.TARGETED® PREDICTION

TECHNICAL MANUAL

2017 EDITION

ANDERS SJÖBERG
SOFIA SJÖBERG

ASSESSIO

Editing and psychometrics: Anders Sjöberg and Sofia Sjöberg

Graphic design: Lina Strand

Introduction

Assessio Psychometrics AB, in collaboration with Psychometrics Sweden, has developed Targeted Prediction©, a framework for evidence-based selection. Targeted Prediction© relies on an evidence-based process to assess individuals for selection purposes. Targeted Prediction© is based on the accumulated empirical research on which individual characteristics should be assessed, how they should be assessed, and how the assessment output needs to be interpreted to achieve the most accurate basis for selection decisions.

The primary purpose of Targeted Prediction© as a framework is thus to predict future individual performance in different jobs or roles based on relevant individual characteristics. Based on the prediction, the available candidates are ranked and the rankings then serve as a basis for the selection decisions.

A key concept in the process is *predictive validity*, which here represents how accurate the assessments, and in turn the hiring decisions, are in predicting future performance. Reaching a high level of predictive validity depends on the quality and precision brought to the data collection and data combination phases of the process (see Sawyer, 1966). There are different approaches to these phases (see Figure 1), which may be more or less cost-effective and also vary with regard to their validity, as discussed in more detail below.

Figure 1. Process steps from collection of data to selection decision

Data collection

All selection decisions are preceded by the collection of information about candidates. To collect relevant data in an efficient way one needs to know the definition of constructs, tools, and predictors, how they are related and what their meaning is within the framework of assessment for selection purposes.

Constructs

A construct defines a psychological attribute. Constructs are latent, theoretical, and cannot be measured directly. For example, it is not possible to directly measure an individuals' level of emotional stability while his or her height is explicit and measurable. The definition of constructs and how they (if multiple) are related and why, is postulated by a theoretical model. Within research, theoretical models serve as empirically testable hypotheses and either accepted (almost never), rejected (often), or revised and set up for testing again. In psychological measurement, the theoretical model provides information of what is being measured and how the constructs relate to each other. The theoretical model also outline why it is relevant to measure the constructs in the specified way and provide explanations for relationships with other phenomenon. Thus, it provides the measurement score with meaning. Without a solid theoretical model with empirical support for the claims and assumptions, a measurement score will not be interpretable regarding psychological meaning.

Tools

There are numerous ways to collect data about candidates. Some are explicit and well known, for example interviews, reference checks, psychological tests. Others are less explicit, for example reading on social networks and heed others' opinions or judgement about candidates. A tool may serve as a predictor on its own, for example, an interview rendering an overall assessment. A tool may also may provide multiple predictors, for example a personality test providing scores on multiple traits.

Predictors

A predictor represents a unit of data. This unit, and others, are after collection taken forward into the data combination process (described below). What predictors represent is a crucial aspect to consider when setting up a selection process. As mentioned above, some predictors represent *tools* (e.g., an interview) while other predictors represent *constructs* (e.g., test score for conscientiousness). In addition, some predictors represent tools that measure more than one construct. The interview for example often measure constructs such as agreeableness, extraversion, and general mental ability. Some predictors also represent tools that in turn include several measurement methods. One example is assessment centers, which traditionally include both exercises (which in turn measure multiple constructs) and general mental ability tests.

Comparisons

Predictors representing *one tool and one single construct*, for example a test measuring general mental ability, make comparisons straightforward, relevant, and possible to evaluate from a utility perspective. This however, is not a characteristic of all predictors which makes it more complicated to compare them from an overall utility perspective. Nonetheless, when establishing the magnitude of the relationships between predictors and general job performance, comparisons between predictors representing single constructs, such as conscientiousness and general mental ability, and predictors representing methods, such as interviews, references, and work samples, are often made (e.g., Hunter & Hunter, 1984; Schmidt & Hunter; 1998). Thus, constructs and tools are confounded (Hough, 2001)

Comparisons between predictors representing constructs and predictors representing tools (e.g., Hunter & Hunter, 1984; Schmitt, Gooding, Noe, and & Kirsch, 1984) are based upon un-equal pre-requisites and thus encourages invalid conclusions about their efficiency. Valid conclusions based on comparison among predictors must be made between predictors representing *either* constructs or methods, thus, within in the domain of predictors.

Comparisons within the domain of predictors (tools and constructs) are equally important but serve different purposes and will provide answers to different questions. Comparing the predictive validity of constructs, such as general mental ability and the five factor model personality traits, is important for increasing our understanding of the underlying nature of these constructs and their relationships with each other and related phenomenon such as job performance. In applied selection practice, this type of knowledge, concerning why the constructs relate to each other and to job performance, is a pre-requisite for making sound choices regarding the design of the selection process. Comparing the predictive validity of different tools is also crucial. It is the predictive validity of the selection tool (not the underlying constructs) and the cost of applying them that drives the main part of the actual gain and cost of the selection process. To evaluate the overall utility of a selection method, the predictors need to represent the selection tools rather than the constructs, although the predictive validity of constructs provides the foundation and theoretical rational for combining selection tools and provide meaning to the assessment score and thus selection decision. At the same time, it is important to remember that it is the predictor score, not the tool(-s) or construct(-s) that is utilized in the data combination phase and thus prediction and selection decision.

Often the choice of tool is made without taking the underlying constructs into consideration causing redundant and expensive assessment processes. One reason for this is the incorrect assumption that different tools (e.g., personality tests, interviews, references) automatically measure different aspects of human nature (thus, different constructs) and/or that different tools contribute with unique information explaining variance in the criterion (e.g., general job performance). This is a common belief despite research showing that different selection tools, such as interviews and tests, often measure the same underlying constructs (Roth & Huffcut, 2013). As mentioned above, working under this assumption could likely lead to the use of needlessly expensive data collection methods, methods which lack incremental validity, or multiple tools with predictors which overlap on the construct level – altogether having a severe negative impact on overall utility.

In summary, a general rule of thumb is that the more standardized the tool is, the greater the certainty of what is being measured, the higher the predictive validity, and the greater the cost-effectiveness. Thus, psychological testing in selection processes has gained its superior position due to:

- the knowledge, awareness and relevance of what is being measured
- reliable and valid measurement
- a strong theoretical and empirical foundation of GMA and five factor model personality traits
- its highly standardized format ensuring consistency across candidates
- its high level of predictive validity
    - the level of GMA and, to a certain extent, some personality traits have been proven to be generic due to their impact on performance in all types of jobs and roles
    - the lack of overlap between the constructs
- its cost-effectiveness compared to other tools measuring the same and/or overlapping constructs

The possibility of choosing which constructs to assess, what tools to use, and what predictors can and should be utilized for, thus provide practitioners with the opportunity to influence the accuracy, cost-effectiveness, and fairness of their selection processes.

Data combination

Simply collecting data is not sufficient for assessing and ranking candidates when multiple data sources are used. When multiple tools produce multiple types of units of data, the mass of information gathered on each candidate becomes unwieldy and requires some further organization: the data needs to be *combined* and unified into an *overall assessment*. This overall assessment will then provide the basis for ranking and decision making. This process applies regardless of the tool(s) used to collect data, the number of data units, and whether it is conducted explicitly or implicitly.

The approach utilized for data combination, and the premises it is based on, is critical in determining the accuracy (validity) of the selection decisions. Within selection processes, it is common to use tools such as psychological tests which are well-documented and have fairly high predictive validity in their *test scores*. Nevertheless, the selection *decisions* may still turn out to have poor validity due to a defective combination of the test scores and/or the addition of units of data beyond the test scores. Targeted Prediction© regulates and avoids this possibility by maximizing the use of the collected data in a standardized fashion. How this is achieved is described in detail later in this manual and will be preceded by a historical review and a description of current practice regarding the process of combining data for selection purposes.

Clinical data combination

The combination of data may be carried out in one of two different ways: clinically or mechanically. The two approaches are exclusive and not compatible with each other. However, a common approach in practice is to combine mechanically in one phase of the selection process before clinically integrating this assessment with other data units in a later phase.

The most common approach in selection practice is what is known as clinical data combination (Viteles, 1925). The starting point for this approach is a job profile that indicates the performance expectations for the position or role. The job profile is then treated as the optimal benchmark. The traditional job profile is usually defined and expressed in qualitative terms, for example, in descriptions of desirable qualities and/or behaviors. Less desirable qualities and behaviors are rarely explicitly stated at this point.

Within the tradition of clinical data combination, test results are typically viewed as a part of a whole, often as a complement to or basis of discussion. An *assessor* (or group of assessors) decides *if* and *how* the test results, and other data, should be combined into the overall assessment. It is also the assessors who, through an *implicit mental process*, carry out the data combination for each candidate. They are then responsible for relating and comparing the overall assessment to the job profile, evaluating the fit (or match), and ranking the candidates according to the degree to which each suits the profile. Processes which involve clinical data combination and the results of such processes, including the ranking of candidates, are rarely explicit.

Mechanical data combination

The alternative approach for combining data is mechanical data combination (Freyd, 1926) which comprises:

- a *pre-determined* specification (corresponding to a job profile) of what will be combined. The chronology is key, as the specification must be formulated before and not after the data has been collected, and it should not (without explicit awareness) be changed afterwards.
- an *explicit* specification – in other words, it should be clearly expressed and documented so that stakeholders have access to and, if necessary, can monitor changes to the specification.
- a specification describing the *logic* behind the combination – this includes a rationale for the inclusion of predictors, describing how they are combined and why. The logic may only involve a simple summary of predictor scores or it could be based on advanced algorithms with a large number of predictors, weighted according to empirical estimates of the relationship between predictors and criteria and taking into consideration the fact that predictors overlap.
- a *mechanically* carried out combination in which the calculations are performed in a standardized fashion via machines in order to guarantee consistency (reliability) across the candidates and leave no room for subjectivity. Other assessment criteria, not related to the job profile, would invariably come into play if subjectivity were to have a role.

Thus, a mechanical approach does *not* allow for the professional assessor to use his/her personal judgment

- to determine the job profile (content or weighting).
- to combine the data (e.g., test scores and outcomes of interviews).
- to match this to the job profile.
- to rank candidates.

Instead, these parts of the process are pre-defined and explicit and occur through a standardized mechanical process.

The superiority of mechanical data combination

As early as in the 1950s, research showed that mechanical data combination is superior to the clinical approach with regard to predicting behavior, including job performance (Meehl, 1954). In a later meta-analysis, Grove, Zald, Lebow, Snitz, and Nelson (2000) found that out of the 136 studies included, 63 found mechanical combination to be advantageous, 8 showed that clinical data combination was superior, and 65 studies showed that the two approaches led to equivalent levels with regard to predictive validity.

Notably, it is the differences between the clinical and the mechanical approaches as such that have the greatest impact on validity, not how the various units of data are weighted (Sawyer, 1966; Grove et al, 2000). To critically fault mechanical data combination on the grounds of not achieving perfect weights is thus not logically sound. The research clearly shows that mechanical data combination leads to higher prediction compared to clinical combination, given that the same predictors are used (Kuncel, Klieger, Connelly & Ones, 2013). The current primary explanation behind these findings is that the clinical data combination approach entails inconsistency across predictors and candidates causing greater error.

In addition to having superior predictive validity, mechanical data combination is in general more cost-effective compared to clinical data combination. This is partly due to the difference in cost for applying each approach. Once an equation is in place using a mechanical approach, the cost per candidate is considerably lower compared to the clinical approach, where the data for each candidate is combined uniquely and individually by the assessor(s). Despite this, the clinical approach is still the most common approach on the market, providing big business for those in the field of workplace assessment.

The *replicability* and *transparency* of the mechanical approach also allow for monitoring, systematic evaluation, and thus continuous improvements. By using an explicit, pre-specified, and standardized process, the monitoring and evaluation of outcomes becomes possible. After the evaluation, predictors may be replaced and weights may be changed etc., which contributes to greater reliability and validity as the empirical basis grows. Moreover, this also enables the possibility of replicating it if done successfully.

The same is not possible with the clinical combination approach, since there is no explicitly stated or defined process for evaluating the outcomes. The subjectivity of the assessor(s) will constantly create shifts in the weighting that do not take the criterion in question into consideration; this is the essence of the clinical approach. An association between the overall decision and the outcome may occur nonetheless, but the specific grounds underlying it will be unknown. Replication is therefore not possible.

With mechanical data combination, replication and transparency are necessary in order to compare candidates according to the same premises; for all candidates, the process consistently takes relevant data into account, while excluding irrelevant data from the assessment. This is especially important in the *recruitment* phase. Being able to explicitly explain beforehand to candidates the standards by which they will be assessed, including how and why they are being assessed, is central for transparency. It ensures that candidates have the opportunity to ask informed questions and decide for themselves if they want to continue with the selection process.

Replication and transparency also make *traceability* possible. It is preferable to be able to explain why one candidate was offered the position over another. An important benefit of traceability is being able to logically trace the selection process, thus making it possible to account for and demonstrate how the decisions were derived.

Some of the above may seem rather obvious, and the clinical data combination approach lay claim to most of these points despite the inherent lack of possibility to achieve them. With its non-standardized format, the clinical data combination approach leaves open the possibilities that relevant data might not be taken into consideration while irrelevant information is, and that the weighting could be arbitrary and carried out in a less suitable way. It carries the inevitable consequence of inconsistency – that the basis for assessment varies between candidates. In practice, this, intentionally or non-intentionally, leads to various kinds of special treatment, which increases the risk of adverse ramifications, including discrimination.

Despite the superiority of mechanical data combination, it is rarely fully implemented in selection practice. The clinical approach remains the standard procedure within practical selection processes today, although the interest and demand for standardized and evidence-based solutions such as mechanical combination have increased in recent years.

The following sections describe how the empirical "profiles", manifested in algorithms, has been developed, thus explaining how the mechanical combination of (predictor) scores from Matrigma and MAP are processed within the Targeted Prediction© framework, and how the results should be interpreted.

Development of the Targeted Prediction© framework

The *purpose* of Targeted Prediction© is to generate candidate rankings for personnel selection decision making with respect to their future level of job performance for a specific role, function, or criterion. The *goal* of Targeted Prediction© is to deliver the most valid and reliable basis for selection decisions, given the predictors GMA and personality. The *approach* for achieving this goal relies on the mechanical data combination of empirical, evidence-based, and well-documented predictor scores and predictor weights – thus algorithms.

Conceptually, the Targeted Prediction© framework is based on pre-determined and pre-developed specifications in the form of algorithms . The algorithms correspond to traditional job profiles, being theoretically based – thus, being a hypothesis, but are instead empirically based and designed for a specific role, function, or criterion. Each specific role, function or criterion represents a "target". Every algorithm is based on the same predictors, test scores representing GMA and personality, and specifies how the predictor scores need to be weighted to achieve the most accurate (valid and reliable) ranking of candidates. The Targeted Prediction© framework also ensures fairness regarding equal treatment since each algorithm is applied in a standardized fashion to each candidate's set of test (predictor) scores.

What should be measured?

The question of *what* should be measured and *how* it should be weighted often become main topics in discussions on selection. Although the answers sometimes depend in part on the specific selection context, certain individual characteristics have been shown to be generic and relevant to all kinds of selection contexts, regardless of the candidates, the job, or the role. As mentioned previously, the level of GMA and, to a certain extent, some personality traits have been proven to be generic – they have impact on performance in all types of jobs and roles (Barrick & Mount, 1991; Barrick, Mount & Judge, 2001; Hurtz & Donovan, 2000; Mount & Barrick, 1995; Salgado, 1997, 2003; Barrick & Mount, 2005). On the strength of these findings, data about these individual characteristics is utilized within the Targeted Prediction© framework.

The predictors that form the basis of all Targeted Prediction© algorithms are derived from the five-factor-based personality test, MAP – Measuring and Assessing Individual Potential (Sjöberg, Svensson & Sjöberg, 2017) and from the GMA test, Matrigma using either the classic version (Mabon & Sjöberg, 2017) or the adaptive version (Mabon, Niemelä, Sjöberg, & Sjöberg, 2017). The development of these tests, their psychometric properties, and their relevance for the prediction of job performance, is outlined in their technical manuals. The scores from the scales (one scale in Matrigma and five scales in MAP) thus constitute the *predictor scores* in each Targeted Prediction© algorithm.

How should the predictors be weighted?

There are various approaches for determining the weights for each predictor. Often it is acceptable to produce a simple summation or average for the data by applying equal weightings on all predictors. In other cases, however, it is beneficial to apply differentiated weights. Research shows, for example, that GMA should be given twice the weighting compared to personality (all traits taken together) when the goal is to predict future job performance (Sjöberg, Sjöberg, Näswall & Sverke, 2012).

Weights may be applied based on theoretical assumptions or hypotheses, or on an empirical basis – the latter being the case with algorithms within the Targeted Prediction© framework. In empirical weighting, the weights have been estimated and determined through *actual measurement* of the links between predictors and what is being predicted, the criterion (e.g., job performance). Empirical weights may be estimated using a number of different approaches and methods, such as local validity studies and meta-analyses, and the statistical analyses may be grounded in any of a diverse range of theoretical assumptions (Le & Schmidt, 2006; Sjöberg, Sjöberg, Näswall & Sverke, 2012). How the empirical weights are determined within the Targeted Prediction© framework and its implications are described below.

Development of Targeted Prediction© algorithms

As mentioned above, the weights assigned to predictors may be determined in a number of ways. Each Targeted Prediction© algorithm is based on empirical weights and thus on actual correlations between the predictors and the target - the criteria representing performance in a role or job. The relationships between predictors and criteria has been measured and established in one or more previous validity studies. A single validity study is referred to as a local validity study, and a meta-analysis is based on the results from a larger number of local validity studies. The advantage of consulting results from meta-analyses is that the estimates are more exact and secure (reliable) and generalizable across jobs, roles, and criteria. Local validity studies may be useful for estimating predictor weights but require large samples, careful design and a thorough statistical processing.

The weights in Targeted Prediction© algorithms are designed to represent the *optimal* weights for each predictor and each specific criterion. Optimal weights are needed to maximize the accuracy (the validity) of the overall score which the Targeted Prediction© algorithms generate. Optimal weights make the maximum use of each predictor, implying that applying other weights for the same predictors would only lower the predictive validity. Optimal weights are determined, in part, by taking into consideration that the predictors covary (overlap) with one another to some extent. The fact that predictors lack independence is rarely taken into account, causing a decrease of the validity of the prediction. The pattern of covariation between predictors differs between Targeted Prediction© algorithms. The degree of covariance is estimated using local studies, compiled into a so-called correlation matrix, and is defined for each algorithm. The correlation matrix is then used in the regression analysis which estimates the relationships between each predictor and the criteria. The estimates from the regression analysis correspond to the weights.

The correlations between the predictors are also needed to formulate a basis for decision making that is founded on a *compensatory approach* (such that low scores on one scale may partly be compensated for by a high score on another). There is, for example, an overlap between test scores on the Agreeableness scale in MAP and the other MAP scales (predictors). This means that only a small part of the variation in the criteria uniquely covary with Agreeableness. The regression analysis takes care of this and simplifies the combination by only taking into consideration the variation in the score for Agreeableness that is unique *and* has predictive validity.

Note that the compensatory approach is often a feature attributed to the clinical data combination approach, as implicit interpretation often aspires to handle such adjustments during an assessment and thus achieve even higher validity. However, it is not possible to apply a compensatory approach by combining data clinically. Note also that several other approaches for interpretation (and decision making), for example applying cut off scores on predictor scores, inherently lack the attributes for generating an overall assessment based on a compensatory approach.

Since the regression analysis takes the relationship between each individual predictor and the criterion into consideration *as well as* the relationships between predictors, the weights in a Targeted Prediction© algorithm may differ from correlations that are estimated without taking these aspects into consideration. Weights not based on these considerations may thus not be optimal.

The Targeted Prediction© framework consist of several algorithms. In the following, information about each specific target along with information about how the algorithm has been developed and what empirical data underlies the algorithm.

Note that each algorithm has a unique label, for example *Targeted Prediction© Leadership*, and the results are referred to as indexes: *Targeted Prediction© Leadership Index*.

Targeted Prediction© Job Performance

Numerous more or less empirically based models and perspectives have been put forth on job performance, including varying definitions of its nature, importance, and scope. The model which currently has the strongest support from research defines job performance within a hierarchical structure. At the top of this hierarchy is a comprehensive factor labeled *general job performance* (Viswesvaran, Schmidt & Ones, 2005). This is the criterion used for the Targeted Prediction© Job Performance algorithm. The general job performance factor encompasses *all actions and behaviors linked and/or contributing to the organization's goals* (Viswesvaran & Ones, 2000). Both performance in actual work tasks and the ability to handle the social aspects of work are thus subsumed within this overarching factor.

General job performance, in turn, functions as an umbrella term for three related but conceptually distinct dimensions: *Task Performance* (TP), *Organizational Citizenship Behavior* (OCB), and *Counterproductive Work Behavior* (CWB). Task Performance encompasses performance on actual tasks which contribute to the production of the organization's goods or services (Rotundo & Sackett, 2002) and which are formally recognized by the employer (Borman & Motowidlo, 1993; Conway, 1999). In general, TP behaviors and activities contribute to taking responsibility for and completing tasks. Carrying out work to a high standard and finishing it in good time are both qualities of high TP.

Organizational Citizenship Behavior is defined as all of the voluntary and positive behaviors that are *not* covered by TP but which still promote the organization's aims in various ways. Organ's (1988) definition is likely the most recognized:

'Organizational Citizenship Behavior are individual behaviors that are discretionary, not directly or explicitly recognized by the formal reward system, and in the aggregate promote the efficient and effective functioning of the organization' (Organ, 1988, p. 4).

This type of deliberate behavior, which contributes to a well-functioning organization by supporting the overall organizational, social or psychological environment, encompasses behaviors which are not directly or explicitly recognized by the organization's formal reward system.

In the hierarchical structure of job performance, a distinction is made between 'good' job performance (represented by TP and OCB) and 'damaging' or counterproductive work behavior which is linked to the work or the organization. The usual and most generally accepted definition of CWB is 'any intentional behavior on the part of an organization member viewed by the organization as contrary to its legitimate interests' (Sackett & DeVore, 2001, p. 145). In other words, CWB is an umbrella term for all conscious and deliberate behavior carried out by an employee of an organization which, in some way, has a negative impact on or damages the organization or its employees.

Targeted Prediction© Leadership

The Targeted Prediction© Leadership algorithm is based on research showing that it is possible to predict overall performance as a leader (Judge, Bono, Ilies & Gerhard, 2002). The model with the strongest support postulates that leadership consists of two main components .The first component represents the extent to which the individual is likely be perceived as a leader by his or her coworkers. This aspect of leadership is known as Leadership Emergence. The second component is known as Leadership Effectiveness and reflects how effective the individual is as a leader, especially in terms of how well the individual provides leadership for his or her coworkers regarding their engaging in activities or behaviors related to and/or contributing to the organization's goals. Together, the two components represent the overall leadership performance criteria in the role of leader, and the Targeted Prediction Leadership algorithm is constructed to predict this overall leadership target (Judge et al., 2002).

Targeted Prediction© Service

The Targeted Prediction© Service algorithm is based on research showing that it is possible and meaningful to predict service performance based on personality and GMA (Hurtz & Donovan, 2000). Successful companies within the service sector has shown to add benefits to their offering that not only satisfy the customers but also surprise and delight them. Delighting customers is a matter of exceeding their expectations, this requires that go beyond the expected level of service. Thus, the Targeted Prediction© Service algorithm aim to predict the potential of customer service defined as adding benefits to their offering that satisfy, surprise and delight customers. The Targeted Prediction© Service algorithm is based on meta-analytic results combining personality and GMA to predict service potential and to facilitate the selection of for service-oriented positions.

Targeted Prediction© Sales

The Targeted Prediction© Sales algorithm is based on Assessio's own research and meta analytic and aim to facilitate the selection of for service-oriented positions. The Targeted Prediction© Sales algorithm is based on the relationship between personality and sales performance defined as the ability to "win" at each stage of the customer's buying process and ultimately earn the business on the right terms and in the right timeframe. Key objectives of the sales performance criteria is to set goals, give feedback to and satisfy customers within the boundaries of the employer's interests. The Targeted

Prediction© Sales algorithm thus predict the potential for performance in jobs and roles targeting these behaviors.

The basis for Targeted Prediction© algorithms

As mentioned above, the development of a Targeted Prediction© algorithm essentially requires two different types of information: the correlations between the *predictors* and the *criterion* (criterion validity), and the correlations *between* the predictors. This information may be obtained in numerous ways and from several different sources. The Targeted Prediction© algorithms are based on meta-analytic estimates for the correlations between the predictors and the criterion and on local validity studies for the estimation of correlations between predictors.

Correlations between predictors

The same correlations *between the predictors* are used in all Targeted Prediction© algorithms. The calculation of correlations between the scales in MAP is based on the Swedish norm group consisting of 569 individuals (see the technical manual for more information). Note that the sample is ideal for this purpose as it comprises a group of individuals being representative for the normal population of Sweden regarding age, gender and educational level.

The correlations between Matrigma and each scale in MAP are calculated using a sample taken from Assessio's database (N=296). These individuals were, at the time of assessment, employed within the Norwegian retail sector and Matrigma was therefore administered with Norwegian instructions. The majority were women (65%), ranging in age between 18 and 60 years (M=32, SD=14). The majority had completed three years of senior high school (56%) or higher (29%).

Correlations between predictors and performance

The correlations between personality predictors, the factors in the five-factor model, and the job performance criterion are taken from the Gonzalez-Mulé, Mount, and Oh meta-analysis (2014, Appendix D). The job performance criterion, as described above, consists of the components TP, OCB, and CWB. How these components are weighted in relation to each other to form an overall measure of performance is described in Gonzalez-Mulé et al. (2014, appendix D).

The correlation between GMA and job performance is based on estimates presented in the meta-analysis carried out by Hunter, Schmidt, and Le (2006). This estimation is based on job performance being of average complexity.

Correlations between predictors and leadership

The correlations between the factors in the five-factor model and the criterion of leadership are based on the results from the meta-analysis carried out by Judge et al. (2002).

Similar to job performance and its components, the two leadership qualities of

Emergence and Effectiveness are combined to form the overarching criterion of leadership. This is described in detail in Judge et al. (2002).

The correlation between GMA and leadership is based on estimates from the same meta-analysis as for performance, Hunter et al. (2006). The estimation is based on leadership being of average complexity.

Correlations between predictors and service

The correlations between the predictors of personality and the criterion of service are based on the results from the meta-analysis carried out by Hurtz & Donovan (2000).

The correlation between GMA and service is based on estimates from the same meta-analysis as for performance, Hunter et al. (2006). The estimation is based on service being of average complexity.

Correlations between predictors and sales

The correlations between the predictors of personality and the criterion of sales are also based on the results from the meta-analysis carried out by Hurtz, G., & Donovan, J. (2000).

The correlation between GMA and service is also based on estimates from the same meta-analysis as for performance, Hunter et al. (2006). The estimation is based on service being of average complexity.

Determining the Targeted Prediction© algorithms

To formulate the Targeted Prediction© algorithms, the predictor weights for the criteria were determined through regression analyses. The results of the regression analyses provide the overall validity for the algorithm as well as the weights (so-called beta weights) for how each predictor should be weighted in each algorithm. The overall validity, expressed as a multiple correlation coefficient, have been estimated to be R=.54 for Targeted Prediction© Job Performance, R=.74 for Targeted Prediction© Leadership, R=.76 for Targeted Prediction© Service, and R=.78 for Targeted Prediction© Sales.

The exact beta weights are not presented in this manual, as these are the intellectual property of Assessio International AB. For Targeted Prediction© Job Performance, however, the results from Matrigma have the heaviest weighting, and Conscientiousness and Extraversion (negative) each have roughly half as much weight, followed closely by Agreeableness. Emotional Stability (negative) is given a limited weighting, and Openness for experience has practically zero weighting.

In the Targeted Prediction© Leadership algorithm, results from Matrigma are given a heavy weighting, Conscientiousness has half as much weighting, and the other scales (Agreeableness (negative), Openness for experience, and Extraversion) have marginal weightings.

In the Targeted Prediction© Service algorithm, results from Matrigma are also given a heavy weighting, Conscientiousness approximately half as much weighting, followed by Extraversion and Agreeableness having a significantly lower weighting, while Emotional stability and Openness for experience have marginal weightings.

In the Targeted Prediction© Sales algorithm the predictor weights follow the same pattern as the Targeted Prediction© Service algorithm regarding heavy GMA weighting and approximately half as much weighting for Conscientiousness. However, Agreeableness has a marginal weight in this algorithm just as Openness for experience and Extraversion.

Interpretation of results

The foundation for all Targeted Prediction© results are the standardized test scores (predictors) from MAP and Matrigma, therefore the quality and norm groups used in the calculation of these test scores is essential. Instructions for administration, information about theoretical background, development, evidence of validity, evidence of reliability, calculations of test scores, and applied norm groups, is outlined in the technical manuals (Sjöberg et al., 2017; Mabon & Sjöberg, 2017). Note that it is the responsibility of the *test administrator* to assimilate this information and to follow the recommendations and guidelines to ensure that requirements for using these assessments are met.

Interpretation of Targeted Prediction© results

The results generated by the Targeted Prediction© framework, regardless of algorithm, is as a C-score for each candidate and is labeled index (e.g., *Leadership Index).* For more information about the C-scale and its properties, see Sjöberg et al. (2017). The C-score provide the basis for the ranking of candidates and the ranking constitutes the direct basis for decision-making. The selection decision is to be made from a top down approach, where *the higher the score, the greater the probability that a candidate will demonstrate better performance in the actual role or job (represented by the criterion) compared to candidates with a lower score*. A Targeted Prediction© score is *not* intended to constitute a basis for discussion, or to be included or incorporated into a clinical data combination approach. Targeted Prediction© scores optimizes the use of GMA and personality for personnel selection decision making, thus implying that the ranking of candidates provided by Targeted Prediction© has the maximum possible accuracy given the predictors and criteria at hand. It is thus not possible to increase the validity in the prediction by interpreting scores on single predictor scores (test scores from MAP or Matrigma).

Feedback reports

Standardized test administrator and candidate feedback reports are available in accordance with regular assessment using Matrigma and MAP, and information concerning this is provided in the technical manuals (Sjöberg et al., 2017; Mabon & Sjöberg, 2017). Thus, there are no standardized individual candidate feedback reports available specifically on Targeted Prediction© results, the reasons for this are several. First and foremost, it is appropriate to provide feedback to candidates on the *constructs* being measured. This imply that regardless if MAP and Matrigma is applied within a Targeted Prediction© process or as single assessments, GMA and personality are the constructs being measured and feedback should be based on these constructs.

Second, Targeted Prediction© is a *process* utilizing assessment scores in accordance with a standardized procedure, providing a score for personnel selection decision making. It is not as assessment tool by itself. The design of the process, the choice of Targeted Prediction© algorithm, and the generated outcome is connected to the selection context. Thus, the implications of the Targeted Prediction© results is dependent upon the selection context (e.g., choice of Targeted Prediction© algorithm, number of applicants, selection ratio, other applicant's scores). This, along with the fact that the Targeted Prediction© score constitutes the direct and recommended basis for decision-making makes standardized candidate feedback on the Targeted Prediction© results inappropriate. Feedback to candidates regarding the implications of the Targeted Prediction© process thus requires to be handled by the test administrator.

On group level, Targeted Prediction© generates a report which contains a compilation of the candidates' results included in a project. In addition to the Targeted Prediction© results the report contain information regarding the target, demographics of the group, and some project administrative information (e.g., assessments used, devices used).

On group level, for Targeted Prediction© Ascend generates a report, labeled Project report, which contains a compilation of the candidate's results included in a project. In addition to the Targeted Prediction© results, the report contains administrative information about the project (e.g., assessment used), information regarding the chosen target, and descriptive information of the group (e.g., age, gender, educational level). The aim of the report is to provide an overview of your Targeted Prediction project with the purpose of simplifying decision-making in personnel selection processes. The project report is intended for the project administrator but may be suitable for all stakeholders involved in the personnel selection decision process.

Summary

Targeted Prediction© enables a standardized, evidence-based personnel assessment and selection process, based on differences in individual characteristics relevant to performance in the workplace.

Targeted Prediction© offers a mechanical data combination process applying empirically based algorithms defining weights for predictor scores; test results from the five-factor-based personality test MAP (Sjöberg et al, 2011) and the non-verbal GMA test Matrigma (Mabon & Sjöberg, 2016).

Targeted Prediction© allows users to follow ISO 10667; Assessment Service Delivery (ISO/IEC, 2011). The client, together with the supplier of the assessment service, can *pre-determine* the applicable Targeted Prediction© algorithm. This algorithm is then applied uniformly to each candidates' set of predictor scores. Based on the combined assessments, the candidates are *ranked* for selection decisions. The risk of intentional and non-intentional discrimination is minimized, and the relevant foundation for the assessment is used in the most optimal way – to maximize the accuracy and fairness of the selection process.

References

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance. *Personnel Psychology, 44*, 1–26.

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.

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18,* 359–372.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco, CA: Jossey-Bass.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84,* 3–13.

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

Gonzalez-Mulé, E., Mount, M. K., & Oh, I.-S. (2014, August 18). A meta-analysis of the relationship between general mental ability and non-task performance. *Journal of Applied Psychology.* Advance online publication. http://dx.doi.org/10.1037/a0037547

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.

Hough, L. M. (2001). I/Owes its advances to personality. In B. W. Roberts & R. T. Hogan (Eds.), *Personality psychology in the workplace* (pp. 19-44). Washington: American Psychological Association.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.

Hunter, J. E., Schmidt, F. L., & Le, H (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, Vol. 91, No. 3,* 594–612.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.

ISO/IEC, (2011). ISO/IEC 10667:2011 Bedömningstjänster i arbetslivet – Processer och metoder för bedömning av människor i organisationer. Del 1: Krav på uppdragsgivare. Del 2: Krav på leverantör. Stockholm: Author.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87,* 765–780.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology, 98*, 1060–1072.

Le, H., & Schmidt, F. L., (2006). Correcting for Indirect range restriction in meta-analysis: testing a new meta-analytic procedure. *Psychological Methods, 11,* 416–438.

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

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis, MN: University of Minnesota.

Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management, 13*, 153–200.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.

Roth, P. L., & Huffcut, A. I. (2013). A meta-analysis of interviews and cognitive ability. *Journal of Personnel Psychology, 12,* 157–169.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology, 87*, 66–80.

Sackett, P. R., & DeVore, C. J. (2001). Counterproductive behaviours at work. In N. Anderson, D. S. Ones, H. K. Sinangil, & V. Viswesvaran (Eds.), *International Handbook of Work Psychology* (Vol. 1, pp. 145–164). London, UK: Sage Publications.

Salgado, J. F. (1997). The five-factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82*, 30–43.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychological, 76*, 323–346.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178–200.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Schmitt, N., Gooding, R. Z., Noe, R. D., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-422.

Sjöberg, S., Sjöberg, A., Näswall, K., Sverke, M. (2012). Using individual differences to predict job performance: Correcting for direct and indirect restriction of range. *Scandinavian Journal of Psychology,* DOI: 10.1111/j.1467-9450.2012.00956.x

Sjöberg, A., Sjöberg, S., & Forssén, K. (2006). *Predicting Job Performance. Manual.*

Stockholm: Assessio International.

Sjöberg, S., Svensson, C., & Sjöberg, A. (2017). *MAP: Measuring and Assessing Individual Potential. Technical manual.* Stockholm: Assessio International.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90,* 108–131.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8,* 216–226.

Viteles, M. S. (1925). The clinical viewpoint in vocational selection. *Journal of Applied Psychology, 9*, 131–13.