MiNT

MEASURING INTEGRITY

TECHNICAL MANUAL

2017 EDITION

SOFIA SJÖBERG
ANDERS SJÖBERG

ASSESSIO

Graphic design: Lina Strand, Johan Larsson

ISBN: 978-91-7418-355-9

Item no.: 005-201

Content

Preface 3

Part 1 5

Counterproductive behavior in organizations 5

Definition of CWB 5

A model of CWB 5

Financial consequences of CWB 9

Individual differences and work-related behavior 10

Personality 10

Personality and job performance 10

Personality and CWB 11

Summary 12

Part 2 14

Integrity 14

The construct of integrity 14

The development of integrity tests 14

The development of personality-based integrity tests 15

Operationalizing the construct of integrity 16

Integrity, CWB, and job performance 17

Summary 17

Part 3 18

The initial development of MINT 18

Item construction 18

Integrity – a hierarchic measurement model 20

Confirmatory Factor Analysis 20

Evidence of reliability 23

Internal consistency 23

Test–retest 24

Part 4 25
Validity – Evidence based on relations to other variables 25
Minnesota Multiphasic Personality Inventory-2 (MMPI-2), Inventory of Interpersonal Problems (IIP), and job involvement 25
Hogan Development Survey (HDS) 28
Measuring and Assessing Individual Potential (MAP), ServiceFirst, and Matrigma 29
Test–criterion relationships 31
Job performance and job satisfaction among traffic controllers 31
Students and grades 32
Job performance and CWB among grocery store employees 33
Job performance among retail employees 33
Summary 34
Part 5 35
Further development, language versions, and norm updates 35
October 2011 35
December 2014 35
Evidence based on internal structure 36
Confirmatory Factor Analysis – Goodness of Fit 37
Multiple-Group Confirmatory Factor Analysis – Goodness of Fit 38
Internal consistency 39
Part 6 40
Financial aspects of personnel selection 40
The purpose of personnel selection 41
Choice of selection method 41
MINT for selection 44
Part 7 45
Instructions for use and interpretation 45
Areas of use 45
Administration and scoring 45
Requirements for testing 45

Information provided to candidates before testing 46
Interpretation and presentation of results 47
Standardized feedback reports 48
Standard error of measurement 49
Summary 50
References 51
Appendix A 57
List of figures and tables 57
Appendix B 59
Development of MINT - Norwegian version 59
Evidence Based on Internal Structure - Confirmatory Factor Analysis 60
Internal consistency and reliability 62
Appendix C 63
MINT – Implications of the hierarchic model for interpretation 63

Preface

It is a well known fact that people are not always productive when they are at work, and this lack of productivity can be measured in a number of ways. Counterproductive Work Behavior, CWB, may be generally defined as any form of behavior that can harm fellow workers or the organization that employs them. It can range from subtle discourteousness to more severe acts of bullying, aggression, and retaliation. A person may engage in CWB by stealing the goods and services being produced, coming to work under the influence of alcohol or drugs, or by sabotaging work efforts.

Surveillance and control systems have been developed in order to prevent these behavior patterns and to catch the guilty parties; however, as these systems have become more advanced, the creativity of those who wish to circumvent them has kept pace. Since this type of negative behavior often leads to significant financial losses for an organization, business leaders and HR practitioners have spent years researching ways to counteract this. The focus has been on finding other methods for identifying uncooperative and hostile people, preferably before they join the organization. But this is no easy task. It is likely that a dishonest job applicant will supply dishonest answers to questions relating to their previous work experience or CWB.

The question then arises as to whether it is possible to measure honesty, or the closely related concept of integrity, using indirect methods. If it can be shown that certain personality traits are related to a lack of honesty or integrity, the logical procedure would be to measure these traits as part of a selection process. In order to increase the probability that applicants respond in a truthful and sincere way that is useful for predicting future behavior at work, the underlying purpose of the items should not be apparent to them.

In essence, this is what personality-based integrity testing is all about. For the last twenty years, attempts have been made, with varying degrees of success, to develop tests for the aforementioned purposes. Measuring INTegrity, MINT, is an attempt to build on this tradition, using a somewhat different psychometric approach. Research has established that some of the personality dimensions in the Five-Factor Model, FFM, can be used to measure the higher order concept of integrity, which in turn predicts CWB. This knowledge has provided the basis for the development of MINT. The test has been developed using items measuring Emotional Stability, Conscientiousness and Agreeableness. In addition, the psychometric development process has applied a so-called hierarchic model, which functions on three levels. The highest level mainly represents Emotional Stability, while the next level is mainly represented by Conscientiousness and Agreeableness. A third level, which comprises different types of CWB, may also be identified. The highest level is obviously the one that is central to the concept of integrity, while the other levels can be studied separately on the basis of the hierarchic model. MINT has in other respects been shown to have excellent psychometric properties.

Hunter Mabon,
Stockholm, November 2007

This manual outlines the initial construction and continuing development of the personality-based integrity test MINT – Measuring INTegrity. The content is primarily presented in chronological order, divided into the stages of development, validation and norm updates.

In Part 1, CWB, which MINT is designed to predict, is theoretically defined and related to behavior in organizations. Part 2 deals with the concept of integrity in a similar way, focusing also on its measurement.

In Part 3, the development of MINT is presented, along with how the hierarchical model is operationalized and a description of the empirical support for the theoretical framework and measurement model.

In Part 4, validity in terms of the relationship between MINT and other variables is presented.

Further development and work regarding norm updates is described in Part 5, followed by Part 6 where the financial impact of employees engaging in CWB is explained. Part 7 provides instructions for using and interpretation MINT and its results.

Part 1

Counterproductive behavior in organizations

This part begins with a definition of Counterproductive Work Behavior, CWB, as well as a description of a theoretical model that describes different types of CWB. It describes the economic consequences of CWB and concludes with research on the relationship between personality traits and work-related behavior, primarily CWB.

Definition of CWB

CWB comprises a number of different types of behaviors, some of which are classified as criminal while others are not. These behaviors have been defined and labeled as deviant behavior, antisocial behavior, and dysfunctional behavior (Kidwell & Martin, 2005). Two of the leading researchers in the field, Sackett and DeVore (2001), define this group of counterproductive work behaviors as follows:

> *Any intentional behavior on the part of an organization member viewed by the organization as contrary to its legitimate interests. (Gruys & Sackett, 2003; p. 30)*

This definition comprises types of behaviors directed at the organization and/or individuals within it; both kinds may have serious consequences for the organization. This definition includes individual behaviors within the organization but not those involving customers or previous employees. Sackett and DeVore's definition is also entirely focused on behaviors, rather than on their consequences, such as the possible damage they give rise to. It is also important to note that the definition only comprises intentional behavior.

A model of CWB

As late as in the 1980's, a considerable amount of research was conducted on individual counterproductive behaviors. There was no common frame of reference, however, nor a theory appertaining to these behaviors, and each ad hoc study looked at specific types of behavior rather than attempting to study CWB as a general phenomenon.

Hollinger and Clark (1982) created an empirically supported conceptual framework in which CWB were divided into two categories. The first of these categories is labeled "Property Deviance" and encompasses behaviors that misuse company assets, as in cases of theft, damage, and abuse of privileges. The second category is called "Production Deviance" and refers to actions that disregard existing norms concerning how work should be performed. This latter category comprises all absences that are not scheduled, such as overlong breaks, as well as behaviors leading to reduced work capacity, including carelessly or slowly performed work. Robinson and Bennett (1995) noted, and confirmed empirically, that Hollinger and Clark's categories lacked an interpersonal dimension that distinguished between types of behavior directed at the organization (Production and Property Deviance) and those directed toward members of the organization. They also noted that the categories lacked a dimension representing a continuum from less to more serious breaches of the law (Gruys & Sackett, 2003). Robinson and Bennett's typology of CWB thus results in four categories.

Behavior towards organizations

Serious offences

Minor offences

Behavior against other organizational members

Serious offences

Minor offences

More recent research (Sackett & DeVore, 2001; Gruys & Sackett, 2003) has shown that CWB should be regarded in terms of a hierarchic model consisting of three levels. The highest level is that of a general CWB factor. At this level, no distinction is made between CWB aimed at individuals and CWB aimed at the organization or tasks. In practice this means that those who engage in some form of CWB, such as theft, are likely to engage in other forms of CWB such as high absenteeism. The reasoning here refers to general probabilities and does not necessarily take into account the relative seriousness of different forms of CWB. Whether or not a person who takes extended breaks will turn to more serious offences such as workplace violence is still to be determined.

The second level in the hierarchy, below the general CWB factor, consists of two dimensions: Interpersonal–Organizational and Task Relevance. The first of these is identical to the one presented by Robinson and Bennett (1995) and represents the extent to which CWB is aimed at either the organization or an individual. The second dimension, Task Relevance, accounts for the CWB criteria that are relevant to the tasks performed within a job.

The third and lowest hierarchical level is comprised of eleven categories of CWB behavior. These categories are themselves comprised of sixty-six more specific CWB behaviors that were identified in a survey of research literature in psychology, management, business, and sociology before being scored into qualitatively related categories (Gruys & Sackett, 2003). This model also has empirical support (Gruys & Sackett, 2003) and can be used to classify different forms of CWB, not only in research but also for practical purposes. These eleven CWB categories can be assigned to the dimensions of Interpersonal–Organizational and Task Relevance as follows:

11 Categories of Counterproductive Work Behavior (CWB)

1. Theft and Related Behavior

   For example, stealing money from the organization or from colleagues; stealing office materials; giving away services or products free of charge

2. Destruction of Property

   For example, destroying property belonging to the organization or colleagues; conscious sabotage of the manufacturing of a product belonging to the organization

3. Misuse of Information

For example, destroying or falsifying documents belonging to the organization; discussing confidential information with outsiders; deliberately withholding essential information from managers or colleagues

4. Misuse of Time and Resources

   For example, taking long breaks without permission; using the Internet for non-work-related purposes; using office equipment for private tasks

5. Unsafe Behavior

   For example, subjecting oneself, colleagues or customers to danger by failing to obey safety regulations or by not knowing what these regulations are

6. Poor Attendance

   For example, being absent from work without legitimate cause; falsely calling in sick; leaving early without permission

7. Poor Quality Work

   For example, consciously performing substandard work; performing one's duties slowly and carelessly

8. Alcohol Use

   For example, arriving at work under the influence of alcohol; consuming alcohol during working hours; performing less competently at work due to the effects of alcohol consumption

9. Drug Use

   For example, arriving at work under the influence of drugs; taking drugs during working hours; performing less competently at work due to the effects of drugs

10. Inappropriate Verbal Actions

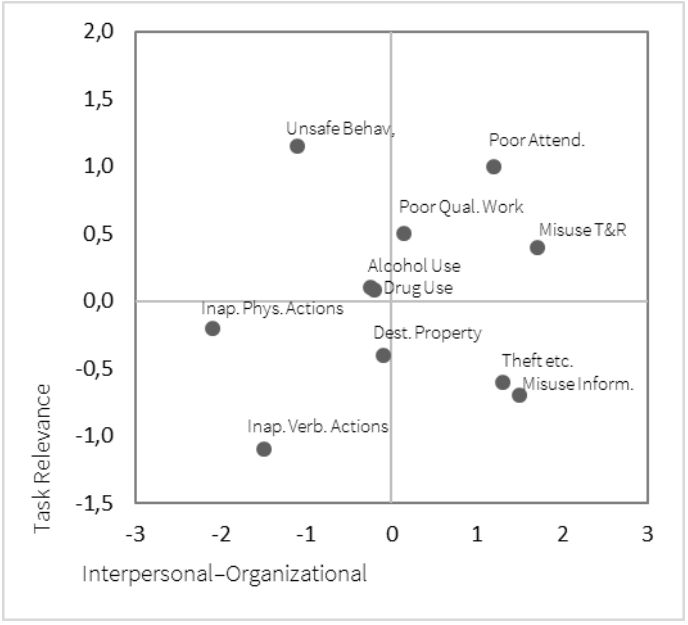
    For example, shouting at or quarreling with customers, colleagues or managers; verbally abusing customers, colleagues or managers; using sexually offensive language at work

11. Inappropriate Physical Actions

    For example, physically attacking customers, colleagues or managers; making unsolicited sexual advances to customers, colleagues or managers

Figure 1.1.

Two-dimensional scatter diagram for the CWB categories in relation to the dimensions of Task Relevance and Interpersonal–Organizational. From Gruys & Sackett (2003)



The figure shows how the 11 CWB categories are related to the dimensions Task Relevance and Interpersonal–Organizational. Categories assigned to the positive end of Task Relevance are relevant to the tasks performed in a specific job, while categories assigned to the negative end of the dimension are not. Categories assigned to the positive end of the Interpersonal–Organizational dimension are directed against the organization, while categories assigned to the negative end of the dimension are directed against individuals within the organization.

As can be seen from figure 1.1 above, the categories of Inappropriate Physical Actions, Inappropriate Verbal Actions, and Unsafe Behavior are to be found to the extreme left, or negative end, of the Interpersonal–Organizational dimension. Since the behaviors in the categories of Alcohol Use and Drug Use are not directed at other individuals or at the organization, they are in a relatively neutral position on this dimension. On the negative end of the Interpersonal–Organizational dimension are Poor Attendance, Misuse of Time and Resources, Misuse of Information, and Theft and Related Behavior.

For the Task Relevance dimension, Poor Attendance, Poor Quality of Work, and Misuse of Time and Resources are all on the positive side of the scale. The categories of Alcohol Use and Drug Use are once again to be found in a relatively neutral position in the middle of the dimension. The categories of Inappropriate Verbal Actions, Inappropriate Physical Actions, Theft and Related Behavior, Destruction of Property, and Misuse of Information are all on the negative side of this dimension scale, indicating that they are not related to work duties.

The correlations among these eleven categories confirm that there is a general CWB factor at a higher level, with the correlations between the various types of behavior ranging from .17 to .71, with an average of .43 (Gruys & Sackett, 2003). These correlations clearly correspond to those Bennett and Robinson (2000) reported between Organizational and Interpersonal Deviance (r=.41). The variation in magnitude of the correlations among different categories of CWB may be explained by the differing nature of the categories with respect to both content and severity. Low correlations show that

these categories are different, while high ones mean that they are similar.

Financial consequences of CWB

It is understandable that serious criminal behavior at work such as theft or violence has a negative impact on organizations. Some forms of CWB, like absenteeism, may seem much more benign by comparison. However, these more common forms of CWB have considerable financial consequences for organizations, and it is, for example, estimated that 20% of all U.S. businesses fail because of this. These behaviors include slow work rates, taking long breaks, socializing instead of working, and a lack of cooperation among colleagues and superiors (Greenberg, 2006).

A number of studies have attempted to determine the extent of counterproductive behavior and the costs for organizations. A study by Bennett and Robinson (2000) states that 75% of all employees have, on some occasion, stolen from their employer. Furthermore, an estimated 33% to 75% of all employees have participated in behaviors such as theft, fraud, vandalism, sabotage, and optional absenteeism. In the U.S., the direct yearly costs of workplace theft are estimated at an average of $779 per employee (Vardi & Weitz, 2004).

In 2005 Swedish retailers reported total losses of over the equivalent of $1 billion, according to The European Retail Theft Barometer[1]. Of this amount, employees reportedly stole $300 million from their employers, amounting to 29% of the total loss. The retail sector has, in addition, spent more than the equivalent of $250 million on security systems to protect against this type of behavior, bringing the total cost to $1.25 billion. It is worth noting that the total losses have not been reduced, despite the massive security expenditure, when compared with the previous year (2004). Sweden is bucking the trend in Europe, where relatively large reductions in losses due to theft are occurring.

According to The Global Retail Theft Barometer for 2014–2015, which presents shrinkage trends across 24 countries in Europe, America, and the Asia-Pacific, the main source of shrinkage (39%) is due to employee theft (www. globalretailtheftbarometer.com). It is also reported that employee theft had increased significantly (by 11 percentage points) in 2015.

Absenteeism also causes significant financial losses to organizations. In 1998, U.S. costs were estimated to be an average of $757 per employee per year (Cascio, 2000). Indirect costs, such as the expense of extra staff, are not included in this figure. The average cost for a company with a hundred employees would thus amount to $75,700. Cascio (2000) defines absenteeism as any failure to be at one's job on schedule, regardless of the reason. The expression "on schedule" is a key phrase in this context, since it excludes absences due to vacations, leave of absence, and maternity/paternity leave. Cascio points out that the confusion surrounding measures of absenteeism can be avoided by simply applying this definition. If an employee fails to appear at work according to the schedule, then it should be classified as absenteeism regardless of the cause. If the employee is not available to carry out his or her regular duties, it will probably lead to the job or tasks being performed less efficiently by a replacement or not at all. This definition

[1] The European Retail Theft Barometer has been undertaken every year since 2002 by an independent British survey consultancy, The Centre for Retail Research. A total of 440 retail chains with 24,572 outlets in twenty-five European countries took part in the 2005 survey. Sixteen retail chains with a total of 1,098 outlets took part in the Swedish survey.

of absenteeism renders other interpretations redundant. The most common reason for absenteeism reported by employees is personal illness. In fact, studies show only one in five employees are actually ill during such absence, while the remaining four are at home for other reasons. The most common real reasons for absenteeism are family-related problems – not the reasons that were reported to employers.

CWB thus presents a serious economic threat to organizations. In the US, its annual costs have been estimated at $4.2 billion for workplace violence and between $40 and $120 billion for theft. Figures between $6 and $200 billion have been ascribed to a broader spectrum of CWB (Bennett & Robinson, 2000).

Individual differences and work-related behavior

Research indicates that individual traits are important in determining the probability that an individual will engage in CWB. As these traits are stable over time, vary among individuals, and influence many behaviors relevant to work performance, they play an important part in recruitment and selection.

General mental ability (GMA) and personality are the center of attention when it comes to individual differences. The relationship between GMA and CWB has been the subject of limited study, although Dilchert, Ones, Davis, and Rostow (2007) obtained a relatively strong negative correlation between GMA and CWB, suggesting that the greater GMA an individual has the lower the probability engaging in CWB. Research literature is more extensive when it comes to personality and CWB.

Personality

A few years ago the general conclusion was that there are few personality differences between individuals involved in CWB and those who are not. Recent research, however, has presented a different picture. Most personality research is based on the Five-Factor Model (FFM), a taxonomy of five broad personality dimensions: Emotional Stability, Extraversion, Openness, Agreeableness, and Conscientiousness.

The FFM is a generally accepted framework for describing personality; in conjunction with the development of meta-analysis[2], it has contributed to considerable progress within several research areas.

Personality and job performance

Extensive research has been devoted to the importance of personality, defined by the FFM, and work-related behavior such as job performance. Research has focused on determining whether or not measures of personality (test scores) can predict job performance. The current consensus is that personality can in fact do so and that some personality dimensions have greater predictive value than others (Barrick & Mount, 2003).

Research has provided rather clear estimates, while the development of the meta-analytical approach has allowed individual studies to be summarized and re-analyzed,

[2] Meta-analysis is a statistical method that uses the results from a number of mutually independent studies, as in this case, with regard to the connection between FFM dimensions and different types of work-related behavior. When the results of several studies, where the same phenomenon has been studied, are brought together in a meta-analysis, the influence of chance on the results is reduced, thereby also decreasing the risk of drawing erroneous conclusions.

MiNT

MEASURING INTEGRITY

TECHNICAL MANUAL

2017 EDITION

SOFIA SJÖBERG
ANDERS SJÖBERG

ASSESSIO

thus providing reliable generalizations. In what is termed as second order meta-analysis, i.e., a meta-analysis of meta-analyses, Barrick and Mount (2003) examined the studies conducted within this area and summarized the correlations between the FFM dimensions and job performance. Their results show that Conscientiousness and, to a lesser extent, Emotional Stability are the two personality dimensions that predict job performance, regardless of job type and job performance measurement criteria.

Individuals who achieve high scores on Conscientiousness are hard working, well organized, and performance oriented and thus generally perform better at work than those lacking these qualities, regardless of the type of job. In a similar way, those who score low on Emotional Stability, irritable and stress-prone individuals with low self-esteem, will in general have lower performance than mentally stable and adaptable people with high self-esteem, regardless of the job.

The pattern is somewhat different for the other personality dimensions. In another study by Barrick, Mount, and Judge (2001) it was shown that the dimensions of Agreeableness, Extraversion, and Openness are also valuable predictors of job performance, but only when it comes to certain types of jobs. For example, it has been shown that Extraversion predicts high performance for jobs involving intensive teamwork but poor performance for jobs where negotiations play an important part. Extraversion also predicts job performance for jobs which require considerable interaction with others (such as sales and management jobs), while having no influence whatsoever for other types of jobs.

Those who score high on the Agreeableness dimension are predicted to have high performance in jobs involving teamwork but low performance in jobs where negotiations play an important part. These three factors are therefore not universal predictors for all jobs and criteria. For further details, refer to Barrick, Mount, and Judge (2001).

Personality and CWB

As our understanding of the important relationship between personality and desirable work-related behaviors has improved, attention has turned to negative, counterproductive behaviors. Job performance as a criterion may to some extent be regarded as the positive opposite of CWB since some aspect of poor job performance overlap with some forms of CWB. This means that an employee who is often absent, works slowly, or produces poor quality work is unlikely to produce consistently good job performance.

There is also compelling research (Dalal, 2005) showing that the correlation between CWB and Organizational Citizenship Behavior (OCB)[3] is not particularly strong (p=.32). This research suggests that these two criteria are not direct opposites, and should in fact be treated as two separate concepts. This conclusion is reinforced by the fact that CWB and OCB show different patterns of relationships when the correlations with common variables such as the FFM factors are studied. Thus, it is not possible to merely change the direction and magnitude of one or multiple personality dimensions in order to predict CWB as opposed to job performance. This would have been possible if the correlation between CWB and OCB had been –1.00. In other words, a high degree of OCB does not

3 The extent to which an employee regards him/herself as part of the organization, identifies with it, and works on the basis of its values and goals.

predict a corresponding low degree of CWB or vice-versa. These two concepts are subject to more complicated patterns that deserve further research.

The significance of these findings for practical selection situations is therefore that, when excluding applicants on the basis of predicted CWB tendencies, those with strong OCB indications should not be viewed as safe in this respect, since even good “organizational citizens” may have tendencies toward CWB.

Table 1.1 shows the results from two studies investigating the relationship between the FFM dimensions and job performance as measured by OCB and CWB. In the first study, Barrick and Mount (2003) demonstrate the relationships between the individual FFM factors and job performance, while, in the second study, Sackett, Berry, Wiemann, and Lacso (2006) present results regarding the associations between FFM factors and both OCB and CWB. The results clearly show that certain FFM dimensions predict CWB and OCB even better than they predict job performance.

As can be seen from table 1.1, the results of the second study in various ways support the hypothesis that OCB and CWB are not direct opposites on the same continuum. This is revealed by the associations with the FFM dimensions, which to varying degrees correlate with both OCB and CWB.

The results of research focusing on CWB have clearly pointed in a single direction. Three FFM dimensions were found to more accurately predict CWB than the other dimensions (Sackett & Waneck, 1996; Kolz, 1999): Conscientiousness, Agreeableness, and Emotional Stability. The first of these dimensions has often been used to explain most of the variation in CWB. The results from Sackett et al. (2006) confirm this and also correspond to the results presented by Ones (1993), which are used to define the concept of Integrity and its connection to CWB.

Table 1.1
Correlations between FFM dimensions and job performance, OCB, and CWB

FFM dimension	Job performance (1)	OCB (2)	CWB (2)
Conscientiousness	.22	.14*	-.31*
Emotional Stability	.12	.04	-.21*
Agreeableness	.07	.24*	-.17*
Extraversion	.12	.15*	.01
Openness	.05	.23*	-.02

Note: *p<.01; Corrected for criterion reliability and restriction of range in the FFM dimensions; (1) From Barrick & Mount (2003) in Validity Generalization – A Critical Review, Kevin Murphy (ed.), page 207; (2) From Sackett, Berry, Wiemann & Laczo (2006). Citizenship and Counterproductive Behavior: Clarifying Relations Between the Two Domains. Human Performance, 19(4), 441–464.

Summary

Counterproductive work behavior (CWB) is generally defined as intentional employee behavior that is contrary to the interests of the organization. Over recent decades a CWB model has emerged based on a three-level hierarchy, with a general CWB factor on top. Eleven categories of CWB, comprising the lowest level, each fall along the scales of the two dimensions that comprise the middle level, Interpersonal–Organizational and Task Relevance.

The financial consequences of CWB have been substantial, with findings indicating that some forms of CWB are very common with extensive losses for organizations across industries.

Research has outlined that certain personality traits may contribute to the prediction of CWB. Through meta-analysis studies, three dimensions of the Five-Factor Model (FFM), Conscientiousness, Agreeableness, and Emotional Stability are reliable predictors of CWB.

Part 2

Integrity

This part begins with a definition of the concept of integrity and provides an introductory overview of the development of integrity tests and, in particular, personality-based integrity tests. The operationalization of the concept of integrity on the basis of three personality dimensions that make up the basis of MINT is described in the final section.

The construct of integrity

Terms such as integrity and honesty are often used synonymously, but there are conceptual differences that are worth clarifying. Honesty is another term for truthfulness, while integrity (and dependability) involves a broader concept covering "A willingness to follow rules, internalized values, norms and expectations" (Murphy, 2000). Integrity tests therefore aim at measuring the presence of these tendencies.

The development of integrity tests

In the development of integrity tests, the focus has been on the overall conceptualization of integrity and on measuring behavior related to the concept. Historically, integrity tests have not been developed on the basis of carefully defining the concept and its relationship to specific issues (in accordance with what is termed the construct-oriented method of test development), which led to confusion over what such tests were really measuring. This also limited the possibilities of systematically studying and identifying the relationships between test scores and relevant criteria, as the instruments themselves contained a number of different types of questions that failed to cover all of the aspects of the concept.

The methodology for measuring integrity was and in some cases still is relatively heterogeneous, requiring a careful analysis of each method and the way it is developed in relation to its theoretical basis. One serious consequence of having a range of methodologies for measuring integrity is that there is no frame of reference for interpretation and feedback based on a coherent latent theory. During the selection process, a candidate may be rejected because of an unsatisfactory score on an integrity scale, but it is difficult to inform the candidate as to why the score is unsatisfactory when it is not entirely clear what the scale is in fact measuring.

Previous studies of integrity tests provided somewhat disappointing results. The progress of personality research and the development of methods such as meta-analysis, however, contributed to a more correct evaluation of these methods. Integrity tests was found to have a good internal consistency and stability over time (reliability[4]) as well as having stable relationships to several criteria that are relevant to working life.

The relationships between test results and work-related behaviors, including CWB, have

4 Reliability refers to the stability of the test in terms of the results not being affected randomly, that they can be expected to remain stable over time, and that an individual will obtain similar results in different test sessions.

been successfully established by examining different situations in repeated studies (Ones, Viswesvaran & Schmidt, 1993). The predictive power of integrity tests with respect to general CWB has been estimated to .47 (Ones, Viswesvaran & Schmidt, 1993). The validity[5] of different tests may differ somewhat, depending on whether it is an overt or personality-based test; on the type of criteria being used, such as ratings of job satisfaction or productivity; and on the complexity of the tasks involved. In practice, validity tends to be higher when more complex tasks are involved. Most integrity tests can predict both job performance and CWB, although it should be stressed that this refers to integrity tests that are sound, carefully developed, and of good quality.

The success of integrity testing in predicting a number of work-related criteria naturally prompts the question of why these tests work. Considerable efforts have been made to identify the personality dimensions underlying integrity tests (Ones, 1993; Murphy, 2001), which will be discussed in greater detail in the next section.

The development of personality-based integrity tests

At present, traditional personality tests and other measures of normal adult personality are often constructed for the purpose of providing broad measures of personality, and general instruments have been created which can be applied in a range of situations.

However, research has shown that it is possible to increase predictive power by developing tailor-made tests targeting specific criteria, which is achieved by combining different personality dimensions (Ones, Viswesvaran & Dilchert, 2005).

The dynamics of personality dimensions are often dependent on the selected criteria, and the total combined effect of the dimensions involved will be greater than the sum effect of the single dimensions. The rationale behind the special design of measures of personality in a work context (Ones, Viswesvaran & Dilchert, 2005; Sjöberg, Sjöberg & Forssén, 2005) is based on several considerations:

- The structure of personality is complicated; it is hierarchic, and there are primary and secondary relationships — at different levels
- There is an imbalance in the hierarchy; some dimensions function at a higher level of abstraction. Emotional Stability and Conscientiousness, in particular, function at a higher level in relation to many criteria, compared to the other dimensions
- The personality dimensions of FFM are not orthogonal (independent), they overlap to varying degrees
- Certain personality traits explain more variance than others when it comes to specific criteria such as job performance

Combining several personality dimensions in a composite measure makes it possible to develop scales with greater predictive value for specific criteria as compared to using the FFM dimensions. Scales of this type are called Criterion-Oriented Personality Scales (COPS).

5 Validity refers to evidence that theory and empiricism support the use of the test scores for correct decision-making.

The predictive value of these scales has been shown to be considerably higher than most other measures of individual differences, and only GMA is higher (Ones & Viswesvaran, 2001). COPS are thus constructed to predict individual differences regarding specific work-related behaviors, while the composition of the scale and its contents may of course vary depending on the criterion.

When it comes to CWB-classified criteria, such as violence, aggression, stress, drug and alcohol abuse, and dishonesty according to the FFM, it can be established that they all rest on common underlying concepts. In CWB-related COPS, 70% to 100% of the variance derives from the three dimensions of Conscientiousness, Agreeableness, and Emotional Stability (Ones, Viswesvaran & Dilchert, 2005), which indicates that these dimensions together are powerful measures with respect to CWB-classified criteria. Therefore, these three dimensions are interesting in regard to method development. The three dimensions are also summarized in Digman's article from 2005, "Higher-Order Factors of the Big Five," where he presents the results of a meta-analysis using correlation matrices from fourteen different studies of various situations. The factor analyses of these studies support the theory that there is a higher hierarchic level above the FFM dimensions which consists of two factors: Factor Alpha and Factor Beta. Factor Alpha is a broad theoretical concept that encompasses the dimensions of Conscientiousness, Agreeableness, and Emotional Stability. A concept that, according to one explanation, may be described as representing the socialization process in the personal development of the individual.

In other words, based on the research described above, the three dimensions, Conscientiousness, Agreeableness, and Emotional Stability, should be the basis for measuring integrity. These dimensions, presented here in order of importance, explain the greatest proportion of the variance in the concept of integrity.

Operationalizing the construct of integrity

In MINT the concept of integrity is operationalized according to the personality dimensions of Conscientiousness, Agreeableness, and Emotional Stability. To further clarify these dimensions, a description of the contrasting characteristics that distinguish high and low levels is provided.

Conscientiousness

High Level
A person with a high level of Conscientiousness is someone who in general is well organized, dependable, and performance-oriented.

Conscientiousness

Low Level
A person with a low level of Conscientiousness usually has a more relaxed attitude to performance, is spontaneous, and acts on the spur of the moment.

Agreeableness

High Level
People with a high level of Agreeableness are often pleasant, easy to get along with, and always believe that others are inherently good.

Agreeableness

Low Level
People with a low level of Agreeableness are in general independent and straightforward, and are often on their guard against others.

Emotional Stability

High Level
People with high Emotional Stability are often calm, confident, and satisfied.

Emotional Stability

Low Level
People with low Emotional Stability are nervous rather than calm, doubt their own ability, and are worried about the future

Integrity, CWB, and job performance

Research has concluded that a measure of integrity combined with a measure of GMA provide the highest validity for predicting job performance (Schmidt & Hunter, 1998). In addition, integrity predicts CWB in a convincing manner. In fact, it predicts CWB even better than it predicts job performance.

The most extensive meta-analysis to date regarding the predictive validity of integrity tests for job performance and CWB was undertaken by Ones, Viswesvaran, and Schmidt (1993). Their results showed that integrity tests have an operational validity of .34 for predicting job performance, while the corresponding figure for CWB is .47. When the effect of Conscientiousness is controlled for, the correlation between integrity and job performance is reduced to .27, which means that Conscientiousness alone does not explain this relationship. Another finding that confirms this is that integrity and Conscientiousness together predicted job performance with r=.36, and when Conscientiousness was removed, the figure decreased, albeit only to r=.34 (Murphy, 2001).

Job performance and CWB are not opposites, nor are they independent of each other; it should therefore come as no surprise that integrity tests tend to predict both these criteria. This increases the potential field of application, since not only can integrity predict CWB but it also helps to identify individuals who are likely to achieve high job performance.

Summary

Integrity refers to a willingness to abide by the rules, follow norms and expectations, and live by internalized values. While the usefulness of integrity tests of the past suffered from poor concept integration, quality tests of today are concept-driven and more reliable.

A combination of personality dimensions, Conscientiousness, Agreeableness, and Emotional Stability, has been found to effectively reflect integrity. These three dimensions have been operationalized in integrity tests, including MINT, which can predict both CWB and job performance, criteria which are neither opposites nor independent of each other.

Part 3

The initial development of MINT

This section describes the construction and initial development of MINT. This work was conducted on the Swedish language version and with Swedish respondents for the empirical analyses. As mentioned, MINT is based on personality theory from the FFM, where the three dimensions of Conscientiousness, Agreeableness, and Emotional Stability constitute both a general integrity factor and two subfactors.

In the development of MINT, it was assumed that there is an imbalance among the three FFM dimensions, with Emotional Stability (Neuroticism reversed) being at a higher level than Conscientiousness and Agreeableness. Furthermore, Emotional Stability has been assumed to account for the majority of the variance in integrity[6].

> *[I]t becomes evident that superordinate personality hierarchy is an unbalanced hierarchy. The current results indicate […] that Neuroticism in particular exists at a different level of abstraction than the other Big Five traits. (Markon, Krueger, Watson, 2005; p 152)*

This does not mean that Conscientiousness and Agreeableness have a less important role to play, only that these two factors must be isolated from the higher-level Emotional Stability before they can be interpreted at the individual level. The underlying model for MINT, derived from Markon, Krueger, and Watson (2005), therefore places Emotional Stability at the top of the hierarchy and Conscientiousness and Agreeableness as subfactors.

Item construction

The items included in MINT have been constructed and selected by taking both theoretical and empirical considerations into account. In total, 206 items were formulated in Swedish, with 50 items representing each of the FFM dimensions of Emotional Stability, Extraversion, and Agreeableness, together with 56 items aimed at measuring Conscientiousness. The construction of items has a dual purpose: (1) to create an instrument for the purpose of measuring four of the five FFM dimensions, therefore items measuring Extraversion was included (PJP; Predicting Job Performance, Sjöberg, Sjöberg, & Forssén, 2006) and (2) to formulate and identify items measuring integrity.

The items were, together with the Swedish version of the NEO PI-R (Costa & McRae, 1985), distributed to a national representative sample of the Swedish population (n=650). After one reminder, 300 completed questionnaires were returned in total. See table 3.1 for descriptive statistics comparing the sample with the Swedish population regarding age, gender and educational level.

Before the item analysis process for MINT commenced, a total of 102 items were

6 Note that integrity with a lowercase letter refer to the construct of integrity, while Integrity with a capital letter refer to the scale (the measure) in MINT.

excluded. All of the Extraversion items were removed, as they were not part of the theoretical model of integrity (Ones, Viswesvaran & Dilchert, 2005). Twenty-three items measuring Conscientiousness were removed as well as 15 items from Agreeableness and 14 from Emotional Stability since they had already been utilized in the development of PJP. Thus, in total 104 items remained for further data processing.

Table 3.1

The distribution of the initial MINT norm group (n=300) by age, sex and educational level compared with the population (www.scb.se)

Age	Norm group	Population
18-30	23%	21%
31-40	19%	19%
41-50	22%	18%
51-60	20%	19%
61-70	9%	13%
71-80	7%	10%
Sex	Norm group	Population
Female	59%	50%
Male	41%	50%
Educational level	Norm group	Population
Primary	19%	25%
Lower secondary	17%	26%
Upper secondary	18%	19%
Post-secondary studies	16%	13%
University degree	29%	16%
Doctoral degree	1%	1%

A number of empirical and theoretical requirements were formulated before the final choice of items. These included:

1. Item content should reflect the factor in question.
2. Each item should correlate most highly with the scale to which it belongs.
3. Item content should not overlap theoretically.
4. The test should be relatively short; each factor should be represented by 20 items.
5. The reliability (internal consistency) should be greater than .90 for each factor and for the general Integrity factor.

In order to determine whether or not the content of each item reflected the dimension in question, a qualitative comparison was made against other items and measures reflecting the corresponding constructs. The comparisons were made against the three FFM dimensions that comprise the components of the concept of integrity as well as against the corresponding dimensions from the Swedish versions of the Hogan Personality Inventory (HPI; Hogan & Hogan, 1997a) and the Hogan Descriptive Scales (HDS; Hogan & Hogan, 1997b).

The next step was to calculate the correlations between each item and the dimension on

the NEO PI-R personality inventory to which the item was theoretically considered to belong. In total, 22 items were excluded because their correlation with the corresponding scale was too weak: three items from Conscientiousness, six from Agreeableness and 13 from Emotional Stability. This left 30, 29, and 23 items respectively after this step.

Items which correlated more highly with one of the other two scales than they did with the intended one were also excluded: seven items from Conscientiousness and four from Agreeableness. This then left 23, 25, and 23 items respectively for the three dimensions.

Further items were then removed for each dimension to reduce the totals to the maximum target limit of 20 items per scale and also to avoid items with overlapping content. This was done using Item Response Theory (IRT). A one-parameter Rasch-model was used to exclude items that were close to each other in terms of difficulty. The purpose of the analysis was to obtain a broader measure of each factor and to exclude items that discriminated at the same scale level. The analytical tool RUMM 2020 was used for the estimation, and the remaining 60 items (20 per scale) were then used for scoring the dimensions that comprise MINT.

Integrity – a hierarchic measurement model

The theoretical model for MINT postulates that integrity is a higher-order construct and that the three underlying personality dimensions do not contribute equally to the overall measure of Integrity. (Ones & Viswesvaran, 1998; Digman, 1997). The model postulates that the three FFM dimensions are weighted with varying degrees of influence, with that Emotional Stability acting as a moderator variable for the other two dimensions (Digman, 1997). In addition, the two independent factors of Interpersonal Orientation and Task Orientation are at the level below the Integrity measure. And, as well as being independent of each other, the two factors are also independent of the higher level Integrity measure.

Confirmatory Factor Analysis

In order to test the underlying theoretical structure in MINT, a hierarchic factor analysis, Nested Factor Analysis (Gustafsson, 2002), was carried out, and the AMOS 6.0 analysis tool (Arbuckle & Wothke, 1999) was used for the estimation. The hierarchic measurement model, recently used more commonly in intelligence research (Mårdberg, Sjöberg, & Henrysson Eidvall, 2000), is based on the identification of a higher-level general factor and a number of residual factors. In this way, the method generates factors that can be measured independently. This is essential for correct profile interpretation but very uncommon in personality research and measurement.

When several dimensions overlap in personality measurement, it is impossible to isolate the separate components when the personality scales are only summated and compared. For MINT, the hierarchic model utilized the Integrity factor as the general factor and Interpersonal Orientation and Task Orientation as the residual factors, producing an orthogonal solution where the factors are zero-correlated. The precision of profile interpretations increases considerably when factors are not related empirically.

In order to identify the hierarchic measurement model, the theoretical structure of MINT was operationalized using two indicators for each latent variable or personality

dimension (Gustafsson, 2002). Two comparable indicators (the split-half method) were therefore created initially for each of the three FFM dimensions, based on factor loadings from three principle component analyses, one each for Emotional Stability (ES), Agreeableness (A) and Conscientiousness (C). The aim was to create six indicators, designated ES1, ES2, A1, A2, C1, and C2.

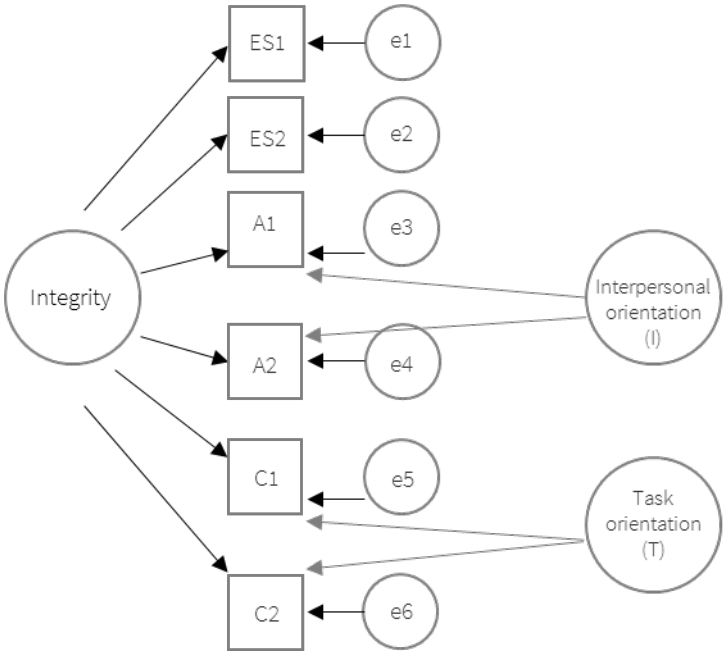
To illustrate the procedure, let us examine Emotional Stability. The item which had the highest loading in this one-factor solution was grouped together with the item which had the lowest factor loading in the same one-factor solution, and this was designated ES1. The item with the second highest loading was grouped together with the one with the second lowest loading, and this was designated ES2. This procedure was repeated until all of the items had been assigned to the indicators ES1 or ES2. This generated two parallel indicators of ES that were then examined using confirmatory factor analysis. These indicators were only generated for statistical, technical purposes; they are not part of the interpretation of test results and have no theoretical implications.

The starting point of the analysis was to assign the six indicators directly to the general factor of Integrity. In addition, two pairs of indicators, A1, A2 and C1, C2, were assigned respectively to the lower-level factors of Interpersonal Orientation and Task Orientation. The measurement model is shown in Figure 3.1 where e represents error.

The value of the results from a factor analysis can be evaluated using different measures of fit. One statistical test used for this purpose is chi square (χ^2). The correlations in the assumed model are compared with the correlations in the empirical data; a significant value means that there is a discrepancy between the theoretical model for MINT and the result of the factor analysis.

Chi square values are, however, too conservative and unrealistic measures of fit between model and data, especially for large samples. Other measures of fit, such as Root Mean Square Error of Approximation (RMSEA), Goodness-of-Fit Index (GFI), Tucker Lewis Index (TLI), and Comparative Fit Index (CFI), have therefore been proposed as alternatives.

Figure 3.1. The measurement model of MINT



The analysis performed shows a good fit between model and data (χ^2, n er li it o has een roposed for RMSEA as an acceptable fit; for the other measures the corresponding figure for a lower limit is .90 (Kelloway, 1998); the higher value, the better fit. The factor loadings are shown in table 3.2.

Table 3.2.
Standardized factor loadings for the model (n=302)

	Integrity	Interpersonal Orientation	Task Orientation
ES1	.91		
ES2	.89		
A1	.35	.78	
A2	.33	.81	
C1	.40		.79
C2	.42		.82

Note: ES=Emotional stability; C=Conscientiousness; A=Agreeableness.

The result supports the hypothesis that the Emotional Stability dimension is a major component in the Integrity measure, as its indicators show loadings on the Integrity factor that are more than twice as high as those on Agreeableness and Conscientiousness. The non-standardized factor loadings on the scales are used to calculate the individual scores, e.g., factor scores. These individual factor scores are calculated by applying regression analysis and by multiplying the regression weights by the sum of the raw scores on the factors. This is essential if the individual scores are to be interpreted using the proposed hierarchic model for MINT. Descriptive statistics for each factor are shown in table 3.3.

Table 3.3
Descriptive statistics for the MINT factors

	Min	Max	M	Md	SD	IC
Integrity	15.35	43.13	33.90	33.85	4.81	.98
Interpersonal Orientation	5.39	23.74	16.86	17.02	3.01	.93
Task Orientation	7.20	25.95	18.74	18.51	2.81	.94

Note: IC=Internal consistency

In order to investigate the underlying structure of MINT in relation to the dimensions in the FFM, the correlations between the FFM dimensions and Conscientiousness, Agreeableness and Emotional Stability in MINT has been calculated. NEO PI-R (Costa & McRae, 1985) has been used as an external measure of the three dimensions; see table 3.4.

Table 3.4
Correlations between MINT and NEO PI-R

	Conscientiousness	Agreeableness	Emotional Stability
Integrity	.41*	.27*	.76*
Interpersonal Orientation	-.10	.65*	-.01
Task Orientation	.59*	-.11	.01

Note: *=P<.01; n=302

The general measure of Integrity is linked to all three dimensions, with Emotional Stability showing the strongest correlation, as expected. Individuals who score high on Integrity tend to be emotionally stable, dependable, and show empathy. The two lower-level factors were each found to have a strong link to one particular dimension: Interpersonal Orientation to Agreeableness, and Task Orientation to Conscientiousness. Neither of these lower-level factors correlated significantly with Emotional Stability nor with the marker for the other dimension. The results thus show, in an expected and convincing fashion, that MINT is linked to the factors in NEO-PI-R.

Evidence of reliability

Reliability concerns the precision of the measurement method. Reliability is based on the evidence that the interpretations based on the results are consistent. Theoretically speaking, a completely consistent and reliable method would contain no measurement error, a situation that never arises in practice. Estimating and being aware of potential measurement errors is essential when evaluating the results for reliability deficiencies. There are a number of different kinds of reliability and numerous ways to estimate the reliability of a measurement; for example, consistency over time, consistency among different raters, and the internal consistency of items or scales all reflect reliability. There is no universally best reliability estimate, as the best approach depends on the context; depending on the decisions being made based on the results, the consequences of different types and levels of reliability deficiencies have varying degrees of relevance and importance.

Internal consistency

Internal consistency is one measure of reliability. The most common method for estimating it is to calculate Cronbach's alpha (**α**). Cronbach's alpha may however only generate an objective reliability measure if the loadings on the common factor are equal. This is very rare in practice. When the scales measure a common factor but the factor loadings (and the error variance) are heterogeneous, the model is called congeneric. This model requires a different approach in order to obtain a correct estimate of internal consistency. If the traditional Cronbach's alpha were calculated for this model, it would generate an underestimation of the internal consistency of the general measure, and alpha in itself is an underestimation of reliability (Reuterberg & Gustafsson, 1992).

When a congeneric model is applied to internal consistency, it is possible to account for both the variance differences in the scale factors and the error variance inherent in each factor. For testing internal consistency in MINT, the reliability of the factors was calculated in conjunction with the confirmatory factor analysis.

By summing all of the true variance and covariance and dividing this by the total variance,

we estimated internal consistency according to the following equation (Reuterberg, & Gustafsson, 1992).

$$\rho_{xx} = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_{ii}} = \frac{proportion\ of\ true\ variance}{proportion\ of\ true\ variance + error\ variance}$$

This yielded an internal consistency of .98 for the general measure of Integrity and .93 and .94, respectively, for the lower-level factors of Interpersonal Orientation and Task Orientation. These results support the internal consistency of all the MINT factors and clearly satisfy the requirements for the accuracy of the measurement.

Test–retest

A test–retest study was conducted (n=97) with students from the Department of Psychology at Stockholm University (72 women and 25 men). All of the students were in the first semester of their Master's Program. The average age of the group was 23 years (SD=5). There were 30 days between the testing sessions. In table 3.5, the test–retest values (r and t) are presented for all of the factors in MINT. The results indicate that these factors were stable for the 30 days and that they correlate highly and significantly and show low, non-significant t-values.

Table 3.5

Test–retest values for factors in MINT (Swedish sample, n=97)

MINT	rtt	t
Integrity	.85*	.10
Interpersonal Orientation	.80*	.77
Task Orientation	.87*	.69

Note: *p<.05; rtt=correlation between the two test occasions; t=paired t-test

An additional test–retest study was conducted in 2012 using the Norwegian version on a sample of n=129 respondents (91 women and 38 men) working in a grocery store chain in Norway. The average age of the sample was 35 years (SD=12); the youngest respondent was 17 years and the oldest 65 years. Table 3.6 presents the results regarding the Relative (r) and Absolute (t-values) stability for Integrity, Interpersonal Orientation, and Task Orientation during the eight-week period.

Table 3.6

Test–retest values for factors in MINT (Norwegian sample, n=129)

MINT	rtt	t
Integrity	.88**	-1.73
Interpersonal Orientation	.77**	-1.24
Task Orientation	.75**	-.37

Note: **p<.01; rtt=correlation between the two samples; t=paired t-test

Part 4

Validity – Evidence based on relations to other variables

The validity of MINT rests on its theory and empirical data supporting the usefulness of the approach for making correct personnel selection decisions based on the probability of CWB. The validation process thus involves collecting evidence supporting the accuracy of interpretations made and the conclusions drawn from the results. In the process of evaluating the extent to which MINT is appropriate for use in personnel selection, correlations with other measures that are vital in assessing job performance, including CWB, have been calculated.

Minnesota Multiphasic Personality Inventory-2 (MMPI-2), Inventory of Interpersonal Problems (IIP), and job involvement

When selecting relevant measures for this purpose, both theoretical and empirical studies have been examined. A number of different behaviors can contribute to poor individual job performance. Obvious counterproductive behaviors such as theft, disloyalty, sabotage, non-permitted absence and violence can have both psychological and financial consequences for companies and organizations. This creates difficulties when collecting data for occurrences of this type; it is not only complicated to obtain access to such data for both the individual and group level but also ethical problems may arise when collecting the data. It is, however, possible to obtain indications of this type of behavior via self-assessment questionnaires, based on the assumption that certain personality traits increase the risk of CWB. For the purpose of validating MINT, data were collected from other well-validated tests and scales that measure relevant behavioral tendencies which theoretically have a negative influence on individual job performance. The data was obtained from Assessio's Swedish database for psychological tests and other relevant measures. Several scales were selected that reflect general measures of CWB and which were hypothesized to correlate negatively with MINT scores.

The first measure, Demoralization, was taken from the MMPI-2 (Minnesota Multiphasic Personality Inventory-2, Hathaway & McKinley, 1989). Demoralization consists of twenty-four items and is in fact the basis of the interpretation of all scales on the MMPI-2. The scale indicates the overall general discomfort reported by the individual. An individual with a high score on this scale will probably describe him/herself as being downhearted, hesitant, and pessimistic and having low self-confidence (i.e., low Locus of Control). Demoralized people expect to fail, or see themselves as having failed, in various aspects of their lives. Those who are seriously demoralized may experience considerable emotional discomfort and a feeling of helplessness – often feeling overwhelmed by and incapable of dealing with the situations in which they find themselves.

The other instrument containing scales that were assumed to correlate negatively with MINT scores is the Inventory of Interpersonal Problems, IIP (Horowitz, Alden, Wiggins, & Pincus, 2000). IIP represents the most apparent individual interpersonal problems with

measures on eight scales: Domineering/Controlling, Vindictive/Self-centered, Cold/Distant, Socially Inhibited, Nonassertive, Overly Accommodating, Self-Sacrificing, and Intrusive/Needy.

All of these scales may be expected to correlate with MINT scores, but the general measure was focused on and considered as a summation of the degree of all interpersonal problems; the higher the score, the greater the difficulties are expected to be experienced in their interpersonal interactions (including at the workplace), although these interpersonal problems may vary in character.

A common feature of the demoralization and Interpersonal problems measures is that they are general measures of individual personality, where extreme scores may be regarded as indicating pathological tendencies and proneness to negative behaviors that are undesirable at the workplace. Both of these measures were therefore assumed to have a negative correlation with the overall Integrity factor in MINT.

Two scales were selected to test the hypothesis that the lower-level factors in MINT, Interpersonal Orientation and Task Orientation, contribute to the explanation of various behaviors. The first of these, the Antisocial Behavior scale, was also taken from the MMPI-2. It contains 22 items that aim to measure different types of antisocial behavior with respect to the past and present. Individuals with high scores on this scale are more likely to engage in different types of antisocial behavior and behave aggressively towards others, and they will likely be perceived as antagonistic, angry, and argumentative. It is also probable that these people are more likely to lie and steal. It is often difficult for them to conform to social norms and expectations, and they may thus also find it difficult to obey rules and regulations.

There is also an increased risk that they will be involved in substance abuse and sexual impropriety. It is common for those who score high on this scale to experience conflict-ridden family relationships and to have a history of substandard performance. The operationalization of Antisocial Behavior is focused on general problematic behavior such as theft and aggression and on relationships with strong conflict. The scale was for this reason assumed to correlate negatively with Integrity and Interpersonal orientation.

The other scale, Job Involvement, which was assumed to correlate positively with Integrity and Task Orientation, was taken from research on work-related attitudes. Consisting of six items, the scale was translated into Swedish and validated for Swedish conditions (Sjöberg & Sverke, 2000; Sverke, Hellgren, Näswall, Chirumbolo, De Witte & Goslinga, 2004). Job involvement, in this context, is defined as “an individual’s identification with a particular job” (Kanungo, 1982, p.342). Individuals with a high degree of job involvement see their work as “an important part of his [their] self-concept” (Lawler & Hall, 1970, p. 311) and feel that the work by definition “defines one’s self-concept in a major way” (Kanungo, 1982, p. 82). Job involvement has been shown to have a negative correlation with absenteeism (e.g., Farrell & Stamm, 1988; Shore, Newton, & Thornton, 1990; Scott & McClellan, 1990), and with the intention to leave (e.g., Baba & Jamal, 1991; Huselid & Day, 1991; Ingram, Lee, & Lucas, 1991; Shore, Newton, & Thornton, 1990). Job involvement is assumed to have a positive correlation with Integrity, as research has shown that people with high scores on this scale are less frequently absent from their job and less inclined to leave when compared with those who have low scores. Job involvement is also assumed to correlate positively with Task Orientation, as people with

high scores on this scale tend to identify with their work and regard the duties that they perform as a central part of their lives.

All of the hypotheses concerning the relations between MINT and other measures are summarized in table 4.1, and the results are presented in table 4.2.

Table 4.1
Hypotheses for the relationships between MINT and other measures of both deviant personality and job involvement

	MINT		
Scale	Integrity	Interpersonal Orientation	Task Orientation
Demoralization (MMPI-2)	-		
Interpersonal Problems (IIP)	-		
Antisocial Behavior (MMPI-2)	-	-	
Job Involvement (JI Scale)	+		+

Note: + and - indicate the assumed direction of the covariation

Table 4.2
Results for the relationships between MINT and other measures of both deviant personality and job involvement

	MINT		
Scale	Integrity	Interpersonal Orientation	Task Orientation
Demoralization (MMPI-2)	-.71*	-.04	.08
Interpersonal Problems (IIP)	-.51*	-.03	.12
Antisocial Behavior (MMPI-2)	-.28*	-.45*	-.02
Job Involvement (JI Scale)	.15*	-.12	.21*

Note: *p<.05

The results in table 4.2 support the hypotheses shown in table 4.1. The correlation between Integrity and the Demoralization scale can be regarded as high. This scale is used as a basis for interpreting all the other scales in MMPI-2; this is in accordance with the theoretical assumption that neuroticism, i.e., the lack of emotional stability, functions as a moderator for most personality dimensions. This is reflected in the high correlation with Integrity. This result shows, as expected, that an individual with a high score on Integrity is less likely to be downhearted, hesitant, or to exhibit low self-confidence than those who have low Integrity scores.

The correlation between Integrity and Interpersonal Problems provides additional support for the interpretation of MINT. It suggests a satisfactory range in the general Integrity measure; the higher the score, the fewer interpersonal problems.

As is shown in table 4.2, Integrity shows a significant negative correlation with the Antisocial Behavior scale. This implies that the higher the score an individual obtains on Integrity, the less probability there is of he or she behaving aggressively towards others and engaging in other behaviors of this type.

One of the lower-level factors in MINT, Interpersonal Orientation, also shows a strong

negative correlation with the scale for Antisocial Behavior, while the other lower-level factor, Task Orientation, shows no correlation with this scale.

This pattern of correlations is to be expected; the Interpersonal Orientation factor is intended to indicate how well a person is able to adapt to situations that involve other people, while Antisocial Behavior focuses on failures in social interaction. The results support the interpretation that low scores on Integrity in MINT together with low scores on Interpersonal Orientation increase the probability that the individual will show deviant social behavior at the workplace.

The correlations with Job Involvement provide further support for MINT, although the estimates are not as high. Integrity shows a significant positive correlation with Job Involvement, implying that people with high scores on Integrity are more likely to report higher Job Involvement compared to those with lower scores on Integrity.

The hypotheses for the two lower-level factors suggesting that Task Orientation should correlate positively with Job Involvement was confirmed. Individuals who identify with their job and see it as an important part of their self-image perform their duties better than those who do not regard their work in this way.

Hogan Development Survey (HDS)

To further investigate the validity of MINT, correlations with the Hogan Development Survey (HDS) (Hogan & Hogan, 1997b) were estimated. The data was collected from Assessio's Norwegian database. HDS is a personality test measuring the tendency for derailment in work life and provides information on factor level and scale level. The factor level is composed of three factors: Moving Away, Moving Against, and Moving Toward. For the Norwegian version of HDS (Hogan & Hogan 2006), only Moving Away and Moving Against could be extracted – the scales Diligent and Dutiful, were not identified as belonging to the Moving Toward factor.

Moving Away represents traits such as being unreliable and distrustful, while Moving Against represents traits such as being overly confident, risk-taking, and manipulative. The two remaining scales, Diligent and Dutiful, represent traits such as extreme focus on details and striving to fulfill obligations, and lack of ability to make independent decisions and being dependent on what others think, respectively. The hypotheses about the relationships between MINT factors and HDS are specified in table 4.3 and the results are presented in table 4.4.

Table 4.3

Hypotheses for the relationships between MINT and HDS factors

	Integrity	Interpersonal Orientation	Task Orientation
HDS Factors			
Moving Away	-		
Moving Against		-	
HDS Scales			
Diligent			+
Dutiful	-		
Skeptical		-	

Note: + and - indicate the assumed direction of covariation

Table 4.4

Correlations between MINT and HDS factors (n=183)

	Integrity	Interpersonal Orientation	Task Orientation
HDS Factors			
Moving Away	-.60*		
Moving Against		-.58*	
HDS Scales			
Diligent			.49*
Dutiful	-.39*		
Skeptical		-.45*	

Note: *p<.01

The results, presented above in table 4.4, provide support for all pre-defined hypotheses. As expected, Integrity correlates strongly and negatively with the Moving Away factor. This hypothesis was based on contrasting content of Integrity compared to that of emotional stability and dependability. The negative correlation between Integrity and Dutiful was also expected due to the fact that Integrity represents stability and confidence in one's own ability while Dutiful concerns insecurity and the fear of acting independently.

Interpersonal Orientation correlated strongly and negatively with Moving Against, the latter representing overly confident, risk-taking, and manipulative tendencies. The traits associated with Moving Against contrast those of Interpersonal Orientation, which include being helpful and supportive of others. Interpersonal Orientation also represents being trustful of others, which explains the strong and negative correlation with the Skeptical scale, which measures the lack of trust in others.

Task Orientation shows a positive correlation with the Diligent scale. Thus, variance concerning being focused on details and on how tasks are to be performed in order to reach maximum results is shared by these two scales. Overall, the results support the construct-related validity of the MINT factors.

Measuring and Assessing Individual Potential (MAP), ServiceFirst, and Matrigma

Additional validation of MINT was conducted against Measuring and Assessing Individual Potential (MAP) (Sjöberg, Svensson, & Sjöberg, 2017), ServiceFirst (Fogli, 2009), and Matrigma (Mabon & Sjöberg, 2017). MAP is a personality test based on the FFM with the five factors labeled as Social Style (representing Agreeableness), Conscientiousness, Emotional Stability, Extraversion, and Openness. ServiceFirst (Fogli, 2009) is comprised of four subfactors: Active Customer Relations, Polite Customer Relations, Helpful Customer Relations, and Personal Customer Relations. Interpretation, however, is based on one overall factor representing service potential. Matrigma (Mabon & Sjöberg, 2009) measures general cognitive ability using progressive matrices as the item format.

Data was collected from employees in a grocery store chain mainly located in or near Oslo, Norway. Descriptive statistics for the sample are presented in table 4.5

Table 4.5.
Means (SD) of the Norwegian grocery store sample

	MAP	Matrigma	ServiceFirst
n	208	202	193
Age	34 (12)	33 (12)	33 (12)
Gender			
Male	33,00 %	34,00 %	34,00 %
Female	67,00 %	66,00 %	66,00 %
Educational level			
Elementary	12,00 %	12,00 %	10,00 %
Lower and Upper secondary	58,00 %	56,00 %	59,00 %
Academic studies	13,00 %	14,00 %	14,00 %
University degree	17,00 %	17,00 %	17,00 %

Note: Due to percentages being rounded up, the total does not equal 100%

The pre-defined hypotheses between MINT and the relevant scales in MAP and in ServiceFirst are outlined in table 4.6, and the results are presented in table 4.7.

Table 4.6.
Hypotheses for the relationships between MINT factors and Social Style, Conscientiousness, and Emotional Stability in MAP and ServiceFirst

Scale	Integrity	Interpersonal Orientation	Task Orientation
Social Style		+	
Conscientiousness			+
Emotional Stability	+		
ServiceFirst	+		+

Note: + and - indicate the assumed direction of the covariation

Table 4.7.
Correlations between MINT and Social Style, Conscientiousness, and Emotional Stability in MAP (an=208) and ServiceFirst (bn=193)

Scale	Integrity	Interpersonal Orientation	Task Orientation
Social Style [a]		.21**	
Conscientiousness [a]			.60**
Emotional Stability [a]	.78**		
ServiceFirst [b]	.55**		.23**

Note: *p<.05; **p<.01

The results show that the correlations are in line with the pre-defined hypotheses, although not all of them were significant. The relationships between MINT and MAP correspond to the theoretical model underlying the two tests: the aspects of the FFM in MINT (Conscientiousness, Agreeableness, and Emotional Stability) were shown to correlate with the corresponding factors in MAP. ServiceFirst was developed in order to provide information about a person's capacity to handle customer relationships; this is supported by its correlation with Integrity. No correlation was found between MINT and Matrigma.

Test–criterion relationships

Several studies have been conducted to investigate the correlations between MINT scores and external criteria. The test data and criterion data were collected independently in these studies, with some of the studies being predictive (data on the criterion collected at a later point in time) and others concurrent (data on the criterion collected at the same point in time).

Job performance and job satisfaction among traffic controllers

One criterion-related study was conducted on 47 candidates applying to positions as traffic controllers at seven different locations in Sweden. Seven respondents had completed two years of high school education, while the rest had completed three to four years of high school education.

All of the respondents had previously been tested with the Swedish version of MINT. The criterion data used were the performance ratings by the managers as well as self-ratings of job performance and job satisfaction.

Each respondent was rated by two managers: an operational manager and a line manager. The mean value of these ratings was used as the criterion data. The rating scale used in these assessments consisted of 19 items focusing on the person's problem-solving skills, interpersonal skills, and level of work commitment.

The factor analysis of the managers' performance ratings revealed a clear first order factor explaining 51% of the variance. The inter-rater reliability was high (ICC=.90), implying that the operational and line managers at each location had a similar rating of how well the respondents were performing

Self-rated work performance was measured by a five-item scale focusing on

commitment, confidence, and competence (Hall & Hall, 1976). Job satisfaction was assessed with a three-item scale adapted to Swedish (Hellgren, Sverke & Isaksson, 1999) from a scale by Brayfield and Rothe (1951).

Table 4.8 shows the number of respondents in the analysis, the inter-rater reliability, Cronbach's Alpha, and the restriction of range of MINT test scores. Table 4.9 presents the observed correlations and the operational validity estimates for all MINT scores and criteria.

Students and grades

Another study was conducted on 79 students from the Department of Psychology at Stockholm University. All of the students were in the first semester of their Master's Program. The students were administered MINT on two occasions (see the Swedish test–retest study, n=97, previously in Part 4) and the average of the two scores was used as the MINT score in the analysis. The grades received on four exams (in biological psychology, developmental psychology, cognitive psychology, and social psychology) were used as the criteria. The grades ranged from 0 to 5 (M=2.79; SD=1.01). The results are presented in table 4.9.

Table 4.8.
Descriptive statistics, reliability and restriction of range results for the sample of Swedish traffic controllers and students

Criterion	N	ICC	Alpha	u^1	u^2	u^3
Job Performance	47	.90	.96	.84	.93	.82
Grades	79		.72	.77	.80	.92
Job Performance (self-rating)	47		.64	.84	.93	.82
Job Satisfaction (self-rating)	47		.75	.84	.93	.82

Note: N=sample size; ICC=Intra-class coefficient; Alpha=Cronbach's Alpha; u1=restriction of range for Integrity scores; u2=restriction of range for Interpersonal Orientation; u3=restriction of range for Task Orientation

Table 4.9.
Observed correlation and operational validity for the sample of Swedish traffic controllers and students

	Integrity		Orientation		Orientation	
Criterion	r	p	r	p	r	P
Others' ratings						
Job Performance	.25	.32	.08	.10	.09	.12
Grades	.11	.17	.17	.25	.26	.33
Self-ratings						
Job Performance	.48	.67	-.29	-.39	.40	.59
Job Satisfaction	.05	.07	.44	.51	.00	.00

Note: r=observed correlation; p=operational validity

Job performance and CWB among grocery store employees

In conjunction with the Norwegian test–retest study described previously in Part 4 (conducted on a sample of n=129 individuals working in a grocery store chain mainly located in Norway), employees were rated by their managers on job performance and CWB in a concurrent validity study. In total, 36 store managers rated 101 employees.

The supervisory rating scheme consisted of nine questions focusing on the employees' level of general job performance (e.g., problem-solving, ambition) and on CWB (e.g., showing up late to work, non-job-related use of the Internet).

The investigation of job performance and CWB ratings revealed bias between raters. This bias was concluded to be systematic and most plausibly due to the fact that store managers can only evaluate their employees relative to the store and group that they manage. Thereby it was decided to only analyze data from store managers who had rated two or more employees. Thus, due to missing data, regarding the CWB rating, 44 employees were included in this analysis. For the analyses, within-group standardized scores were used. The store manager thus represented a unit (sample) for analysis and the scores from each unit were used as the criteria in the analysis.

In validation studies of this kind it is likely for the validity to be underestimated due to unreliability in the dependent variable. It was not possible, however, to estimate the reliability in the rating since each employee was only rated by one supervisor. Therefore, an estimate for the reliability in this type of supervisor rating was taken from a meta-analysis (Viswesvaran, Ones & Schmidt, 1996) providing the reliability estimate for supervisory ratings of job performance (.52) and of CWB (.66). Information regarding restriction of range was not possible to estimate and is thus not corrected for in the operational validity estimates presented in table 4.10. The results show that the MINT factors predict both general job performance and CWB, thereby indicating that MINT is useful for predicting both negative and positive behavior in a selection context.

Table 4.10.
Correlations between MINT and job performance (n=101) and CWB (n=44) ratings in the grocery store chain sample

	Job Performance		CWB	
MINT	r	p	r	p
Integrity	.19**	.26	-.28*	.35
Interpersonal Orientation	-.12	-.17	-.13	-.16
Task Orientation	.18*	.25	.02	.02

Note: **p<.05; *p<.10; r=observed correlation; p=operational validity

Job performance among retail employees

A concurrent validity study was also conducted on warehouse workers and truck drivers in a Norwegian retail organization. Initially, 101 employees were included in the study, before reducing the final sample size to 68 employees after eliminating those for whom complete data on both MINT and the performance criteria were not available. A large majority of the employees (83%) had completed three years of upper secondary education or lower education. The mean age for the sample was 40 years (SD=11).

The criteria measure consisted of managerial performance ratings on a 7-point Likert scale (Cronbach's alpha α ith19 items focusing on problem-solving skills, interpersonal skills, and work commitment. The results show that Interpersonal Orientation has the strongest relationship with general job performance.

Table 4.11.

Correlations between MINT and job performance ratings in the retail sample (n=68)

	Job Performance	
MINT	r	ρ
Integrity	.13	.18
Interpersonal Orientation	.24	.33
Task Orientation	-.05	-.07

Note: r=correlation between MINT and criterion without corrections; ρ=operational validity, r adjusted for measurement error (.52) in the criterion (Viswesvaran, Ones & Schmidt, 1996)

Summary

Validity is defined as evidence that the theory and empirical findings support the use of MINT in order to make correct interpretations and decisions. The overall evidence that MINT may be useful for individual assessment in the prediction of external criteria rests on the strength of its co-variations with other measures and relationships with external criteria. The findings are compelling and suggest that MINT scores can play an important part in a selection context by predicting CWB.

Part 5

Further development, language versions, and norm updates

In the following part, norm updates are outlined along with information regarding the psychometric qualities of the additional language versions that MINT has been translated into and adapted for over the years. Note that updating the norms for MINT requires not only the re-calculation of means and standard deviations but also the re-calculation of factor scores.

October 2011

At this point it was decided to update the existing norm group (n=1,545) with data (n=20,662) collected via Assessio's web platform during 2007–2010. In total, the updated norm group consisted of n=22,207.

All of the respondents in the norm group completed MINT as part of a selection process. The large majority of these assessments had been administered in Swedish (n=19,941), while n=744 had been administered using the Norwegian version (developed according to the description in Appendix B). The remaining had been administered using new language versions: English (n=968) and Finnish (n=732).

The entire norm group consisted of 13,352 women and 8,855 men and the average age of the group was 34 years (SD=11). 578 respondents (2.6%) had completed elementary school as their highest educational level, 1,977 respondents (8.9%) had completed a two-year lower secondary program, 4,843 respondents (21.8%) had completed a three- or four-year upper secondary program, 3,526 respondents (15.9%) had completed a higher educational program of at least three years, 11,008 respondents (49.6%) had completed a higher educational program of more than three years, and 275 respondents (1.2%) had conducted doctoral studies. In order to investigate psychometric qualities, equivalence across language versions – and thus the possibility of applying a global norm group – a multiple-group confirmatory factor analysis (MG-CFA) was conducted, following the principles established by Cheung (2008) (for further technical details regarding this analysis, please see the previous MINT Technical manual in Sjöberg & Sjöberg, 2012).

Overall, the results provide support for stable psychometric properties and for equivalency across the various language versions of MINT. Thus, it was decided to implement one global norm group of n=22,207.

December 2014

After the norm update in October 2011, the popularity of the instrument increased heavily and at the end of 2014 an extensive amount of additional data had been collected. This provided the opportunity to update the existing norm group of n=22,207 by adding data from n=102,292 which had been collected between 2011 and 2014. The total sample then consisted of 124,499 respondents.

The data from the 124,499 respondents was collected using 13 different language versions (the newly developed language versions were Danish, Estonian, Hungarian, Latvian, Lithuanian, Russian, Slovak, Spanish, and Thai) and came mainly from private-sector assessments for selection purposes. The age of the respondents ranged from 18 to 75 years (M=30.61, SD=10.95) and the overall analysis showed a low positive correlation between age and Integrity (r=.14; p<.001), a low negative correlation between age and Task Orientation (r=-.14; p<.001), and a positive correlation between age and Interpersonal Orientation (r=.11; p<.001). 48 percent of the sample were women, who scored somewhat higher on all three scales compared to the men (Integrity, r=.02, p<.001; Interpersonal Orientation, r=.06; Task Orientation, r=.05; p<.001). In table 5.1, the number of assessments per language version across gender is presented along with the means and standard deviations for age.

Table 5.1.

er o assess ents across lan a e ersion a e an en er lo al nor ro n

	Age		Gender		
Languages	M	SD	Female	Male	Total
Danish	33	12	2,094	2,475	4,569
English	31	10	9,467	10,321	19,788
Estonian	31	7	22	6	28
Finnish	29	10	1,215	848	2,063
Hungarian	27	5	8	12	20
Latvian	29	9	286	165	451
Lithuanian	30	7	141	101	242
Norwegian	26	9	10,256	9,753	20,009
Russian	32	8	279	210	489
Slovak	33	7	6	4	10
Spanish	32	7	779	540	1,319
Swedish	32	11	34,303	39,979	74,282
Thai	28	4	831	398	1,229
Total group	30	8	59,687	64,812	124,499

Evidence based on internal structure

The extensive amount of data available provided both the opportunity to re-analyze the internal structure in order to confirm the validity of the updated norms (mean values and standard deviations) and enabled an investigation into the levels of equivalence across the language versions (existing and additional). For this purpose, the following evidence based on internal structure was analyzed:

- Goodness of fit measures from Confirmatory factor analysis (CFA)
- Goodness of fit measures from Multiple-group CFA (MG-CFA)
- Internal consistency of Integrity, Interpersonal Orientation and Task Orientation

Confirmatory Factor Analysis – Goodness of Fit

As outlined previously, CFA is a factor analysis performed with the purpose of confirming a hypothesized factor structure. For MINT, the hypothesized theoretical structure corresponds to a hierarchical model composed of the higher order construct of Integrity and the two lower-order factors of Interpersonal orientation and Task orientation.

The variance–covariance matrices of raw scores were used as input for the hierarchical nested factor analysis (NFA) that was carried out (Gustafsson, 2002). The analysis tool of Mplus (Muthén & Muthén, 1998-2012) was used for maximum-likelihood estimation.

The hierarchical structure of the measurement model, composed of a higher-level general factor and a number of residual factors, allow for the factors to be measured independently. As explained previously, this is essential for correctly interpreting profiles (thus scores) comprised of different personality measures. When dimensions overlap, which they inevitably do in personality measurement, it is impossible to isolate the separate components when the personality scales are only summated and compared. The hierarchical model lends itself to an orthogonal approach, in which the factors are nearly zero-correlated (not related empirically), which considerably increases the precision of profile interpretations. For a more detailed description of the operationalization of the MINT model using six indicators, see previously in Part 5.

The results from the language-specific CFA's are presented in table 5.2 [7]. Along with the chi-square statistics and four goodness-of-fit indicators, the Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Standardized Root Mean Square Residual (SRMR), and Root Mean Square Error for Approximation (RMSEA) are presented. As mentioned previously, CFI and TLI values .90 often is used as a lower limit for defining an adequate fit. For the SRMR, values lower than .05 indicate a good fit, and for the RMSEA, .10 represents the upper limit for an acceptable fit (Hu & Bentler, 1997; Kelloway, 1998). Although the results differ somewhat when comparing the various language versions, the overall results (see the last column in table 5.2) show an acceptable fit between the proposed model and the data. The standardized factor loadings for the model, using the total sample of n=124,499, are presented in table 5.3.

[7] MINT was initially developed within the CFA framework and therefore this approach was used in the norm updates. Note, however, that an accepted alternative approach for testing this type of nested model is exploratory factor analysis (EFA). Asparouhov and Muthén (2009) have proposed a hybrid between CFA and EFA, namely ESEM. The ESEM approach is an EFA measurement model with rotations that can be used in a structural equation model. By setting the ES indicators to only load on the integrity factor, setting the other indicators to load on all three factors, and then rotating the solution using a Geomin rotation, will give identical results to the CFA approach.

Table 5.2

Confirmatory factor analyses, goodness of fit estimates, and reliability estimates across language versions for the total norm group of n=124,499.

	Language version													
	Danish	English (US)	Spanish	Estonian	Finnish	Hungarian	Lithuanian	Latvian	Norweigan	Russian	Slovak	Swedish	Thai	Total
	Sample sizes													
	4,569	19,788	1,319	28	2,063	20	242	451	20,009	489	10	74,282	1,229	124,499
CFA														
$c^{2\,(df=7)}$	154.96*	790.512*	68.335*	a	73.132*	a	32.749*	28.513*	442.565*	25.315*	a	1, 948.661*	94.124*	3, 565.925*
CFI	0.982	0.968	0.973	a	0.984	a	0.948	0.972	0.986	0.982	a	0.987	0.947	0.984
TLI	0.961	0.931	0.942	a	0.965	a	0.888	0.940	0.971	0.620	a	0.971	0.887	0.965
SRMR	0.043	0.051	0.053	a	0.043	a	0.073	0.049	0.038	0.043	a	0.041	0.080	0.044
RMSEA	0.068	0.075	0.082	a	0.068	a	0.123	0.083	0.056	0.073	a	0.061	0.101	0.064
95% CI Lower	0.059	0.071	0.065	a	0.054	a	0.083	0.052	0.051	0.044	a	0.059	0.083	0.062
95% CI Upper	0.078	0.080	0.100	a	0.082	a	0.167	0.115	0.060	0.105	a	0.063	0.119	0.064
Reliability														
Integrity	0.961	0.930	0.918	0.941	0.941	0.900	0.924	0.904	0.915	0.630	a	0.921	0.940	0.922
Interpersonal Orientation	0.831	0.805	0.836	0.895	0.831	0.800	0.809	0.801	0.826	0.858	a	0.845	0.784	0.888
Task Orientation	0.849	0.881	0.881	0.703	0.900	0.771	0.869	0.869	0.901	0.913	a	0.894	0.918	0.839

Note: *p<.05; df=(degrees of freedom); a=not possible to estimate due to small sample size; CFI=Comparative Fit Index; TLI=Tucker Lewis Index; SRMR=Standardized Root Mean Square Residual; RMSEA=Root Mean Square Error of Approximation

Table 5.3

Standardized factor loadings for the model (n=124,499)

	Integrity	Interpersonal Orientation	Task Orientation
ES1	.84		
ES2	.86		
A1	.38	.72	
A2	.35	.68	
C1	.45		.76
C2	.50		.72

Note: ES=Emotional stability; C=Conscientiousness; A=Agreeableness

Multiple-Group Confirmatory Factor Analysis – Goodness of Fit

Multiple-Group Confirmatory Factor Analysis (MG-CFA), following the principles suggested by Cheung (2008), was performed to test for measurement equivalence across 10 of the 13 language versions of MINT[8]. Measurement bias occurs when the relationship between the construct and the observed scores differs between two or more groups. For example, a test would suffer from measurement bias if the observed scores reflected Integrity for one group but not for another (i.e., MINT measured different constructs as a function of the specific language version). Likewise, measurement bias would also be evident if a set of scores were shown to be reliable for one group of respondents but not for another, indicating that the same construct, (e.g., Integrity) had been measured to different degrees. Measurement bias can be conceptualized as a set of questions regarding the equivalence (or invariance) of different parameters of a measurement model.

Based on a baseline model with no constraints (Unconstrained) of equality between the groups, constraints were gradually imposed: metric invariance (same factor loadings), scalar invariance (same intercepts), invariant factor variances, and invariant uniqueness (full uniqueness). The results regarding the χ^2 al esan the itin icescan eta enas evidence that MINT is free from measurement bias across the 10 language versions (see table 5.4). Even the most restricted model, which imposed equality constraints on the measurement error terms, still provides a close (albeit not exact) fit between the data and

[8] MINT language versions in Estonian, Hungarian, and Slovak had too small sample sizes in order to be included in the CFA.

the model, thus suggesting that the measurement model parameters appear to be stable across language versions. In conclusion, the results show support for the use of the same factor scores across language versions.

Table 5.4.

Multiple-Group Confirmatory Factor Analysis across the 10 language versions (n= 124,441)

	Unconstrained	Metric	Scalar	Variances	Uniqueness
x^2 (df)	40,657,035(70)	40,657,038(115)	40,657,053(169)	40,657,057(196)	40,657,063(250)
		.030(45)	.015(54)	.004(27)	.006(54)
CFI	.987	.987	.987	.987	.987
TLI	.973	.983	.989	.990	.992
RMSEA	.022	.017	.014	.013	.011
95% CI Lower	.021	.017	.014	.013	.011
95% CI Upper	.022	.017	.014	.013	.011

Note: df=degrees of freedom; CFI=Comparative Fit Index; TLI=Tucker Lewis Index; RMSEA=Root Mean Square Error of Approximation

Internal consistency

The reliability of the factors was computed according to the principles outlined in Part 3, and in conjunction with the CFA using Mplus software (Muthen & Muthén, 2012). In this case, the factor determinacy coefficient is the correlation between the estimated and the true factor scores. Using Reuterberg and Gustafsson's (1992) equation, the reliability was .92 for the general measure of Integrity and .89 and .84 for the lower-level factors of Interpersonal Orientation and Task Orientation respectively (N=124,499); see table 5.2 These results support the internal consistency of the MINT factors and clearly satisfy the requirements of precision.

Part 6

Financial aspects of personnel selection

Most companies are aware that employees differ with respect to job performance and are thus also aware of the importance of recruiting high performing individuals and the consequences of hiring unsuitable employees for the organization. An economical and carefully set up selection process could be decisive for the competitiveness of an organization, its profitability, and its future survival.

The challenge for many organizations is to accept and implement valid and financially beneficial selection methods when making employment decisions. Research focusing on employee selection and its relation to job performance and financial utility has been extensive and the findings are unambiguous. Nevertheless, it has been difficult to accept among practitioners. Regardless of geographic location and type of market, some practitioners and organizations have been more open to this research-based knowledge – and to the improvement that could be gained from a pragmatic evaluation of their selection process – than others. One of the most distinguished researchers in the field, Frank Schmidt, writes about personnel selection in The Industrial-Organizational Psychologist:

Some years back I realized that personnel selection had become a churning arena of constant irrationality. There is a huge disconnect between what we know to be true from research and what people pretend to be true. There is a serious corruption of scientific truth caused by legal and ideological intrusions into the field of selection and the failure of the profession to respond appropriately to these intrusions. (Schmidt, 2006, p. 27)

Schmidt is referring to the difficulties in accepting the research results in this area. His view is that this reluctance depends on a number of factors, including ideology, entrenched conceptions, ideas about the area, and notions of which methods and processes are the “good ones.”

These conceptions and ideas may be based on methodological features that no longer have, or perhaps never had, research support. Since scientific advancement is an ever-changing, iterative process, it is extremely important to continually and critically re-evaluate the methods utilized in light of new research. In many cases, people have staked their professional identity on and invested years of work, money, and time into a method or theory they firmly believe in – regardless of research support. In addition, some of the selection methods used at present have no scientific basis, and evaluations of methods are often based on subjective feelings that the method “works.” Such a belief takes hold when organizations consider that they have successfully managed to avoid recruiting obviously counterproductive individuals.

Subjective opinions of this sort are a barrier to conceiving how much better overall employee job performance might have been with the aid of more efficient selection procedures. Today, the use of the research-based knowledge now available implies accepting and taking responsibility to contribute to the professionalism within this field.

The purpose of personnel selection

The purpose of personnel selection is to identify and choose those candidates who have the highest probability of performing well at the job. There is, in turn, a linear relationship between job performance and economic utility: the higher job performance, the greater financial contribution to the organization.

As mentioned, research on employee selection and its relation to job performance and financial utility is both extensive and homogenous. Briefly and simply, the results may be summarized as follows: GMA is the strongest single predictor of job performance, while some aspects of personality also contribute to predicting job performance.

Identifying the individual traits most suitable for predicting job performance is essential for determining whether or not the method is economically viable; this is because the financial outcome is a direct linear function of the predictive validity.

Choice of selection method

Individual traits such as GMA and personality may be measured using several different methods. GMA can, for instance, be measured through an intelligence test or by administering assessment center exercises to measure problem-solving ability. Personality may be measured through self-reports or candidate interviews. Methods differ with respect to both validity and reliability. Some provide more accurate information, are more valid, than others, and some are considerably more expensive to both develop and administer. The most financially advantageous method, combining high validity and lowest cost, would seem the obvious choice for measuring these traits. This is seldom the case in practice however, for the reasons mentioned previously.

When designing a selection process and choosing selection methods, it is important to make sure that the method in question measures the aspects of cognitive ability and personality that are most likely to predict job performance. Tools that gather information about a specific aspect of an individual's cognitive ability may be less useful, since research shows that the more comprehensive GMA factor is best for predicting job performance (Schmidt & Hunter, 2004). Regarding personality, some dimensions predict job performance better than others. The general conclusion is that some methods are more effective than others when it comes to measuring the relevant aspects of traits such as GMA and personality with respect to job performance. This further underlines the importance of choosing the correct method for the specific context.

When choosing a selection method, an organization often considers practical aspects such as whether the method is available on the web, how many different language versions there are, whether it can be administered in an unsupervised form, and if the test results generate extensive reports (reinforcing the impression that more valid and reliable information is being obtained about the candidates). Even factors such as payment flexibility may influence the choice of method. These practical requirements must, however, be viewed in relation to the economic utility of the method. Otherwise, an organization may find that the practical considerations influenced the choice of method at the expense of the method's predictive validity. Such situations often result in selection processes eventually causing a financial loss for the organization. The financial gain is a direct function of a method's validity and cost of implementation; choosing methods

relying heavily on practical considerations may thus result in using a poor method that costs just as much as a valid one, or more. The cost may exceed the return for the company based on the job performance of a new employee over several years, since the validity of the selection will be severely diminished (close to zero) and any recruitment of high-performing employees is thus coincidental.

Although no clear link has been established between quality, i.e., validity, and the cost of selection methods available on the market, the cost of methods with low validity is often as high or even higher than that of their more valid counterparts. One explanation might be that the developers of poorer methods choose not to invest in extensive and time-consuming development and validation processes, prioritizing, rather, investing in numerous language translations, various result reports, and high-tech delivery solutions. Such offerings may give the erroneous impression that they can lead to increased validity by appearing convincing, of high quality, and advanced. These methods often lack documentation outlining their development work, the links with theory, the design of their measurement model and its interpretation, and the rational for its purported practical applicability. Few customers would accept such a situation otherwise – one in which they lack a manual or any other information about the qualities of a product or service. This situation within the field of personnel selection has fascinated the academic community and has become a research area of its own.

One complication when choosing selection methods is that the suitability of one method over another may depend on certain practical circumstances. As can be seen in table 6.1, the use of work samples is one of the methods that shows very high validity. However, it is seldom possible to apply the method in practice since, for many occupations, the work would have to extend over a period of time (i.e., probationary employment) in order to obtain a work sample. In addition, it is if not impossible at least impractical to collect and use work samples for a large number of applicants, and it may also be the case that the job does not require previous experience.

There are few types of jobs, such as architects, graphic designers, and carpenters, that have formal requirements which can be attested for in a work sample. In the final analysis, the practical aspects are such that it would often be too time-consuming and expensive for the organization to utilize such a selection method for a large number of applicants.

It may be necessary to set up a selection process that uses different methods at different stages to gradually arrive at a shortlist. A common procedure at present is to review candidates' formal criteria to ensure that, for instance, requirements relating to education and experience are satisfied. In situations where such requirements are not relevant, resumés (CVs) and personal letters, which most candidates are asked to submit, are used in the first phase of the process. This approach, however, has been shown to have very little value when assessing the suitability of candidates (low validity and reliability) and it is also time-consuming (thus expensive). Nevertheless, it is used in most selection processes.

Table 6.1.

Predictive validity for a number of methods

Method	r	R	Incremental validity	Increase in percent	GMA tests (r)
GMA tests	.51				
Integrity tests		.65	.14	27%	.00
Work sample		.63	.12	24%	.38
Structured inter view		.63	.12	24%	.30
Unstructured inter view		.55	.04	8%	.38
Assessment center		.53	.02	4%	.50
References		.57	.06	12%	.00
Graphology		.51	.00	0%	.00

Note: From Schmidt and Hunter (1998)

After this first stage, candidates are typically interviewed. Selection processes without an interview at some stage of the selection process is extremely rare and would likely be regarded with great skepticism by practitioners and organizations. From a scientific perspective, interviews can differ in quality. Interviews may be of high quality and cost-effective, or they may be of poor quality and lead to a negative return on investment for the organization. Interviews of high quality are usually characterized by being standardized, in that the questions have been tested, evaluated, and structured, and the same questions are given to all candidates in the same manner. The responses are then interpreted and evaluated on the basis of predetermined criteria. Interviews of this type may be of high quality (high validity) and are, to a large extent, specifically designed for the job in question as they have a clear element of work sampling. High-quality interviews of this type are, however, comparatively unusual in practice. As interviews are time-consuming and therefore expensive, this is often remedied by limiting the number of interviews and relying instead on the information provided in the resumés and the applicants' letters. As mentioned, resumés and applicant letters has been shown to have very low validity, as it appears that candidates called to an interview are in no way better with respect to job performance than those who are not. This means that not only may some unsuitable candidates be included in the selection process, but also that some suitable candidates will be excluded. By this stage, a considerable amount of time (and thus money) has typically been spent on reading applications and assessing and classifying candidates, a process that provides little information as to future job performance.

When a number of different methods are used in a selection process, it is important to establish not only the validity of each method, but also the relationships among these methods. The reason for using a variety of methods is to provide as much relevant information as possible about certain individual traits in order to predict future job performance. The methods chosen should therefore complement each other by providing unique information about the candidate without overlapping, as this would in practice mean receiving the same information from two sources. Receiving the same information using different methods also means paying several times for the same information. This is a poor financial choice, but it also has an ethical aspect, as candidates redundantly provide the same personal information on numerous occasions.

Using methods which are known to have low validity – or which generate no valid information as to their future job performance – is a questionable practice. In selection, we are looking for information that has a bearing on the future job performance of the candidates, not information in general. The information collected should be relevant, relate to future job performance, and be obtained in as fair, reliable, and cost- and time efficient manner as possible.

When it comes to measuring the GMA and personality traits that are central to predicting job performance, psychological tests play a central role for the following reasons.

- GMA and personality tests may be administered to any candidate, irrespective of profession, organizational level, and background (they do not need experience of the job for which they are applying).
- Compared to other methods with the same level of predictive power (work sampling and the structured interview), psychological tests are the most cost-effective.
- Testing both GMA and personality provides the most effective combination of selection methods, since GMA and the relevant personality dimensions do not overlap.

In the distant past, tests were developed for the purpose of testing large numbers of applicants with no formal qualifications, such as for the U.S. Army, and not for the assessment of a few shortlisted candidates at the end of a selection process.

MINT for selection

MINT has been developed to satisfy qualitative (validity), practical, and financial criteria. It aims to produce a high level of validity and accuracy when ranking applicants with respect to future work behavior. In so doing, it also aims to be applicable in practice to a large number of applicants as an early stage of screening at a modest cost – a cost that is expected to be less than the financial benefits clients gain through the increased job performance resulting from hiring candidates based on MINT scores.

Screening based on relevant personality traits which are measured in an effective, standardized, and reliable way may well be the best alternative with respect to quality. Excluding applicants at an early stage, when there is little probability that they will perform well on the job (or if they are found likely to engage in counterproductive behavior), provides considerable financial advantages. Candidates assessed for early elimination will therefore not take part in later stages of selection and ultimately not have the opportunity to harm the organization.

Note however, that although MINT is suitable for screening, there are no barriers for using it at a later stage of the selection process.

Part 7

Instructions for use and interpretation

Areas of use

MINT has been developed for the purpose of providing evidence-based information for making personnel decisions based on candidates' likelihood of demonstrating CWB. As the traits being measured are universal, MINT may be used to screen for CWB among applicants for any position, in any industry and business, and for jobs at all levels. MINT is designed to be applied in a personnel selection context and is most suitable in the initial stage as a screening tool. MINT is not intended for development purposes, such as manager and employee development, career guidance, team building, and coaching, or meant for use within a clinical context.

Administration and scoring

MINT is available via Ascend by Assessio and via partner systems using the Ascend API[9]. The candidates respond to items shown on the screen, and the web system computes raw scores, converts them into standardized scores, generates the results, and provides standardized feedback reports. The use of MINT requires a trained test administrator who may choose to either administer MINT remotely by sending a link to the respondent via e-mail, or to administer MINT on-site. For remote administration, the e-mail addresses of the candidates are required.

Requirements for testing

The requirements for administration and conditions of testing are the following:

- MINT has no time limit for testing. A candidate may thus suspend the testing temporarily and resume it at a later point in time. It is recommended, however, that candidates complete MINT in a coherent session.

- A basic reading comprehension is required. All of the needed test-taking instructions for MINT are available on the screen throughout the testing. Although the instructions aim to be simple and straightforward, they nevertheless require a basic level of reading comprehension. For the Swedish version, a reading index, "Läsbarhetsindex" (LIX), has been used to obtain a measure of the complexity of a text, and is based on the average number of words per sentence and the proportion of long words (i.e., those with more than six letters) expressed as a percentage. There are five generally agreed levels of difficulty linked to different LIX values, ranging from very easy to read (children's books) to very difficult to read (formal bureaucratic texts). The LIX value for the MINT items is 23 and instructions is 38, which means that MINT in terms of comprehension can be categorized as "very easy to read" and "easy to read" respectively. The language may be easy, but factors such as serious reading and writing difficulties, dyslexia, and taking the test in a non-native language may still

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

negatively impact understanding and thus potentially influence the results. At present, there is no specific information on if and how MINT scores may be affected by these kinds of factors. It is therefore important that the test administrator, as best as possible, ensures that candidates have sufficient reading comprehension skill to complete the MINT items.

- Other impairments that are likely to have a negative effect on the test results should be identified, including but not limited to perceptual, visual, and cognitive impairments.
- A non-distracting testing environment is needed. Public environments, e.g., internet cafés and public transportation, are not suitable for taking MINT.
- A personal computer is recommended since MINT has been visually adapted and developed for administration on a full-sized computer screen. Test-taking via a tablet, smartphone, or similar device may affect the test results.
- A stable internet connection is needed for the full duration of the testing to ensure a valid result.
- Basic computer skills are necessary for candidates to be able to use a mouse and/or keyboard to complete MINT. The test administrator should ensure that the technical aspects do not cause any difficulties for the candidate, as this may have a negative effect on the results.

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

Information provided to candidates before testing

If MINT is administered remotely, a web link is sent to the candidates' e-mail addresses. The automatically generated e-mail is based on a template with basic information about the testing. This e-mail is editable so that the test administrator may insert specific information for a single candidate or group of candidates. The e-mail should always include the following information:

1. The purpose of testing
2. What type of test MINT is and why it is being used in the present context
3. How MINT will be administered and what is required for completing the test (see Requirements for testing previously in Part 7)
4. How the results will be used and saved, by whom, for how long and why
5. Notification that the candidate has the right to choose whether the test score may be included as part of the information about him or herself used in the selection process
6. Whether feedback will be provided to the candidate and, if so, when and in what format (e.g., standardized on screen, personal feedback, face-to-face meeting, over the phone), and what the feedback will contain
7. Contact details to the test administrator

More information about the rights and obligations of test distributors, test administrators, and candidates 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.

Interpretation and presentation of results

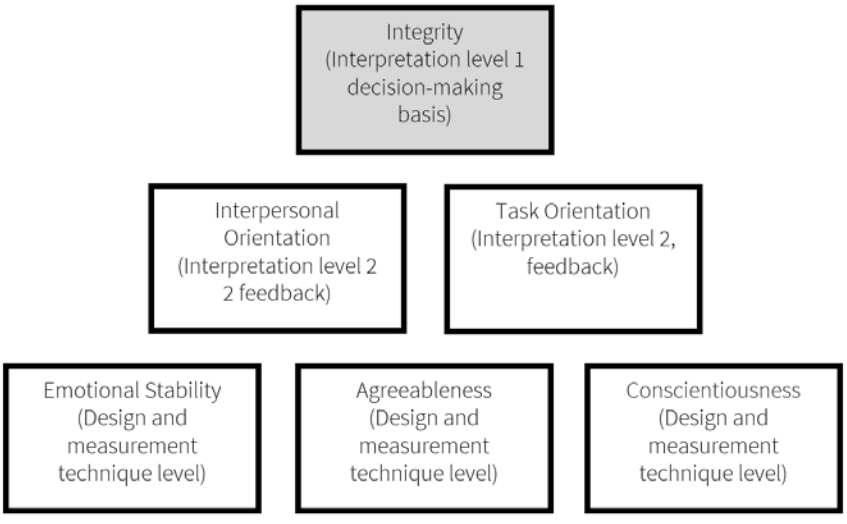
The results are presented on a standard scale known as the C-scale, which makes it possible to compare the scores of candidates. The C-scale ranges from 0 to 10, with a mean of 5 and standard deviation of 2. The C-scores for Integrity are divided into three levels in order to facilitate interpretation: Low scores (0–2 C-scores), Average scores (3–6 C-scores), and High scores (7–10 C-scores).

The higher score on Integrity, the less likely it is that the person will demonstrate CWB; conversely, the lower the score, the more likely it is that this behavior will occur. From an absolute perspective, this implies that when interpreting scores on the Integrity scale, the test administrator should be especially attentive of those with low scores (0–2 C-scores) due to their elevated probability of engaging in CWB.

Results on the subscales of Interpersonal Orientation and Task Orientation are also presented in C-scores, although without being divided up into different levels. The C-score for each scale gives an idea of the extent to which the candidate applies the approach in question in the workplace. The higher the score, the more likely the person will be to take that specific approach.

Test scores must always be interpreted by using the theoretical model on which the test is based. MINT is based on a hierarchic model; see figure 7.1, with a general factor, the Integrity scale, and two independent subfactors – the Interpersonal Orientation and Task Orientation scales.

Figure 7.1. Hierarchic model of MINT



The interpretation of test scores must therefore be subject to the restrictions inherent in this model. An account is given in the following section regarding the importance of the hierarchic model for the interpretation of test scores. In addition, the test administrator is provided with a number of "rules of thumb" for practical use of MINT. For a more detailed account of the theoretical background of the hierarchic model, refer to Appendix C.

The structure of MINT is comprised of a general factor, represented by the Integrity factor, and two independent subfactors, represented by the scales of Interpersonal Orientation and Task Orientation. The general factor, Integrity, is made up of a weighted combination of three of the personality dimensions from the FFM. The combination of the three dimensions provides an estimate of individual Integrity and thus of the likelihood that the person will demonstrate CWB. In addition, the two independent subfactors Interpersonal Orientation and Task Orientation may be derived from the model. These subfactors are not only independent of each other; they are also independent of the general factor, Integrity. This entails that there is no overlap; a result on one scale does not influence the score on the other scales. Thus, it is possible for a respondent to have high or low scores on each of the subscales, or the same results on both. It is also possible for a respondent to have low scores on both subfactors and a high score on the general Integrity factor.

Note also that the two subfactors, Interpersonal Orientation and Task Orientation, are derived from a lower level in the hierarchic model. In other words, they are theoretically at a lower level than the general Integrity factor and thereby characterized a lower reliability and reference to a narrower construct.

In addition, these subscales only cover some parts of the personality dimensions that make up the general factor. This implies that the Integrity score always should be interpreted in the first instance; high scores on the Interpersonal Orientation and Task orientation subfactors can never compensate for a low score on the general Integrity factor.

Standardized feedback reports

After the testing is completed, the web system will present the C-scores in a project overview. The project overview, available for the test administrator, presents the C-scores of all candidates in a project in the form of an on-screen list, with each candidate's name together with his/her scores for each scale, Integrity, Interpersonal Orientation, and Task Orientation. The web system also enables the ranking of candidates on the basis of their scores. The intention of this project overview is to provide a basis for decision-making at the group level.

The web system will also, for each candidate, generate two types of standardized feedback reports: the Interpretive Report and Your Results. The standardized feedback report labeled Interpretive Report is intended for the test administrator. This report provides a candidate's C-score for each of the three scales. The result for the Integrity scale is also given in terms of levels: whether low, moderate, or high along with a description of what the level represents. In addition, the report also provides information about MINT, what the scales measure, the meaning of scores, and how to interpret score levels. This report is made available to the test administrator in pdf format.

The standardized feedback report labeled Your Result is intended for the candidate, and

it is optional for the test administrator to provide candidates with this feedback. If the test administrator chooses to distribute this report to a candidate, it is displayed on the candidate's screen after the testing is completed.

This report graphically illustrates the candidate's C-scores on Interpersonal Orientation and Task Orientation along a continuum and thus displays if the candidate has a stronger orientation towards Interpersonal Orientation or towards Task orientation or if it is equal towards both. The graphic illustration is followed by a description of what the results mean and a list of typical and likely behavior. One of three possible interpretations is reported: (1) preference for Interpersonal Orientation, (2) preference for Task Orientation, or (3) equal preference for Interpersonal Orientation and Task Orientation.

The reason for not displaying the score or level of result on Integrity is that it is recommended as the basis for decision, which in turn is dependent on the context and group of candidates.

The Your Result report is designed as a standardized, stand-alone feedback report and does not require personal feedback. Note, however, that this does not exclude the possibility for a test administrator to provide personal feedback as well.

In addition, the Your Result report also provides information about MINT, what the Interpersonal Orientation and Task Orientation scales measure, the meaning of scores, and what to take into consideration when reading the results.

Standard error of measurement

All psychological measurements, including tests, are subject to measurement error. Some measurement error is possible to estimate and may thus be taken into account when interpreting individual scores. This error, the Standard Error of Measurement (SEM), is based on the reliability estimate of internal consistency. By using the reliability coefficient, it is possible to determine a range (interval) of scores around the candidates test score within which an individual would be expected to score if she or he were tested an infinite number of times. The interval is set at a certain level of confidence, usually at 68% or 95%. This percentage represents the level of probability that the interval covers the "true" score.

Based on the average reliability of the factors (.95) and the standard deviation of the C scale (2 C-scores), SEM is estimated to 0.5 C-scores for a confidence interval of 68% and to 1 C-score for 98%.

Example: A candidate receives a score of 5 on the Integrity scale. This means that the candidates' "true" score, when measurement error is taken into account, can with 95% probability be expected to be within the 4–6 range.

Summary

- Test scores on the general Integrity scale are to be interpreted initially and are recommended to serve as the basis for selection decisions.
- The Integrity score provides an indication of the likelihood that a candidate will engage in CWB (score levels: Low (0–2 C-scores), Average (3–6 C-scores), and High (7–10 scores)).
- The general Integrity factor is independent of the lower-level factors of Interpersonal Orientation and Task Orientation. Thus:
    - ✓ The general Integrity factor is independent from the subfactors, and it is therefore possible to obtain any combination of subfactor scores without it affecting the general Integrity score.

In turn, the subfactors are also independent of each other. Thus:

- ✓ The Interpersonal Orientation and Task Orientation subfactors are independent of each other, and it is therefore possible to obtain any combination of scores on these two scales (Examples: 2-2, 1-8, 7-2, 8-8 etc.).

The independence between factors has several implications for interpretation:

- ✓ Low scores on the Integrity factor cannot be compensated for by high scores on one or both of the subfactors.

This is due to the fact that test scores on the subfactors do not provide information about the likelihood of the candidate demonstrating overall CWB.

- ✓ High scores on a subfactor represent a stronger orientation, while low scores indicate a less strong orientation.

Scores on a subfactor indicate indicate whether CWB – regardless of the individual's overall Integrity score and thus likelihood of engaging in CWB – is more likely to occur within a interpersonal domain, e.g., bullying and inappropriate verbal utterances, or within a task-oriented domain, e.g., poor attendance and poor quality work.

• An individual test score should always take into account the measurement error inherent in any psychological testing, which means in this case that an individual's "true" score on MINT may be expected to be within the range C +/–1 with a probability of 95%.

References

Anderson, C. R., Hellriegel, D., & Slocum, J. (1977). Managerial response to environmentally induced stress. *Academy of Management Journal*, *20*, 260-272.

Arbuckle, J.L., & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago, IL: Small Waters Corporation.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438,

Baba, V. V., & Jamal, M. (1991). Type A behaviour, prevalence, and consequences. *Human Relations, 44*, 1213-1228.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*, 45-87.

Barrick, M. R., & Mount, M. K. (2003). Impact of meta-analysis methods on understanding personality-performance relations. In Murphy, K. R (Ed.), *Validity generalization. A critical review*. NJ: Lawrence Earlbaum Associates.

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85*, 349-360.

Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology, 35,* 307–311.

Brown, S. P., & Leigh, T. W. (1996). A new look at psychological climate and its relationships to job involvement, effort, and performance. *Journal of Applied Psychology, 81*, 358-368.

Cheung, G.W. (2008). Testing equivalence in the structure, means, and variances of higher-order constructs with structural equation modelling. *Organizational Research Methods, 11*, 593-613.

Cascio, W. F. (2000). *Costing human resources: The financial impact of behaviour in organizations* (4th ed.). Cincinnati, OH: South-Western College Publishing.

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behaviour and counterproductive work behaviour. *Journal of Applied Psychology, 90* (6), 1241-1255.

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. Swedish version: *Revised NEO Personality Inventory NEO PI-R™*. Stockholm: Psykologiförlaget AB, 2003.

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

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive workplace behaviours. *Journal of Applied Psychology, 92*, 616-627.

Elloy, D. F., Everett, J. E., & Flynn, W. R. (1991). An examination of the correlates of job involvement. *Group and Organization Studies, 16* (2), 160-177.

Farrell, D., & Stamm, C. L. (1988). Meta-analysis of the correlates of employee absence. *Human Relations, 41*, 211-227.

Fogli, L. (2009). *Service First. Adaptation and validation of a Service Instrument. Danish, Estonian, Finnish, Latvian, Lithuanian, Norwegian, Polish, Russian, Swedish, and US English versions.* Stockholm. Assessio International.

Gerpott, T. J. (1990). Intracompany job transfers: An exploratory two-sample study of the buffering effects of interpersonal support. *Prevention in Human Services, 8*, 113-137.

Greenberg, J. (May, 2006). Counterproductive work behaviour: A scientist-practitioner workshop. Workshop presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology. Dallas, TX.

Greguras, G. J., & McCook, K. (in press). An investigation of the antecedents and consequences of job attitudes. *Group and Organization Management*.

Gruys, M. L., & P. R. Sackett. (2003). Investigating the dimensionality of counterproductive work behaviour. *International Journal of Selection and Assessment, 11* (1), 30-42.

Gustafsson, J-E. (2002). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement*. Mahwah, NJ: Erlbaum.

Hall, D.T. & Hall, F.S. (1976). The relationship between goals, performance, success, self-image, and involvement under different organization climates. *Journal of Vocational Behavior, 9, 267-278.*

Hathaway, S. R., & McKinley, J. C. (1989). *Minnesota Multiphasic Personality Inventory-2, MMPI-2*. Minneapolis: University of Minnesota Press. Swedish version: *Minnesota Multiphasic Personality Inventory-2*, MMPI-2. Stockholm: Psykologiförlaget AB, 2004.

Hattrup, K., O'Connell, M. S., & Labrador, J. R. (2005). Incremental validity of locus of control after controlling for cognitive ability and conscientiousness. *Journal of Business and Psychology, 19* (4), 461-481.

Hellgren, J., Sverke, M., & Isaksson, K. (1999). A two-dimensional approach to job insecurity: Consequences for employee attitudes and well-being. *European Journal of Work and Organizational Psychology, 8,* 179–195.

Hogan, R., & Hogan, J. (1997a). *Hogan Personality Inventory*. Tulsa, OK: Hogan Assessment Systems. Swedish version: *Hogans Personlighetsinventorium*.

Stockholm: Psykologiförlaget AB, 2002.

Hogan, R., & Hogan, J. (1997b). *Hogan Development Survey*. Tulsa, OK: Hogan Assessment Systems. Swedish version: *Hogans Deskriptiva Skalor.* Stockholm: Psykologiförlaget AB, 2002.

Hollinger, R. C., & Clark, J. P. (1982). Formal and informal social controls of employee deviance. *The Sociology Quarterly*,*23*, 333-343.

Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *Inventory of Interpersonal problems, IIP*. San Antonio, TX: Psychological Corporation. Swedish version: *Inventory of Interpersonal problems, IIP*. Stockholm: Psykologiförlaget AB, 2002.

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.

Huselid, M. A., & Day, N. E. (1991). Organizational commitment, job involvement, and turnover: A substantive and methodological analysis. *Journal of Applied Psychology, 76*, 380-391.

Ingram, T. N., Lee, K. S., & Lucas, G. H. (1991). Commitment and involvement: Assessing a sales force typology. *Journal of the Academy of Marketing Science, 19*, 187-197.

Kanungo, R. N. (1982). Measurement of job and work involvement. *Journal of Applied Psychology, 67*, 341-349.

Kelloway, E. K. (1998). *Using Lisrel for structural equation modelling: A researcher's guide*. Thousand Oaks, CA: Sage Publications.

Kidwell, R. E. Jr., & Martin, C. L. (2005). *Managing organizational deviance*. Thousand Oaks, CA: Sage Publications.

Kolz, A. R. (1999). Personality predictors of retail employee theft and counterproductive work behaviour. *Journal of Professionals Services Marketing, 19* (2), 1999.

Lawler, E. E., & Hall, D. T. (1970). Relationship of job characteristics to job involvement, satisfaction, and intrinsic motivation. *Journal of Applied Psychology, 54*, 305-312.

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

Mathieu, J. E., & Farr, J. L. (1991). Further evidence for the discriminant validity of measures of organizational commitment, job involvement, and job satisfaction. *Journal of Applied Psychology, 76*, 127-133.

Markon, K. E., Krueger, R. R., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach.

Journal of Personality and Social Psychology, 88 (1), 139-157.

Mumford, M. D., Connelly, M. S., Helton, W. B., Strange, J. M., & Osburn, H. K. (2001). On the construct validity of integrity tests: Individual and situational factors as predictors of test performance. *International Journal of Selection and Assessment, 9* (3), 240-257.

Mumford, M. D., Gessner, T. E., Connelly, M. S., O'Connor, J. A., & Clifton, T. C. (1993). Leadership and destructive acts: individual and situational influences. *Leadership Quarterly, 4*, 115-148.

Murphy, K. R. (2000). What constructs underlie measures of honesty or integrity? In Goffin, R. D., & Helmes, E. (Eds.), *Problems and solutions in human assessments*. MA: Kluwer Academic Publishers.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide.* Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Mårdberg, B., Sjöberg, A., & Henrysson Eidvall, S. (2000). *BasIQ begåvningstest. Manual*. Stockholm: Psykologiförlaget AB.

Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9*, 31-39.

Ones, D. S., & Viswesvaran, C. (1998). Integrity testing in organizations. In R. Griffen and A. O'Leary-Kelly (Eds.), *Dysfunctional behaviour in organizations: Violent and deviant behaviour*. Monographs in organizational behaviour and industrial relations, Vol. 23, parts A & B (pp. 243-276). Stamford, CT: JAI Press.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance, 18* (4), 389-404.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78* (4), 679-703.

Paterson, J. M., & O'Driscoll, M. P. (1990). An empirical assessment of Kanungo's (1982) concept and measure of job involvement. *Applied Psychology-An International Review, 39*, 293-306.

Reuterberg, S-E., & Gustafsson, J-E., (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement, 52* (4), 795-811.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviours: A multidimensional scaling study. *Academy of Management Journal, 38*, 555-572.

Robinson, S. L., & Bennett, R. J. (2000). Development of a measure of

workplace deviance. *Journal of Applied Psychology, Vol. 85* (3), 555-572.

Sackett, P. R., Berry, C. M., Wiemann, S. A., & Laczo, R. M. (2006). Citizenship and counterproductive behaviour: Clarifying relations between the two domains. *Human Performance, 19* (4), 441-464.

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: Sage Publications.

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49*, 787-829.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *ASA Proceedings of the Business and Economic Section,* 308-313.

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* (2), 262-274.

Scott, K. D., & McClellan, E. L. (1990). Gender differences in absenteeism. *Public Personnel Management, 19* (2), 229-253.

Shore, L. M., Newton, L. A., & Thornton, G. C. (1990). Job and organizational attitudes in relation to employee behavioural intentions. *Journal of Occupational Behaviour, 11*, 57-67.

Sjöberg, S., Sjöberg, A., & Svensson, C. (2012). *Measuring Integrity.* Technical Manual. Stockholm: Assessio International AB.

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

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

Spielberger, C. D., & Vagg, P. R. (1999). *Job Stress Survey, JSS*. Odessa, FL: Psychological Assessment Resources. Swedish version: *Job Stress Survey, JSS. Skala för arbetsrelaterad stress*. Stockholm: Psykologiförlaget AB, 2002.

Sverke, M., Hellgren, J., Näswall, K., Chirumbolo, A., De Witte, H., & Goslinga, S. (2004). *Job insecurity and union membership: European unions in the wake of flexible production*. Brussels: P. I. E. – Peter Lang.

The European Theft Barometer. (2005). *Monitoring the costs of shrinkage and crime for Europe's retailers*. Nottingham: Centre for Retail Research.

The European Theft Barometer. (2004). *Monitoring the costs of shrinkage and crime for Europe's retailers*. Fourth report to the retail industry. Nottingham:

Centre for Retail Research.

Vardi, Y., & Weitz, E. (2004). *Misbehaviour in organizations: Theory, research, and management*. Mahwah, NJ: Lawrence Earlbaum Associates.

Appendix A

List of figures and tables

Figure 1.1. Two-dimensional scatter diagram for the CWB categories in relation to the dimensions Task Relevance and Interpersonal – Organizational. From Gruys & Sackett (2003).

Figure 3.1. The measurement model of MINT.

Figure 7.1. Hierarchic model of MINT.

Table 1.1. Correlations of the FFM dimensions and job performance, OCB and CWB.

Table 3.1. The distribution of the initial MINT norm group (n=300) by age, sex and educational level compared with the population. www.scb.se

Table 3.2. Standardized factor loadings for the model (n=302).

Table 3.3. Descriptive statistics for the MINT factors.

Table 3.4. Correlations between MINT and NEO PI-R.

Table 3.5. Test-retest values for factors in MINT (Swedish sample n=97).

Table 3.6. Test–retest values for factors in MINT (Norwegian sample n=129).

Table 4.1. Hypotheses for the relationships between MINT and other measures of both deviant personality and Job involvement.

Table 4.2. Results for the relationships between MINT and other measures of both deviant personality and job involvement.

Table 4.3. Hypotheses for the relationships between MINT and HDS factors.

Table 4.4. Correlations between MINT and HDS factors (n=183).

Table 4.5. Means (SD) of the Norwegian grocery store sample.

Table 4.6. Hypotheses for the relationships between MINT factors and Social style, Conscientiousness and Emotional stability in MAP and ServiceFirst.

Table 4.7. Correlations between MINT and Social style, Conscientiousness and Emotional stability in MAP ([a]n=208) and ServiceFirst ([b]n=193).

Table 4.8. Descriptive statistics, reliability and restriction of range in the sample of Swedish traffic controllers and students.

Table 4.9. Observed correlation and operational validity in the sample of Swedish traffic controllers and students.

Table 4.10. Correlations between MINT and job performance (n=101) and CWB (n=44)

ratings in the grocery store sample.

Table 4.11. Correlations between MINT and job performance ratings in the retail sample (n=68).

Table 5.1. Number of assessments across language version, age and gender (global norm group, n=124 499).

Table 5.2. Confirmatory factor analyses, goodness of fit estimates and reliability estimates across language versions for the total norm group of n=124, 499.

Table 5.3. Standardized factor loadings for the model (n=124, 499).

Table 5.4. Multiple-Group Confirmatory Factor Analysis across the 10 language versions (n=124, 441).

Table 6.1. Predictive validity for a number of methods.

Appendix B

Development of MINT - Norwegian version

This chapter describes the translation and adaptation process of the Norwegian version of MINT from the original Swedish version (for a description of the development of the original Swedish version, please see Part 3). The translation and adaptation of items was carried out in parallel with the adaptation work of another of test: Predicting Job Performance, PJP (Sjöberg, Sjöberg & Forssén, 2006). In total, 206 items were translated from Swedish into Norwegian by a professional translator. In a second step, the items were reviewed by an external subject matter expert and suggested changes were put forward and discussed with Assessio's Norwegian experts within the field of test development. Necessary revisions of items were made and, in the third and last step, items were language reviewed in order to ensure high linguistic quality. After the finalized translation and adaptation of the 206 items into Norwegian, they were administered to a group of individuals that also constituted the first norm group for the Norwegian version of MINT.

The collection of this data was carried out in the spring of 2005. The 206 items were sent to 537 individuals from Assessio's database of test subjects, a database of subjects compiled as a representative sample of the Norwegian population. At regular intervals, these individuals responded to Assessio's occupational psychology instruments during the development and/or standardization process. In addition to the 206 items, NEO PI-R (Costa & McCrae, 2003) was also distributed. NEO PI-R is a self-reported personality measure based on the FFM of Personality. NEO PI-R thus measures Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. In total, 237 surveys were returned, giving a response rate of approximately 44%. Descriptive statistics concerning this norm group are shown in table B1.

In the work with the Swedish version of MINT 60 items out of the 206 items were included in the original Swedish version of MINT. These 60 items were selected based on a number of qualitative and quantitative criteria described in more detail in Part 3 of this manual. After the Norwegian data collection had been carried out, the items that fulfilled the same criteria as for the Swedish version of MINT were analyzed further. These analyses are described below.

Table B.1.
Age, gender, and education level distribution of the original MINT norm group (n=237), compared to the Norwegian population.

	Norm group	Population
Age group		
ears	23%	17%
ears	28%	21%
ears	21%	19%
ears	15%	17%
ears	6%	9%
ears	3%	6%
Unspecified	4%	
Gender		
Male	35%	50%
Female	65%	50%
Educational level		
Elementary and lower secondary	8%	33%
Upper secondary	16+30%	42%
University or college, short (up to and including four years)	32%	19%
University or college, long (more than four years and/or PhD)	14%	6%

Evidence Based on Internal Structure - Confirmatory Factor Analysis

In order to test the underlying structure of the Norwegian version of MINT, a similar factor analysis to that described in Part 3 was carried out (see Part 3 for an explanation of the technical concepts). The two indicators identified in the principle component analysis for each of the personality dimensions in the Swedish version were also used in the Norwegian version. As mentioned in Part 3, these indicators are only derived from a statistical perspective, and they are not included in the interpretation of test results. They have no implications for the interpretation of test scores. The theoretical model is presented in figure 3.1 (Part 3).

The results show a good fit between the theoretical model and the data collected (χ 7=21.581, p>.05; RMSEA=.094; GFI=.970; TLI=.952; CFI=.977). As mentioned in Part 3, an upper limit of .10 has been proposed for RMSEA as an acceptable fit, and .90 as the corresponding figure for the lower limit of the other measures. The factor loadings are presented in table B2.

Table B.2.

Standardized factor loadings for the MINT model

	Integrity	Interpersonal Orientation	Task Orientation
ES1	.95		
ES2	.85		
A1	.35	.67	
A2	.37	.75	
C1	.43		.77
C2	.40		.75

ES=Emotional Stability, C=Conscientiousness, A=Agreeableness

The results support the hypothesis that the dimension of Emotional Stability is a major component in the Integrity measure, as its indicators show loadings on the Integrity factor that are more than twice as high as those from the dimensions of Agreeableness and Conscientiousness. This can be compared to the Swedish version, where Emotional Stability was also the strongest component. The non-standardized factor loadings on the scales are shown to calculate individual values (factor scores). These individual factor scores can be calculated by applying regression analysis and multiplying the regression weights by the sum of the raw scores for the factors. This is essential if the individual scores are to be interpreted using the proposed hierarchic model for MINT. Descriptive statistics for each factor are shown in table B3.

MINT's relationship with the dimensions of the Five-Factor Model, measured with NEO PI-R (Costa & McCrae, 1985), has been investigated to ensure that the underlying structure of the Norwegian version of MINT is sufficient. The correlation pattern shown in Table B4 is the same as in the Swedish version, as the general Integrity measure correlates with all three FFM dimensions. Emotional Stability shows the strongest correlation. Interpersonal Orientation correlates with Agreeableness,

Table B.3

Descriptive statistics for the factors in MINT

	Min	Max	M	Md	SD	IC
Integrity	17.82	39.80	29.61	29.65	4.20	.98
Interpersonal Orientation	3.92	19.77	14.10	14.15	2.42	.84
Task Orientation	4.10	24.93	19.40	19.25	2.30	.88

Table B.4

Correlations between the factors in MINT and NEO PI-R

	Conscientiousness	Agreeableness	Emotional Stability
Integrity	.47*	.25*	.76*
Interpersonal Orientation	-.03	.42*	.05
Task Orientation	.49*	-.10	.09

*=p<.01, n=302

and Task Orientation correlates with Conscientiousness. Neither of these lower-level factors correlates significantly with Emotional Stability, nor with the marker for the other

dimension. The results thus show, in a convincing fashion, that the Norwegian version of MINT correlates as expected with the NEO PI-R factors.

Internal consistency and reliability

Reliability, in the form of internal consistency, measures the accuracy with which a test measures what it is intended to measure. The reliability of the Norwegian version was studied using the congeneric model for the calculation of internal consistency. With this model, both the variance differences in the scale factors and the error variance inherent in each factor can be taken into consideration. As described in Part 3, this model uses a different approach for obtaining a correct estimate of internal consistency than the traditional Cronbach's alpha method (which would generate an underestimate of the internal consistency of the general measure). The internal consistency was calculated using the formula presented in Part 3 (Reuterberg & Gustafsson, 1992).

The reliability of the general Integrity measure was estimated to be .98, and is therefore as high and as satisfactory as that of the Swedish version. The corresponding coefficients for Interpersonal Orientation and Task Orientation are .84 and .88, respectively – slightly lower than the coefficients in the Swedish version but still satisfactory. These results support the internal consistency of all of the MINT factors and clearly satisfy the requirements for the precision of the measurement.

Appendix C

MINT – Implications of the hierarchic model for interpretation

This section aim to explain, in more detail, the characteristics of pre-requisites for interpreting scores generated the type of hierarchical measurement model that MINT is based upon. The information in this appendix is intended for those who use MINT and interpret scores and presupposes a good orientation of psychometrics and measurement theory.

Test scores are only meaningful if they are considered in a relevant context, such as in relation to an applicable theoretical concept or for predicting a relevant criterion. If a test score cannot be linked to a theoretical framework, it is meaningless and cannot tell us anything as we do not know how to interpret it.

The theoretical framework of MINT is based on predicting tendencies towards counterproductive work behaviors, which is done by measuring the personality dimensions of Emotional Stability, Agreeableness, and Conscientiousness.

Many existing tests are constructed so as to generate a general interpretable score as well as scores on individual items, which can be interpreted in terms of profiles in order to provide information about the traits being measured.

Some tests can also be interpreted at a sub-test level. When a test is constructed on the basis of several interpretable levels of abstraction, the theoretical model or framework is called a hierarchic one.

The levels in the hierarchic model are called strata, meaning level or layer. In order for traits to be interpreted at different strata, the test requires sub-tests that are specially constructed to measure traits at a lower stratum while also accounting for how they influence traits at a higher stratum.

It is important to point out that there is a difference between a total score that is the sum of sub-test scores and a total score constructed to describe a theoretical model. A general and higher-level score must always be linked to a meaningful theoretical concept such as Integrity. This means that a test developer has to be systematically selective when creating the higher-level factor, as choosing items for the factor at random would be fruitless.

The general factor in MINT is clearly defined as Integrity. The variation in this factor is extracted from the two lower-level factors, Interpersonal Orientation and Task Orientation. In the test development process the traits that are independent of the general factor can be identified, which makes it possible to interpret scores for strata other than the general one. It is, however, difficult to interpret and comprehend scores that represent different hierarchic strata when these have shared variance, i.e., when they overlap. In such cases it is impossible to know whether the score in question refers to a

higher or lower stratum. To avoid this problem, it is necessary to systematically extract the residuals to be found in the data and thus isolate the factors being defined.

Interpreting test scores based on residuals from hierarchic models presents certain problems, however. Even if the test scores measure real traits, from the conceptual perspective they are artificial and they do not directly reflect the real traits of the individual. This is because the trait is affected by factors on several different hierarchic strata. A person who, for example, has a low score on the residual based score Interpersonal Orientation may show a high degree of agreeableness if she or he has a high score on the general Integrity factor.

Since residual scores are basically difference scores, scales of this type tend to have lower reliability than raw-score scales. Accordingly, the interpretation of lower stratum scores in a hierarchy becomes successively more uncertain. One example of this is the interpretation of scores at the sub-test level. If one succeeds in attaining high reliability and low inter-correlation (both of which are problematic when interpreting difference scores), then a high degree of specificity has been obtained. Specificity refers to the variation in test scores that is reliable and unique to the scale or sub-test in question. This appears in the interpretation as the frequency of observable differences between strengths and weaknesses. Tests with low specificity make it harder to identify strengths and weaknesses (Gustafsson, 2002). The model on which MINT is based proposes that the relationships among individual Integrity differences can be described using a hierarchic model with two strata.

It is, in the case of hierarchic models, important to distinguish between two types of conceptual perspectives. The “bottom-up” perspective implies that the lower-strata factors are indivisible and that the higher-strata factors have no influence on the lower-strata. The “top-down” approach, as represented by the MINT-model, means that the lower-strata factors are separated from the variance derived from the higher-strata factors, which in this case means that Integrity is not contaminated by the variance of Interpersonal Orientation and Task Orientation.

The factors at the lower level strata are thus divided into two parts, one with unique variance for the lower factor and another that can be assigned to the higher stratum. These two perspectives on the nature of factors at the lower-strata level are essential to the measurement and thus the interpretation, of these factors.

When test scores are interpreted using a hierarchic model, the rule is that the lower down in the hierarchy an interpretation is made, the lower the “referent generality” of the score. Referent generality refers to how much of the frame of reference is covered by the construct in question as well as to the variation in behaviors or mental activities to which the construct is related and the extent to which the construct relates to them. The fact that measurements and constructs differ with respect to referent generality has considerable similarities to the hierarchic structure of individual differences. It is not readily apparent how constructs such as low and high referent generality should be measured out, but it is to be expected that the methods will differ.

To measure constructs with high referent generality, it is necessary to use heterogeneous tools in order to avoid what is termed construct under-representation, a measurement issue in which none or not enough of the aspects of the underlying construct have been

captured. Potential problems with using a homogenous test for measuring a construct with high referent generality is that low validity may arise via construct irrelevance and the test always measures several dimensions. This construct-irrelevant variance is not a problem in well-developed heterogeneous measures because of the effect of aggregation.

A number of paradoxical problems may arise when a hierarchic model is used to measure a construct. This gives, in conjunction with the differences in interpretable results generated by the model, rise to several questions about measurement methods. The model is only meaningful when measures of constructs with low referent generality are interpreted as indicators of non-observable traits or abilities. If, however, a test score is regarded as a sample from a domain, there is no point in dividing the variation into separate parts depending on the degree of generality. Test scores using the model in question here are regarded as measuring a unitary construct.

A hierarchic model is more restrictive (have more constraints) than an oblique one (less constraints) and is based on further assumptions. If, however, the hierarchic model fits the data and the assumptions are satisfied, this type of model tends to be more informative.

The process of selection and recruitment has been focused on identifying high-performing individuals who will enhance the value of the companies and organizations that employ them. Personnel selection has not focused enough on pinpointing individual probability of engaging in dishonest or unreliable behavior. Theft, unsanctioned absenteeism, and other forms of CWB result in considerable financial losses worldwide, and many companies invest heavily in expensive monitoring systems, often with less than satisfactory results.

MINT is a personality-based test developed for indicating the likelihood of candidates' engaging in behavior harmful to the organization. MINT is based on modern industrial psychology research and is designed to measure the extent of individual integrity and thus predict tendencies towards counterproductive work behavior. This manual contains information regarding the theoretical background of MINT, the extensive development process, the research findings providing evidence underpinning the quality of MINT, and guidelines for administration and interpretation of test results.