Mean	SD	alpha
Scale	AG	48.21	4.42	.70
	CO	51.98	4.41	.81
	ES	49.64	5.20	.82
	EX	45.63	5.01	.79
	OP	43.61	5.17	.74

Table C6. Mean values, standard deviations, reliability (Cronbach's alpha) for the norm group of the Norwegian language version (N=4 821)

		Mean	SD	alpha
Scale	AG	47.26	4.31	.73
	CO	49.08	4.44	.79
	ES	48.20	4.83	.80
	EX	44.48	4.62	.78
	OP	42.49	4.86	.74

Table C7. Mean values, standard deviations, reliability (Cronbach's alpha) for the norm group of the Finnish language version (N=8 883)

		Mean	SD	alpha
Scale	AG	47.46	3.80	.63
	CO	51.40	4.37	.81
	ES	47.78	5.14	.80
	EX	46.30	4.53	.76
	OP	45.81	4.87	.74

Table C8. Mean values, standard deviations, reliability (Cronbach's alpha) for the norm group of the US English language version (N=3 919)

		Mean	SD	alpha
Scale	AG	49.93	4.74	.67
	CO	53.71	4.27	.73
	ES	49.55	5.69	.76
	EX	46.76	5.05	.68
	OP	48.64	4.56	.65

Appendix D

Report Overview

MAP	Report name	Available languages*
A comprehensive Five Factor Model (FFM) personality test. Provides results on five overall scales and 25 subscales.	**On-screen feedback report** Intended recipient: Respondent Type: On-screen feedback (Also downloadable as a PDF) FFM and facet level.	Arabic Chinese (Simplified) Danish English Estonian Finnish Flemish French German Italian Latvian Lithuanian Norwegian Spanish Swedish
	Descriptive report Intended recipient: Respondent Type: PDF report Thorough report intended as a basis for one on one feedback. FFM and facet level.	English Finnish Norwegian Swedish
	Interpretive report Intended recipient: Test administrator Type: PDF report Descriptive and development report for individuals. FFM and facet level.	English Finnish Norwegian Swedish
	Managerial report Intended recipient: Test administrator Type: PDF report Descriptive and development report focused on individual Leadership skills. FFM and facet level.	English Finnish Norwegian Swedish
*Additional languages added continuously, for an updated list of available languages contact your sales representative.		

MAP Essence	Report name	Available languages*
A Five Factor Model (FFM) personality test. Provides results on five overall scales.	**On-screen feedback report** Intended recipient: Respondent Type: On-screen feedback (Also downloadable as a PDF) FFM level only.	Arabic Chinese (Simplified) Danish English Estonian Finnish Flemish French German Italian Latvian Lithuanian Norwegian Spanish Swedish
	Descriptive report Intended recipient: Respondent Type: PDF report Thorough report intended as a basis for one on one feedback. FFM level only.	Arabic Chinese (Simplified) Danish English Finnish French German Italian Lithuanian Norwegian Spanish Swedish
	Interpretive report Intended recipient: Test administrator Type: PDF report Descriptive and development report for individuals. FFM level only.	Arabic Chinese (Simplified) Danish English Finnish French German Italian Lithuanian Norwegian Spanish Swedish
*Additional languages added continuously, for an updated list of available languages contact your sales representative.		

Appendix E

Development and psychometric properties of MAP – Norwegian language version

This section describes the translation and adaptation process from the original Swedish language version (please see Chapter 5) into the Norwegian language version. This work was carried out between 2011-2012.

Translation and adaptation

In the first step, instructions and items were translated from Swedish into Norwegian by a professional translator, a native speaker of both Swedish and Norwegian. In the second step, the items were reviewed by an external subject matter expert to primarily ensure equivalent psychological content and meaning between the two language versions. Suggested changes were put forward and discussed with Assessio's Norwegian test development team which included experts in psychology, linguistics, and psychometricians. Necessary revisions of instructions and items were made, and in the final step, all content was reviewed again to ensure high linguistic quality.

After finalizing the translation and adaptation of the instructions and items, the content was implemented in Assessio's web-based platform and the new version was administered to a pilot group of 10 individuals with Norwegian as their mother tongue. No changes were made after this and the Norwegian language version was taken into use for further data collection.

Descriptives and reliability

When data from n=380 had been collected a psychometric evaluation was conducted. The sample of n=380 consisted of employees in the grocery store sector. In Table E1 the sample is described regarding age, gender and educational level. Mean values, standard deviations, minimum and maximum values, as well as reliability (Cronbach's alpha and test retest) for scales and subscales respectively, are presented in Table E2. Table E3 presents the correlations between the scales and age, gender and educational level, and in table E4 the correlations between all scales and subscales for the Norwegian grocery store sample are presented.

Table E1

The distribution of the Norwegian normgroup (n=380) across age, gender and educational level

Age group	**Normgroup**
18-30 years	57%
31-40 years	20%
41-50 years	14%
51-65 years	9%
Gender	
Female	69%
Male	31%
Educational level	
Elementary school	11%
Elementary school and 2 years of upper secondary school	56%
Elementary school and 3-4 years of upper secondary school	14%
Less than 3 years of tertiary education	16%
3 or more years of tertiary education	3%

E2

Descriptive statistics (means, standarddeviations, minimum, maximum) and reliability (Cronbach's alpha and test-retest (n=147)) for scales and subscales for the Norwegian grocery store sample (n = 380).

Scale / subscale	M	SD	Min	Max	Alpha	r_{tt}
AG	120.40	9.95	91	147	.82	.86
CO	122.75	12.19	86	151	.90	.88
ES	113.11	13.58	75	147	.90	.91
EX	114.16	13.10	68	147	.90	.91
OP	104.70	11.63	70	142	.84	.89
AG1	23.82	3.23	13	31	.71	.81
AG2	23.06	2.99	14	30	.86	.78
AG3	25.58	2.77	18	32	.66	.79
AG4	22.91	2.85	14	29	.59	.78
AG5	25.03	3.01	17	32	.65	.76
CO1	24.95	3.07	15	31	.79	.86
CO2	26.23	2.81	18	32	.65	.85
CO3	26.13	3.60	17	32	.85	.78
CO4	23.26	3.47	13	32	.78	.78
CO5	22.18	2.85	14	30	.63	.80
ES1	22.49	3.99	10	32	.80	.80
ES2	23.25	4.11	12	32	.78	.82
ES3	22.29	4.23	9	32	.82	.86
ES4	21.22	3.09	12	29	.62	.92
ES5	23.84	3.15	16	32	.71	.71
EX1	24.69	3.63	13	32	.80	.84
EX2	20.82	4.31	9	31	.85	.88
EX3	23.20	3.19	13	31	.70	.83
EX4	20.17	3.47	11	30	.70	.81
EX5	25.29	3.54	13	32	.84	.83
OP1	20.81	3.99	10	31	.73	.84
OP2	18.64	4.32	9	31	.79	.87
OP3	23.56	2.64	16	31	.56	.72
OP4	22.31	3.10	14	31	.66	.73
OP5	19.39	4.27	8	32	.83	.81
Average values					.73	.81

E3

Correlations between scales and age, gender and educational level in the grocery store sample (n = 296)

Scale	Gender	Age	Educational level
AG	-.19*	.17*	-.01*
CO	-.03	-.19*	.12*
ES	.13*	.30*	.08*
EX	-.06	-.12*	.14*
OP	-.09*	-.19*	.13*

Note. *p<.05

E4

Correlation matrix for the Norwegian grocery store sample (n = 380)

Skala	AG	CO	ES	EX	OP	AG1	AG2	AG3	AG4	AG5	CO1	CO2	CO3	CO4	CO5
AG	1.00														
CO	.40	1.00													
ES	.34	.53	1.00												
EX	.35	.35	.32	1.00											
OP	.21	.26	.00	.37	1.00										
AG1	.66	.26	.50	.26	.03	1.00									
AG2	.60	.37	.37	.10	.00	.35	1.00								
AG3	.74	.30	.10	.26	.17	.29	.23	1.00							
AG4	.60	.11	-.05	-.00	.20	.12	.18	.44	1.00						
AG5	.75	.29	.18	.54	.30	.38	.25	.55	.33	1.00					
CO1	.30	.81	.51	.48	.30	.21	.29	.23	-.01	.28	1.00				
CO2	.41	.76	.36	.11	.03	.29	.35	.31	.19	.22	.47	1.00			
CO3	.29	.82	.40	.52	.34	.19	.25	.21	.01	.30	.74	.46	1.00		
CO4	.35	.81	.48	.17	.12	.23	.34	.27	.16	.18	.56	.60	.53	1.00	
CO5	.18	.63	.28	-.00	.18	.06	.18	.14	.10	.12	.32	.44	.36	.38	1.00
ES1	.21	.32	.84	.32	-.03	.46	.23	-.03	-.08	.07	.35	.21	.26	.31	.10
ES2	.34	.34	.75	.11	-.04	.44	.27	.18	.09	.13	.30	.31	.18	.31	.22
ES3	.29	.38	.74	.59	.17	.37	.29	.07	-.15	.34	.45	.15	.44	.29	.09
ES4	.14	.43	.57	-.24	-.13	.20	.27	.02	.02	-.04	.25	.37	.20	.41	.46
ES5	.26	.51	.73	.26	.06	.32	.29	.13	-.05	.13	.50	.32	.39	.50	.21
EX1	.42	.14	.16	.72	.17	.25	.11	.30	.12	.61	.20	.11	.23	.02	-.02
EX2	.13	.29	.27	.75	.31	.12	.07	.03	-.13	.35	.43	.02	.45	.14	.04
EX3	.31	.49	.38	.77	.25	.24	.19	.28	-.01	.34	.62	.24	.60	.31	.08
EX4	-.03	.00	.01	.66	.35	.01	-.15	-.02	-.10	.17	.16	-.20	.20	-.05	-.15
EX5	.45	.33	.32	.71	.27	.35	.17	.37	.13	.47	.35	.24	.40	.20	.04
OP1	-.07	-.11	-.37	.17	.66	-.21	-.18	.00	.07	.09	-.03	-.17	.03	-.20	-.09
OP2	.09	.12	-.04	.16	.76	-.01	-.03	.04	.15	.14	.14	.02	.16	.05	.07
OP3	.51	.23	.15	.18	.43	.29	.27	.39	.34	.44	.14	.21	.17	.19	.20
OP4	.27	.39	.26	.52	.49	.14	.12	.19	.12	.32	.44	.15	.42	.30	.18
OP5	.03	.27	.11	.22	.71	.00	-.04	.05	.02	.09	.29	-.00	.34	.14	.24

ES1	ES2	ES3	ES4	ES5	EX1	EX2	EX3	EX4	EX5	OP1	OP2	OP3	OP4	OP5	**Skala**
															AG
															CO
															ES
															EX
															OP
															AG1
															AG2
															AG3
															AG4
															AG5
															CO1
															CO2
															CO3
															CO4
															CO5
1.00															ES1
.53	1.00														ES2
.62	.32	1.00													ES3
.30	.41	.18	1.00												ES4
.53	.43	.44	.31	1.00											ES5
.14	.08	.37	-.18	.08	1.00										EX1
.25	-.03	.61	-.15	.21	.38	1.00									EX2
.34	.16	.48	-.04	.41	.39	.51	1.00								EX3
.13	-.10	.26	-.38	.04	.32	.38	.44	1.00							EX4
.31	.30	.38	-.12	.21	.52	.32	.50	.27	1.00						EX5
-.35	-.26	-.19	-.35	-.21	.07	.09	.00	.30	.14	1.00					OP1
-.04	-.06	.04	-.05	-.06	.11	.13	.07	.13	.13	.44	1.00				OP2
.07	.11	.15	.08	.13	.24	.10	.09	-.07	.29	.09	.31	1.00			OP3
.25	.12	.29	-.03	.28	.28	.34	.49	.46	.32	.11	.15	.04	1.00		OP4
.07	.03	.16	.02	.12	-.06	.31	.20	.26	.05	.29	.35	.13	.32	1.00	OP5

The Norwegian grocery store sample (n=380) was applied as a norm group in the Assessio Select platform. However, the data collected on the Norwegian language version, described in section 5 (MAP) and section 6 (Essence), is now applied as the norm for the Norwegian language version.

Validity

Validity is an overarching concept and perspective embracing multiple aspects of a psychological assessment. Collecting and outlining evidence for validity is an ongoing process and may never be finalized in an absolute sense. Evidence supporting the validity of MAP regarding content-, construct-, and criterion-related validity is presented in the following sections.

Factor structure

To get an indication of the overall validity for the Norwegian language version a confirmatory factor analysis (Confirmatory Factor Analysis, CFA) in the AMOS 18.0 program was computed. The analysis was conducted with the correlation matrix in table E4 as input. As for the corresponding analysis for the Swedish language version (please see a detailed description in Chapter 5, the section Factor Structure), the theoretical model postulates five factors (scales) with five facets (subscales) under each factor. This model was tested first. The results from the analysis showed a statistically significant difference between model and data (df = 265; χ^2 = 2073.16, p <.001) and the fit indicators (less sensitive to sample size), RMSEA = .13 and CFI = .58, indicated that the measurement model would benefit from some improvements to reach fully acceptable levels. The factor loadings for the model are presented in Table E5 and the correlations between the theoretical constructs are presented in Table E6.

E5

Confirmatory Factor Analysis loadings, Norwegian normgroup (n = 380)

AG		**CO**		**ES**		**EX**		**OP**	
AG1	.46	CO1	.87	ES1	.80	EX1	.60	OP1	.52
AG2	.34	CO2	.58	ES2	.60	EX2	.61	OP2	.61
AG3	.66	CO3	.84	ES3	.73	EX3	.77	OP3	.37
AG4	.40	CO4	.67	ES4	.39	EX4	.46	OP4	.33
AG5	.84	CO5	.42	ES5	.69	EX5	.68	OP5	.44

Note Chi Square = 2073.158, (df = 265); RMSEA = .134; NFI = .548; TLI = .523; CFI = .578

E6

Correlations between theoretical constructs, Norwegian normgroup (n = 380)

	AG	**CO**	**ES**	**EX**	**OP**
AG	1.00				
CO	.42	1.00			
ES	.29	.61	1.00		
EX	.63	.64	.54	1.00	
OP	.35	.40	.01	.47	1.00

Note. All correlations are significant at p<.05

The relationship with job performance

In year 2012, a concurrent criterion-related validity study was conducted using the Norwegian version of MAP. This sample also consisted of employees within the grocery store sector in Norway. In total, 36 store managers assessed 130 employees. The criterion in the study consisted of the employee's level of work performance assessed by each employee's store manager. Each store manager rated the employee's overall work performance on 9 questions (e. g., "solve problems on his/her own ","is competitive" and "remember important things") using a Likert-scale.

When evaluating the properties of the job performance ratings, it became evident that a bias existed between the raters (store managers). It was considered likely that the bias was due to systematic measurement errors as store managers may only compare employees working at a single workplace (store). It was therefore decided to only use ratings from store managers who had rated more than two employees. Further, the ratings used were standardized within each rater. In the analysis, each store manager thus represents a unit and is treated as a separate sample. The scores from each unit was used as the criterion in the analysis.

Validity studies of this type may however underestimate the validity due to reliability deficiencies in the observed performance ratings. Correcting for this unreliability when estimating the criterion validity may thus be of importance. An employee, however, is only rated by one manager and it is consequently not possible to estimate the reliability. That would require multiple raters. Instead, an estimate from a recognized and comprehensive meta-analysis on performance ratings was applied. In this study, the average reliability of job performance ratings conducted by managers/supervisors has been estimated to .52 (Viswesvaran, Ones & Schmidt, 1996).

A concurrent validation may also lead to an underestimation of the validity as organizations rarely select individuals for employment purely by chance. Their selection process has likely rendered some kind of systematic selection. It is therefore highly likely that there is limited variation in the data, and this will lower the validity estimate. This phenomenon, restriction of range, may however be estimated and corrected for when conducting concurrent criterion-related validity studies.

To investigate the restriction of range, the variation (standard deviation, SD) for the five factors in the Swedish standardization sample, SD=2, was compared to the variation in the validation sample. By dividing the SD from the validation sample with the SD from Swedish standardization sample, one can estimate the limitation in range in each factor. The estimates may then be applied to correct the validity estimates in the current study. This correction is important due to the test scores in practice being applied to job seekers and not to individuals already selected and employed within the organization. This type of correction is called "correction for indirect range restriction" as one corrects for both known and unknown factors in observable relationships (correlations) between predictors and the criterion (Schmidt, Shaffer, & Oh, 2008).

For a validation study of this kind, there is no valid rational for having an exploratory (thus, less demanding) approach regarding the outcome. There is massive research outlining the

relationships between the FFM factors and criteria such as job performance. Based on this research (Schmidt, Shaffer, & Oh, 2008, Appendix D), the following hypotheses were formulated:

All FFM factors in MAP should correlate positively with job performance. Based on meta-analytic research (Schmidt, Shaffer, & Oh, 2008, Appendix D) the estimates are hypothesized to be of the following magnitudes: Agreeableness, ρ = .08; Conscientiousness, ρ = .22; Emotional Stability, ρ = .12; Extraversion ρ = .09; Openness ρ = .04).

The results, presented in Table E7, provides support for the criterion validity of MAP regarding job performance. All factors correlate positively with job performance. Comparison of results show that Extraversion, Emotional Stability and Openness in the current study have stronger correlations with job performance ratings compared to the meta-analytic estimates. Agreeableness (in MAP; Social style) on the other hand, show a weaker positive correlation with work performance in the current study compared to the meta-analytic estimates. In line with the hypothesis (and thus meta-analytic results), Conscientiousness is the factor with the strongest correlation with job performance at the workplace.

E7
Results from validationstudy with job performance criteria (N=130)

	Job Performance				
Factor	**α**	**υ**	**r**	**ρ^1**	**ρ^2**
AG	.82	.91	.02	.04	.08
CO	.90	.92	.15	.23	.22
ES	.90	.89	.12	.19	.12
EX	.90	.89	.14	.22	.09
OP	.89	.85	.05	.09	.04

Not. α = Cronbach Alpha for the Norwegian nomrgroup. υ = Restriction of range in validation sample. r = Observed relationship. ρ^1 = operational validity corrected for restriction of range in predicctors and reliability deficiencies in the criteria. ρ^2 = Estimated correlation i population (Schmidt, Shaffer, & Oh, 2008, Appendix D).

Relationships with other assessments

To investigate the construction validity, the five scales were correlated with the results from two other assessments: ServiceFirst™ and MINT™.

ServiceFirst™ is an online-based assessment measuring service mindedness and was developed to be used for selection to professions with customer contact. ServiceFirst™ aim to predict service mindedness by measuring, among other things, responsiveness and customer focus. ServiceFirst™ (Fogli, 2011) is not based on the FFM of personality but is based on work analyses of service professions. The hypothesis, however, is that all five

factors correlate positively with the overall score in ServiceFirst™. The study was conducted on same group of employees in the grocery store sector as described above (see the section Relationship with job performance) and on an additional group of employees from the same organization (N = 181).

The test Measuring INTegrity (MINT™; Sjöberg & Sjöberg, 2007) is a personality-based assessment developed for personnel selection. MINT™ conceptualizes and measures the personality-based construct of integrity with the aim of predicting counterproductive work behavior, CWB. The overall score gained from MINT™, the Integrity score, is based on three of the FFM factors: Emotional Stability, Conscientiousness and Agreeableness. Positive correlations were thus expected between these three MAP scales and the overall Integrity factor in MINT. In the overall Integrity factor however, Emotional Stability has the heaviest weighting (Sjöberg & Sjöberg, 2007, Table 5.2, page 51), therefore it was hypothesized that the strongest correlation would be between these two factors.

The results in Table E8 shows that all hypothesized correlations were significantly different from zero, except for the correlation between Openness and Integrity in MINT™. These results support the hypotheses, with the exception of Extraversion in MAP which unexpectedly correlated positively with MINT™. This is likely due to the partial overlap between external and emotional balance in MAP.

E8

Correlations between MAP scales and Service First™ and MINT™(N=181)

	Measure	
Scale	**Service First™**	**MINT™**
AG	.47*	.35*
CO	.48*	.43*
ES	.51*	.79*
EX	.48*	.38*
OP	.22*	-.09

Note. = p<.05. r = observed correlation.

Altogether, the results contribute to supporting that MAP in general and the Norwegian language version in particular, measures what it is intended to measure, that it measures the FFM in the intended way, and that scores from MAP are related to job performance.

Appendix F

Psychometric properties of MAP – Swedish language version

This section describes the work regarding the Swedish language version of MAP. During 2019 it was decided to conduct several studies to further ensure the validity and reliability of the Swedish language version. This work was completed and finalized in 2020 and it consists of studies regarding both construct- and concurrent validity as well as test-retest reliability.

Relationship with job performance

In 2019, a concurrent criterion-related validity study was conducted using the Swedish language version of MAP. The reason for conducting the study was to investigate the validity measures in conjunction with the norm group update to ensure the quality of the Swedish language version of the assessment.

There is massive research outlining the relationships between the FFM factors and criteria such as job performance. Based on meta-analytic research (Schmidt, Shaffer, & Oh, 2008) regarding the FFM factors, it is hypothesized that all FFM factors in MAP should correlate positively with job performance. The estimates in the general population are suggested to be of the following magnitudes: Agreeableness, ρ = .08; Conscientiousness, ρ = .22; Emotional Stability, ρ = .12; Extraversion ρ = .09; Openness ρ = .04 (Schmidt, Shaffer, & Oh, 2008, Appendix D).

A study to investigate the relationships between the FFM factors of MAP and job performance was conducted in collaboration with an insurance company in Sweden. The employee's level of job performance was assessed by either the employee's manager or a person in a similar position within the company. The employees were evaluated on their overall work performance based on two criteria, the first being *future potential* for the organisation, and the second being daily *contribution* to the organisation. The managers rated the employee's performance using several standardized statements, each with three different rating levels – low, medium and high. The data collected using the assessments of potential and contribution were combined into a single overall estimate that was analysed together with the employee's MAP result.

In total 87 employees were assessed over a six-month period. Table F1 shows descriptive information for the insurance employee sample regarding age, gender and educational level. Table F2 show the results regarding the combined performance rating and the FFM personality factors.

The results provide support for the criterion validity of MAP regarding job performance as all

factors correlate positively with the performance rating. In the sample four of the FFM factors, Conscientiousness, Extraversion, Emotional Stability and Openness show stronger correlations compared with the meta-analytic results of Schmidt, Shaffer, & Oh (2008). However, only the correlations between the criteria and Conscientiousness and Extraversion are statistically significant. In line with the hypothesis, Conscientiousness has the strongest correlation with job performance at the workplace.

Table F1
The distribution of the insurance employee sample (N=87) across age, gender, and educational level

Age group	**Insurance employees**
18-30 years	33%
31-40 years	31%
41-50 years	17%
51-65 years	19%
Gender	
Male	52%
Female	48%
Educational level	
Middle or junior high school	44%
Less than three years of post-secondary education	22%
Three or more years of post-secondary education	25%
Other	2%

Table F2
Results from insurance employee study with job performance criteria (N=86)

	Performance rating				
Factor	**α**	**υ**	**r**	ρ^1	ρ^2
AG	.89	1.02	.07	.07	.08
CO	.92	.90	.32*	.36*	.22
ES	.92	.87	.16	.18	.12
EX	.93	1.02	.28*	.27*	.09
OP	.87	.85	.14	.17	.04

Note. * p<.05. α = Cronbach's alpha. υ = Restriction of range in validation sample. r = Observed relationship. ρ^1 = Operational validity corrected for restriction of range in predictors and reliability deficiencies in the criteria ρ^2 = Estimated correlation in population (Schmidt, Shaffer, & Oh, 2008, Appendix D).

Relationship with other assessments

In 2019, a construct validity study was conducted using the Swedish version of MAP. The purpose of this study was to evaluate the construct validity of the Swedish language version of MAP by comparing the results with the Swedish version of another assessment that measures the FFM. The assessment chosen for this was NEO-PI-3 (McCrae & Costa, 2010), widely regarded as the golden standard of FFM assessments.

The hypothesis is that all factors in MAP should correlate positively with the factors in NEO-PI-3. Neuroticism in NEO-PI-3 has been reversed to make interpretation of the results easier. The construct definition of the Agreeableness factor has a few differences between the two assessments and therefore the correlation is expected to be lower compared with the other FFM factors.

Participants for the validation study were recruited using Asessio's social media channels and web page and an incentive was used for recruitment. For each participant that completed both MAP and NEO-PI-3 Assessio donated 100 SEK to the Swedish Red Cross. A total of 143 participants completed both assessments during a four-week period.

Table F3 shows descriptive information for the validation study sample regarding age, gender and educational level. Table F4 show descriptive statistics for the FFM factors of MAP and table F5 show the results of the relationship between the FFM factors of MAP and NEO-PI-3. For interpretation purposes the scale Neuroticism in NEO-PI-3 has been reversed to reflect the scale Emotional Stability in MAP.

The results indicate a good convergence between the Swedish language version of MAP and the Swedish language version of NEO-PI-3. As expected, Agreeableness show a weaker correlation (.67) compared to the other FFM factors (.85 – .89). The most likely explanation being that the construct definition of Agreeableness differs between the two assessments. In conclusion, the results support the construct validity of the Swedish language version of MAP.

Table F3

The distribution of the validation sample (N=143) across age, gender, and educational level

Age group	**Validation sample**
18-30 years	24%
31-40 years	39%
41-50 years	21%
51-65 years	16%
66+	1%
Gender	
Male	40%
Female	57%
Prefer not to say	3%
Educational level	
Elementary school	1%
Middle or junior high school	16%
Less than three years of post-secondary education	7%
Three or more years of post-secondary education	72%
PHD	2%
Prefer not to say	2%

Table F4

Descriptive statistics of the MAP FFM factors for the sample group (N=143)

Scale	M	SD	Min	Max	Alpha
AG	122.46	12.95	78	150	.90
CO	129.98	12.58	94	157	.91
ES	118.83	15.71	78	149	.93
EX	113.13	11.91	62	155	.94
OP	112.50	13.70	82	142	.87

Table F5

Correlation between MAP scales and NEO-PI-3 scales (N=143)

			NEO-PI-3		
Scale	AG	CO	ES	EX	OP
AG	.67				
CO		.85			
ES			.87		
EX				.89	
OP					.89

Note. All correlations are significant at p<.05; The Neuroticism scale in NEO-PI-3 has been reversed to reflect the scale of Emotional Stability in MAP.

Test-retest reliability

In 2020, a test-retest study was conducted using the Swedish language version of MAP. The reason for conducting the study was to investigate the repeatability of the Swedish language version to further ensure its quality.

Due to the popularity of MAP in the Swedish market the data required for a test-retest study could be conducted using existing data from the database of Assessio's web-platform Ascend.

Several criteria was set up to either exclude or include data based on, minimum and maximum time between the two assessments (between 2 weeks and 2 months), Swedish required as test language, only respondents reporting Swedish as native language, only assessments completed as part of a recruitment or development process and only assessments completed starting January 2019 through April 2020.

The data selection process yielded a sample of 7 482 cases that met the above criteria. The data was then analysed to see how well the FFM results from the first and second assessment correlate with each other. Table F6 shows descriptive information for the sample regarding age, gender, and educational level. Table F7 show the result of the test-retest analysis of the FFM scales.

The results show correlations ranging between .77 and .81 for the FFM scales. Overall acceptable levels supporting the test-retest reliability of the Swedish language version, the analysis also shows there is room for further improvement.

Table F6
The distribution of the test-retest sample (N=7 482) across age, gender, and educational level

Age group	**Test-retest sample**
18-30 years	39%
31-40 years	18%
41-50 years	10%
51-65 years	8%
N/A	25%
Gender	
Male	35%
Female	39%
Prefer not to say	.01%
N/A	26%
Educational level	
Elementary school	5%
Middle or junior high school	29%
Less than three years of post-secondary education	11%
Three or more years of post-secondary education	27%
PHD	1%
Other	2%
N/A	26%

Table F7
Test-retest correlations, mean values and standard deviations for the test-retest sample (N=7 482)

	r	M^1	SD^1	M^2	SD^2
Agreeableness	.77*	127.63	12.02	128.99	12.42
Conscientiousness	.76*	137.62	11.71	139.00	11.81
Emotional stability	.78*	130.21	13.68	132.09	13.73
Extraversion	.81*	119.19	13.34	120.32	13.57
Openness	.80*	113.92	12.16	114.54	12.56

Note. *p<.001, r = Observed correlation, M^1 = Mean value of first assessment, SD^1 = Standard deviation of first assessment, M^2 = Mean value of second assessment, SD^2 = Standard deviation of second assessment.