Diekmann, J., & König, C. J. (in press). Personality testing in personnel selection: Love it? Leave it? Understand it! In I. Nikolaou & J. Oostrom (eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice*. Hove, UK: Psychology Press.

PERSONALITY TESTING IN PERSONNEL SELECTION: LOVE IT? LEAVE IT? UNDERSTAND IT!

Janina Diekmann

Cornelius J. König

Universität des Saarlandes

LOVE IT OR LEAVE IT: THE VALIDITY DISCUSSION OF PERSONALITY TESTING

The quality of selection procedures is judged primarily by looking at predictive validity results, as the prediction of performance at work is clearly the most important issue for the practice of personnel selection. Based on these results, researchers have made recommendations to improve methods such as the interview (Kepes, Banks, McDaniel, & Whetzel, 2012; McDaniel, Whetzel, Schmidt, & Maurer, 1994) or have contributed to the abandonment of methods with no predictive quality, such as graphology (Driver, Buckley, & Frink, 1996).

Although most established selection methods such as mental ability tests or assessment centers have been found to be valid, the situation is significantly different with regard to personality testing. Discussions about whether personality tests are valid instruments began 60 years ago, with studies finding moderate but profession-dependent results at best (Ghiselli & Barthol, 1953) and generally troubling results at worst (Guion & Gottier, 1965). This discussion was intensified when, in their Big Five meta-analysis, Barrick and Mount (1991) found

conscientiousness to be the only trait that was generally and at least moderately predictive of work performance, whereas the other four Big Five traits showed only small correlations which varied between different occupations. Currently, the debate about whether or not one should use personality tests in personnel selection procedures is dominated by two perspectives, both of which are supported by good arguments.

On the one hand, there are those advocators of personality tests who "love it" (e.g., Bartram, 2004; Ones, Viswesvaran, & Dilchert, 2005): The findings of Barrick and Mount (1991) as well as further meta-analyses (Hurtz & Donovan, 2000; Salgado, 1997) and a second-order meta-analysis (Barrick, Mount, & Judge, 2001) are used to argue that there are consistent correlations and to support the central role of conscientiousness and (in part) of emotional stability in predicting job performance. Although the other Big Five traits were not related to overall work performance, they were able to predict specific professions or criteria. Numerous studies and meta-analyses explored the personality-performance relationship. For example, they examined the longitudinal impact of the Big Five on career success (Judge, Higgins, Thoresen, & Barrick, 1999) using specific criteria such as job satisfaction (Judge, Heller, & Mount, 2002), or specific occupations or roles such as social professions (Blickle & Kramer, 2012) or leadership roles (Judge, Bono, Ilies, & Gerhardt, 2002). In these studies and meta-analyses, researchers frequently found high validities (for a detailed overview of research, see Rothstein & Goffin, 2006). Moreover, advocates of personality measures in personnel selection argue that personality traits particularly predict

typical performance, whereas general mental ability particularly predicts maximum performance (e.g., Marcus, Goffin, Johnston, & Rothstein, 2007).

Some debate within the "love it" group concerns the preference for broad or narrow personality traits: While some researchers recommend using all relevant personality traits together to maximize validity (Barrick & Mount, 2005) or using so-called compound personality traits (Ones et al., 2005; Ones & Viswesvaran, 1996) to predict overall job performance, others believe that narrow traits (and specific criteria) with well-considered theoretical assumptions of the trait-performance relationship will lead to better predictions (Dudley, Orvis, Lebiecki, & Cortina, 2006; J. Hogan & Holland, 2003; Tett, Steele, & Beauregard, 2003). Nevertheless, there is a group of advocates of personality testing who feel that "personality matters" (Barrick & Mount, 2005, p. 359).

On the other hand, there are researchers (e.g., Murphy & Dzieweczynski, 2005) who are more drawn to a "leave it" position. They argue that the correlations found in the above-mentioned meta-analyses are quite small and that there is a lack of convincing general theories that relate personality constructs to job performance (Murphy & Dzieweczynski, 2005). Even those who see themselves as more or less impartial (Morgeson et al., 2007) are concerned about the low validity, which is sometimes "pimped" by corrections for predictor unreliability (Campion in Morgeson et al., 2007). They therefore advise against the use of most personality tests in selection contexts or recommend the additional use of tests of general mental ability. In addition to this validity issue, critics often also point to the problem of faking. There is little doubt that applicants can, and actually do, fake answers when completing a personality test (e.g., Birkeland, Manson,

Kisamore, Brannick, & Smith, 2006). Although some researchers consider this to be unproblematic (e.g., J. Hogan, Barrett, & Hogan, 2007), faking does seem to change rank orders and therefore affects actual selection decisions (Stewart, Darnold, Zimmerman, Parks, & Dustin, 2010). Common correction methods such as lie scales do not provide a satisfactory solution to the problem either (e.g., Campion, Dipboye and Schmitt in Morgeson et al., 2007), although assessors believe that they do (Robie, Tuzinski, & Bly, 2006).

As this "love it or leave it" debate continues, so too does the use of personality tests (Bartram, 2004). Research clearly shows that organizations use personality tests: Personality testing is quite popular in Belgium, France, Greece, Ireland, Netherlands, Portugal, Spain and Britain (Bruchon-Schweitzer & Ferrieux, 1991; Eleftheriou & Robertson, 1999; Hodgkinson, Daley, & Payne, 1995; Hodgkinson & Payne, 1998; Ryan, McFarland, Baron, & Page, 1999; Schuler, Frier, & Kauffmann, 1993; Shackleton & Newell, 1994; Williams, 1992; Zibarras & Woods, 2010). It is also known to be a regularly used instrument in several other countries such as Germany, Italy, Scotland, and the USA (Harris, Dworkin, & Park, 1990; Piotrowski & Armstrong, 2006; Ryan et al., 1999; Rynes, Orlitzky, & Bretz, 1997; Scholarios & Lockyer, 1999; Schuler et al., 1993; Schuler, Hell, Trapmann, Schaar, & Boramir, 2007; Shackleton & Newell, 1994).

UNDERSTAND IT: THE PRACTICE OF PERSONALITY TEST USE

Against this background, we believe that it is time to set out on a new research path that concentrates on the practice of personality test use in organizational settings. Apart from the highly important questions of validity and faking,

research should find out which tests are being used in which ways and for what reasons in order to optimize our recommendations to practitioners.

To our knowledge, only few authors have been interested in which tests are actually used by organizations or (Industrial and Organizational) psychologists in general (e.g., Brown, 1999; Evers et al., 2012; Furnham, 2008; Muñiz & Fernández-Hermida, 2010; Muñiz, Prieto, Almeida, & Bartram, 1999; Ryan & Sackett, 1987, 1992; Sneath, Thakur, & Madjuck, 1976; Steck, 1997). Even fewer have explored which tests are used for personnel selection in particular (Berchtold, 2005; Di Milia, 2004), even though the criticism has been raised that personality tests are "poorly chosen" (Murphy & Dzieweczynski, 2005, p. 343). Therefore, we ask, which tests are really used in selection procedures? Is it the often-mentioned (e.g., Hülsheger & Maier, 2008; Murphy & Dzieweczynski, 2005) Myers-Briggs Type Indicator (MBTI), a Big Five instrument or do other tests play an important role?

Taking into account those studies which survey general test use in organizations and those conducted by Industrial and Organizational psychologists without a specific focus on selection (Berchtold, 2005; Brown, 1999; Di Milia, 2004; Furnham, 2008; Muñiz & Fernández-Hermida, 2010; Ryan & Sackett, 1987, 1992), the evidence so far shows that the tests most frequently mentioned across studies are the 16 Personality Factor Questionnaire (16 PF), the Myers-Briggs Type Indicator (MBTI), the Occupational Personality Questionnaire (OPQ), the Minnesota Multiphasic Personality Inventory (MMPI), the Big Five Personality Inventory (NEO), the California Psychological Inventory (CPI), and the Thomas Assessment/ Personal Profile Analysis (PPA). This is in line with information

from job websites or free personality test websites listing the supposed main personality tests (Donston-Miller, n.d.; Free Personality Test, n.d.). However, many more tests are mentioned in these studies, reflecting the huge variety of tests which exist (there are an estimated 2,500 publishers in the United States alone, see R. Hogan, 2005; Hough & Oswald, 2005; Psychometric Success, 2013), operating in a $500 million industry (Psychometric Success, 2013).

A closer look at the two studies that exclusively considered tests used in personnel selection procedures (Berchtold, 2005; Di Milia, 2004) reveals that there may be differences in test use that could be due to regional preferences or the fact that some tests have only a national range. Examining personality test use in selection procedures of Australian organizations, Di Milia (2004) found not only the OPQ, MBTI, NEO and 16PF to be frequently used, but also questionnaires, such as the Personal Characteristics Inventory (PCI), the Fifteen Factor Questionnaire, the Occupational Personality Profile (OPP) and the DISC. Swiss organizations (Berchtold, 2005) also use the MBTI, 16PF, Thomas Assessment, OPQ and NEO, supplemented by tests like the Master Person Analysis (MPA), Insights Discovery or MDI, the Bochum Inventory for profession-related personality description (BIP), the DISG (DISC) or the Herrmann Brain Dominance Instrument (HBDI). All in all, 173 companies were found to use 52 different personality tests for selection purposes in Switzerland.

To complement the existing studies and to survey the current state of personality testing in Germany, we conducted our own study, questioning HR practitioners in

companies of all sizes across Germany[1]. We found that personality tests were used in 15.1% of the surveyed companies (see Figure 1 for the application frequency of all selection methods). This is slightly less than the 20% which has usually been found in Germany over the last twenty years (Schuler et al., 2007) but can probably be explained by the fact that we also had smaller companies in our sample (41.6% had fewer than 500 employees). Respondents found personality tests to be moderately useful for promotion, planning of personnel development activities, assistance in team development activities and for personnel selection at the employee level, and to be somewhat more useful for personnel selection at the management level (see Figure 2). Actual test users found personality tests to be significantly more useful for all purposes than did non-test users. Concerning the question of which personality tests were used, in accordance with the studies mentioned above, we found a huge variety of different methods, including Insights Discovery or MDI, the BIP, the PPA, the 16 PF, the DISC, the Hogan Personality Inventory (HPI) and the Predictive Index (PI). For an overview of all mentioned tests, see Figure 3.

[1] We randomly called 769 companies; in 605, we were able to talk to employees or managers who worked in conducting the selection process. 403 people were interested in participating in the study and were invited to take part in the online survey by e-mail. A total of 166 persons (37.3% male, 56.6% female, 6.0% did not specify their gender) actually completed the whole survey (292 dropped out). Respondents had been in their current jobs for an average of 12.7 years (*SD* = 8.8) and most (71.7%) had a university education, with the majority being trained in business administration (58.0%) and only 5.0% in psychology. On average, they had been involved in 41.3 selection procedures during the last year (*SD* = 111.9), and a total of 77.1% had decision-making rights concerning the choice of selection methods. Companies had approximately 904.4 (*SD* = 1608.9) employees (7.8% had up to 50 employees, 16.3% between 51 and 250, 24.1% between 251 and 500 and 39.2% had over 500 employees; 12.7 did not answer this question), 72.3% were operating internationally, mostly in manufacturing, wholesale and the retail trade, financial and insurance activities or personnel services. The survey consisted of three main parts: First, we wanted to know which selection methods the companies used. Second, we asked participants about the purposes for which they found personality tests to be useful. Third, we concentrated on personality test use in personnel selection and asked for preferences of 15 different criteria that can be used to distinguish these tests.

These studies provide a first impression of the world of selection by personality testing. The MBTI is clearly one of the most frequently used personality tests; it is not only mentioned in various different studies, but is also high in the rank order of frequently used tests within these studies. Although the NEO personality inventory is also used in several countries, it generally ranks (far) below the MBTI (Berchtold, 2005; Di Milia, 2004; Furnham, 2008). This points towards the so-called research-practice gap in personnel selection, which describes the fact that research contents and recommendations of researchers are not always in line with the current implementation practice (e.g., Rynes, Giluk, & Brown, 2007): While we as researchers focus very much on the Big Five and instruments measuring these personality traits, practitioners seem to prefer other instruments like the MBTI although there is great doubt about its validity (e.g., Ones et al., 2005). Moreover, the three studies concentrating on selection (Berchtold, 2005; Di Milia, 2004; and our own study) clearly show that there is much more to personality testing than the MBTI and NEO (surprisingly, neither the MBTI nor the NEO are among the tests used in Germany). These three studies demonstrate the huge variety of personality tests in existence and use, some of which are restricted to certain countries/languages (for example the BIP, which was developed in Germany) and some of which are probably not appropriate in selection procedures.

Personality tests by comparison: What's it all about?

Let's take a closer look at the above-mentioned personality questionnaires: In the following section, we describe and discuss several important criteria beyond standard criteria such as reliability and validity (because previous research has

shown that these criteria are not the only criteria important to practitioners, see, e.g., König, Klehe, Berchtold, & Kleinmann, 2010) that concern characteristics of the personality test and its presentation of results, aspects of application, description of quality criteria and the process of finding a personality test that might influence the allure of often-used personality tests for practitioners.

Test characteristics and presentation of results

A first distinguishing criterion is whether the test results in a personality type (e.g., MBTI, DISC, HBDI) or in a dimensional personality profile (e.g., 16 PF, NEO, BIP, MPA). Whereas dimensions reflect the idea that a person usually shows all traits to a certain degree on a continuous scale, types group people into discrete classes (Gangestad & Snyder, 1985). The measurement of dimensions is widespread in psychological research, but there seems to be a nagging distrust of types, which are often seen as an (over)simplification, a trigger of stereotyped thinking, or even pure invention (e.g., Gangestad & Snyder, 1985). Moreover, it is often difficult to decide where to set theoretically or empirically meaningful cut-off points that assign a person to one type or the other without misclassification, and there is the general question of whether a person can exclusively be assigned to one type (Robins, John, & Caspi, 1998; York & John, 1992). Even defenders of the MBTI believe that people can belong to more than one type and that the test alone will not find the “right” type, but that one needs to talk to the test taker (Bayne, 2005). At the same time, type-tests may have advantages over dimensional personality tests. For example, the reduction of information and complexity into one type may be easier to interpret and therefore more appealing. Whereas a dimension-based test reports many scales with a person varying on all

of these scales, a type includes all information in an economical manner and makes it easier to differentiate between applicants. A schema-like categorization system may also match the human knowledge structure of cognitive schemata (Smith & Queller, 2008) and limitations of cognitive capacity (Tversky & Kahneman, 1974). Both approaches to personality testing may thus have their advantages and disadvantages (and may not only co-exist but even benefit from each other; Robins & Tracy, 2003). In our survey, we also asked the practitioners whether they preferred dimension-based personality tests or type tests and whether they preferred the results to be aggregated into one comparable value or to be presented in multiple comparable facets[2] (see Figure 4). Results of one-sample *t*-tests, testing for differences to the scale middle of 3.5, showed a significant[3] preference for types rather than dimensions (M = 3.89, SD = 1.53) on the one hand and a significant preference for facets rather than an aggregation to one value (M = 3.97, SD = 1.51) on the other. This indicates that a mixture of both types of results may be most attractive. Interestingly, actual test users (n = 28, M = 3.21, SD = 1.62) preferred dimensions, whereas those who did not use personality tests (n = 138, M = 4.03, SD = 1.48) showed a strong preference for types. This suggests that a certain expertise concerning personality tests leads to a difference in preferences (but given the small sample of test users, this result should be treated with caution).

[2] Each preference item had two poles on a one- to six-point scale, e.g. “Would you prefer…” and “… a dimensional representation of measured traits” on one pole and “… the aggregation of measured traits in types” on the other pole.
[3] Whenever we speak of significance, we mean at least $p < .05$.

A second criterion concerns the report. The user is confronted with a type or a profile that he needs to interpret and compare with an ideal type or profile and/or other applicants. On the one hand, this compact alternative has the advantage that the user does not have to read a long report but can focus on the aspects that are important to him or her. On the other hand, if a practitioner is interested in an interpretation, he or she is left alone with this task. That can be a considerable problem if he or she is not a psychologist with appropriate training in test interpretation. A manual can be very helpful, but may not always be easy to understand. The other option, which is usually provided automatically with online test versions, consists of detailed narrative reports, which offer the advantage of an extensive, easy-to-understand and quick evaluation that is less prone to mistakes regarding subjectivity and the difficult task of simultaneously processing several variables (Bartram, 1995; Snyder, 2000). So-called computer-based test interpretations (CBTIs) have been used and discussed for decades now, especially in clinical psychology (e.g., Butcher, Perry, & Dean, 2009; Fowler, 1985). They are almost standard in reports of commercial test publishers as well as in science-based personality tests like the NEO (at least in some versions: in Germany, a narrative report is available for the NEO-PI R+, but not for the NEO FFI) and the BIP. These narrative reports can differ in terms of various aspects, for example the extent to which text and graphs are integrated, the involvement of interpretation of configurations and interactions, or the possibility to adapt a test to the context (e.g., development or selection) (Bartram, 1995). The gain of being provided with an interpretation is often bought with the uncertainty about accuracy and validity of these interpretations, and narrative reports of different

tests probably differ in their accuracy (Kellett, McCahon, & James, 1991). Especially in the case of tests from commercial publishers, it is often difficult to evaluate how these interpretations are generated, which statistical methods and which interpretive rules or algorithms are used to combine test results and text modules, or how these text modules were developed. Frequently, the report cannot be modified or adapted to the current test context (Bartram, 1995), and even if this were the case, it is questionable whether non-trained personnel staff would be able to do so appropriately. Some reports may even take advantage (knowingly or not) of the Barnum effect: They make such broad statements that people usually feel that the report is accurate, scientifically precise and offers good reasons for decisions, but it is actually too general for a practitioner to make well-grounded judgments (Guastello, Guastello, & Craft, 1989; Guastello & Rieke, 1990; Snyder, 2000). Unfortunately, there is barely any research concerning the issue of narrative reports in an organizational context or addressing the huge variety of tests in use. Our survey found a significant preference for a profile (M = 3.82, SD = 1.41) rather than a narrative report. Perhaps there is a stronger need for quick comparisons in selection procedures, making narrative reports less important than, for example, in consulting and development activities.

Another criterion concerns the development and background of a test. Although test development can have different backgrounds, there seem to be two major variations: A personality test can be based on a personality theory or on a statistical approach. The MBTI, for instance, is an example of the theory-based approach. It was developed by Katherine Briggs and her daughter Isabel Briggs Myers, under the influence of C. G. Jung's typology (Briggs Myers & Myers,

1993; Jung, 1960). Another influential theory concerns William Marston's (1979) behavioral types – Dominance (D), Inducement (I), Submission (S), and Compliance (C). This led not only to the DISC assessment but also to the development of other personality tests such as the Personal Profile Analysis. Insights MDI used both models as a background (Spieß, Eckstaller, & Woschée, 2004). The HBDI, by contrast, was developed by Ned Herrmann (1989), taking into account brain hemispheres theory (e.g., Mintzberg, 1976) and MacLean's (1985) theory of the "triune brain." It results in four thinking styles, reflected by a four-quadrant brain model. Another (main) way of developing a test, which is favored by most scientists, is based on a statistical approach. The NEO, for example, has such a statistical, non-theoretical background. It is based on the so-called lexical approach, and the Five Factors measured in this test were developed through factor analytical methods (e.g., McCrae & Costa, 1997). There are good reasons why practitioners might be attracted by both approaches. On the one hand, the statistical, factor analytical method is an empirical one. This alone may give a personality test a serious appearance, meeting needs of legal security. On the other hand, people have a strong need for explanations, in particular explanations of human behavior (Keil, 2006; Lombrozo, 2006; Malle, 2004), and although the above-mentioned theories probably do not deliver such an explanation, they may serve as compensation. At least they suggest that there is more to a test than just a description of traits, and people may usually not require a scientifically tested theory (Keil, 2003, 2006; Rozenblit & Keil, 2002; Wilson & Keil, 1998). Moreover, such a general structure as derived in the NEO may not meet practitioners' requirements, as it does not refer to work-related applications such

as personnel selection (Hough & Oswald, 2005). In our study, we also asked practitioners whether they preferred a theory-based or statistically-based development of traits. Results indicate that practitioners significantly favored a statistically-based development (M = 3.87, SD = 1.32). It thus appears that practitioners do understand the importance of a scientific approach.

Mode of delivery

The most apparent point of application concerns the presentation of the test: the "classic" paper-and-pencil form and the application at the computer with a local test system or via the internet. The advantages of an electronic application are obvious: The testing material as well as test and response time can be controlled, items can be easily adapted, application and evaluation of results are highly objective, printing costs and unwieldy paper copies are eliminated and feedback is available in an instant (Bartram, 2000; Lievens & Harris, 2003). What is more, the internet provides a high flexibility, as applicants can be tested independently of place and time (Lievens & Harris, 2003). At the same time, there are some difficulties that have to be faced, which have been discussed to different degrees in the literature: Problems such as connection problems during internet testing or a lack of computer or internet access are likely not as serious as they were a couple of years ago, but are probably still an issue. Moreover, practitioners should keep in mind that people have different levels of affinity to computers and the internet, which might lead to discrimination of some groups such as older people or ethnic minorities (Bartram, 2000). The ethical question of security of data transfer and confidential management of test results also remains important. A further question concerns the transferability of paper-and-pencil tests to the computer format.

Currently, computer-based tests are usually still the same as their paper-and-pencil predecessors (Bartram, 2000). However, it is necessary to ensure that the psychometric properties are the same for two reasons: First, companies may use both versions and compare applicants undergoing paper-and-pencil and computer-based assessments, and second, equivalent scores are required in order to use the norms traditionally gleaned from the paper-and-pencil version (Meade, Michels, & Lautenschlager, 2007). Most studies found encouraging results (Bartram & Brown, 2004; Chuah, Drasgow, & Roberts, 2006; Salgado & Moscoso, 2003) and even some benefits of web-based testing (e.g., more normal distribution or higher reliabilities, Ployhart, Weekley, Holtz, & Kemp, 2003). Nevertheless, there are differences (e.g., concerning means, Ployhart et al., 2003), and Meade et al. (2007) warn that comparability cannot be taken for granted. Practitioners in our sample strongly preferred a computer application over a paper-and-pencil application (M = 2.38, SD = 1.62), but were indifferent as to whether the test should be applied via the internet or on-site (M = 3.51, SD = 1.93). Moreover, there is no preference regarding who (the company/ the practitioner or the test publisher) evaluates test results (M = 3.51, SD = 1.97), meaning that the focus seems to be on an automated process and not on the way in which this automation is delivered (by an external provider, on-site or via the internet). On the other hand, actual test users do prefer an application by internet (n = 28, M = 2.61, SD = 1.77) compared to non-users (n = 138, M = 3.70, SD = 1.87), meaning that people who already use personality tests seem to perceive the advantages of this medium.

Declaration and description of quality criteria

Quality criteria, especially measures of reliability and validity (which we will subsume with the term quality criteria in the following), are very important to researchers, who consequently present these measures in extensive test manuals, as do some commercial test publishers. However, considering the huge amount of personality tests available, the extent to which publishers are interested in measuring and providing quality criteria likely varies. Besides, the existence of quality criteria does not mean that practitioners have access to such information before buying a test. There is a huge variety of ways in which quality criteria can be reported: According to our experience, information on publishers' or distributors' websites is (a) seldom extensive, (b) often only brief, (c) sometimes only available on demand or by buying the manual, or (d) not available at all. A brief description of quality criteria may be an alternative which is more convenient to practitioners, as they probably do not have the time to read long manuals. In our study, practitioners significantly preferred succinct statements about quality criteria rather than extensive information (M = 4.42, SD = 1.36), and brief information about benefits rather than detailed reports (M = 2.88, SD = 1.49). Nevertheless, they do not seem to be naïve in terms of believing these statements, as they strongly prefer to check this information rather than trusting the declarations of the author (M = 2.72, SD = 1.49). At the same time, actual test users significantly preferred more detailed reports about the benefits of a certain test (n = 28, M = 3.57, SD = 1.69) compared to non-users (n = 138, M = 2.74, SD = 1.41), whereas there was no difference concerning the length of quality criteria information. Consequently, there is perhaps more to selling personality tests than

numerical criteria. Moreover, no significant results were found regarding the question of whether practitioners would prefer a theoretical explanation of why the measured traits should be important for their employees' professional performance compared to statistical measures (M = 3.61, SD = 1.53). Once again, this indicates that both kinds of information are needed, and more is needed to convince practitioners of the benefit of personality tests in personnel selection than the scientists' mere focus on proving validity data.

Finding a personality test

An additional criterion that distinguishes personality tests is where and how practitioners can find information about them. In our survey, practitioners significantly preferred to inform themselves by searching websites and flyers rather than professional journals and magazines[4] (M = 3.26, SD = 1.56). Answers to an open question concerning sources revealed that most used the internet (35.5%), information and recommendations from their personal network (12.7%), and professional (HR-related) magazines (12.0%). They strongly favored tests used by many companies rather than tests that set them apart from other companies (M = 2.69, SD = 1.29), a confirmation of the finding of König et al. (2010). Most commercial publishers seem to take advantage of this practice of using recommendations, by citing referees who predominantly work in well-known companies on their websites. These references do not necessarily contain any information about the frequency and reason of use in the respective company. Moreover, our sample preferred to compare a small pre-selection of tests rather

[4] In our survey we used the German word „Fachzeitschrift" that includes professional and peer-reviewed journals as well as magazines.

than many different tests (M = 4.70, SD = 1.28), even more so when they were not currently using a personality test (n = 138, M = 4.80, SD = 1.22) than when they were already using one (n = 28, M = 4.21, SD = 1.45), which might not be too surprising considering the huge amount of tests available.

Another criterion that may affect the selection of a personality test is whether practitioners have to gain a certificate to use a special test (i.e., some publishers do not sell their inventories or at least part of them to people who are not trained and certified, and others offer training as an additional service, i.e. the MBTI the HBDI certification). Practitioners in our sample did not have a particular preference for or against certification (M = 3.60, SD = 1.72), although actual test users prefer certification (n = 28, M = 2.89, SD = 1.77) compared to non-users (n = 138, M = 3.74, SD = 1.68). Offering training seems reasonable, at least for non-psychologists, who have probably not had such training during their education, because otherwise, there is no guarantee that users are really informed about the proper application and interpretation of results.

An additional factor which is important in the decision-making process but is not covered in our survey[5] concerns the promotion of personality tests. Promotion strategies may differ to various extents: For example, there is "classic" advertisement in HR journals or stands at HR fairs. In addition, some may rely on a factual strategy, while others may (consciously or unconsciously) emphasize special characteristics of their tests in the sense of a unique or emotional selling proposition (e.g., the HBDI stresses a metaphorical connection to the brain)

[5] Practitioners probably do not know anything about promotion strategies and they cannot consciously evaluate the effect of promotion on their decision.

(Herrmann International, n.d.) or point out the model of personality upon which the test is based (i.e., they can highlight that their tests rely on well-established models, for instance the MBTI on Jungian theory). Whatever their strategy, commercial test publishers probably invest a lot in their promotion strategies in order to stand out from the crowd of personality tests.

FUTURE PROSPECTS AND CONCLUSIONS

We were able to show that – at least in Germany – there is definitely more to personality testing than just the Big Five or MBTI, and we believe that it is necessary to gain a broader overview, an international appraisal of actual personality test use rather than to focus solely on particular single measures. Not only is there a large range of personality tests offered to practitioners, but many of them are also in use in the context of personnel selection. We discussed the influence of different criteria on the decision-making process, such as certain test characteristics, the different ways of presenting results or aspects of application. We believe that a deeper understanding needs to be gained of this decision-making process, the requirements and needs of practitioners and the advantages and disadvantages of the manifold alternatives. For instance, we know nothing about the quality and actual handling of narrative reports in the selection process. Moreover, we concentrate strongly on dimension-based tests without even considering whether types might somehow meet practitioners' needs. Although the development of the Big Five certainly has great advantages in terms of comparability, it may not fit with categories of practitioners in personnel selection. Other traits or competencies may be more important to them because they are meaningful in terms of showing an intuitive theoretical relation to job

performance. We need to find out a lot more about how personality tests are actually used, what may influence the decision for implementing personality tests in the selection process, and how attitudes to personality tests may change before and after this implementation. Our survey was only a first attempt to learn something about practitioners' needs and requirements concerning the use of personality testing in personnel selection and to initiate a change in perspectives – away from believing that reliability and validity are the only criteria important to practitioners towards an understanding of the existence of multiple influences. Possibly, there may be many more criteria according to which personality tests can be differentiated (e.g., whether items relate to organizational contexts, to clinical contexts or neither, the costs of one or several applications, the number of dimensions or types measured, item format, how dimensions and types are named, whether they are special tests for different roles like leaders or salesmen, etc). It will be the task of future research to use this new perspective to develop arguments for propositions and specific hypotheses concerning the influence of different criteria to the decision-making process of practitioners.

Moreover, it is not enough merely to survey practitioners, as questionnaires are prone to socially desirable responding (as it probably happened in our question whether practitioners would prefer to check quality information or trust declarations of the author). Rather, practitioners' decision making needs to be experimentally analyzed.

In addition, it may be necessary to take a step away from pure research and to try to diminish the research-practice gap in personnel selection. One such step may be to simplify the search and comparison of different personality tests by setting up

national websites that list personality tests categorized according to their benefit for different purposes (e.g., development, selection, general assessment of personality, …) and provide the most important information and professional and independent evaluations of common tests. Another step may be to develop training programs for different personality tests in different organizational contexts in order to improve actual test use.

As personality tests continue to be used – no matter how scientists evaluate this – it is important to understand this use and make adequate recommendations and offers to practitioners. We should not ignore the needs and requirements of practitioners and should therefore try to adapt our research priorities accordingly.

REFERENCES

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, *18*, 359–372. doi:10.1207/s15327043hup1804_3

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9–30. doi:10.1111/1468-2389.00160

Bartram, D. (1995). The role of computer-based test interpretation (CBTI) in occupational assessment. *International Journal of Selection and Assessment, 3*, 178–185. doi:10.1111/j.1468-2389.1995.tb00025.x

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, *8*, 261–274. doi:10.1111/1468-2389.00155

Bartram, D. (2004). Assessment in organisations. *Applied Psychology: An International Review*, *53*, 237–259. doi:10.1111/j.1464-0597.2004.00170.x

Bartram, D., & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ 32i scores. *International Journal of Selection and Assessment*, *12*, 278–284. doi:10.1111/j.0965-075X.2004.282_1.x

Bayne, R. (2005). *Ideas and evidence : critical reflections on MBTI theory and practice*. Gainesville, FL.: CAPT.

Berchtold, M. (2005). Häufigste Auswahlverfahren in der Personalselektion [Most frequently used selection methods in personnel selection]. *HR Today, 12/2005*, 37–39.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*, 317–335. doi:10.1111/j.1468-2389.2006.00354.x

Blickle, G., & Kramer, J. (2012). Intelligenz, Persönlichkeit, Einkommen und Fremdbeurteilungen der Leistung in sozialen Berufen [Intelligence, personality, income, and job performance assessments by others in social welfare jobs: A validation study]. *Zeitschrift für Arbeits- und Organisationspsychologie*, *56*, 14–23. doi:10.1026/0932-4089/a000070

Briggs Myers, I., & Myers, P. B. (1993). *Gifts differing: Understanding personality type*. Palo Alto, CA: Consulting Psychologists Press.

Brown, R. (1999). The use of personality tests: A survey of usage and practice in the UK. *Selection and Development Review*, *15*, 3–8.

Bruchon-Schweitzer, M., & Ferrieux, D. (1991). Une enquête sur le recrutement en France [An inquiry on personnel recruitment in France]. *European Review of Applied Psychology*, *41*, 9–17.

Butcher, J. N., Perry, J. N., & Dean, B. L. (2009). How to use computer-based reports. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 693–706). Oxford, UK: Oxford University Press.

Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality*, *40*, 359–376. doi:10.1016/j.jrp.2005.01.006

Di Milia, L. (2004). Australian management selection practices: Closing the gap between research findings and practice. *Asia Pacific Journal of Human Resources*, *42*, 214–228. doi:10.1177/1038411104045364

Donston-Miller, D. (n.d.). Top Personality Tests Used in Hiring. *The Ladders*. Retrieved August 26, 2013, from http://www.theladders.com/career-advice/top-personality-tests-hiring

Driver, R. W., Buckley, M. R., & Frink, D. D. (1996). Should we write off graphology? *International Journal of Selection and Assessment*, *4*, 78–86. doi:10.1111/j.1468-2389.1996.tb00062.x

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, *91*, 40–57. doi:10.1037/0021-9010.91.1.40

Eleftheriou, A., & Robertson, I. (1999). A survey of management selection practices in Greece. *International Journal of Selection and Assessment*, *7*, 203–208. doi:10.1111/1468-2389.00120

Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., … Urbánek, T. (2012). Testing practices in the 21st century. *European Psychologist*, *17*, 300–319. doi:10.1027/1016-9040/a000102

Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology*, *53*, 748–759. doi:10.1037/0022-006X.53.6.748

Free Personality Test. (n.d.). Personality tests: The main personality tests (Myers Briggs, Big 5, 16 PF ...). *Personality test free*. Retrieved August 26, 2013, from http://www.personalitytestfree.net/personality-tests.php

Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment*, *16*, 300–305. doi:10.1111/j.1468-2389.2008.00436.x

Gangestad, S., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review*, *92*, 317–349. doi:10.1037/0033-295X.92.3.317

Ghiselli, E. E., & Barthol, R. P. (1953). The validity of personality inventories in the selection of employees. *Journal of Applied Psychology*, *37*, 18–20. doi:10.1037/h0059438

Guastello, S. J., Guastello, D. D., & Craft, L. L. (1989). Assessment of the Barnum effect in computer-based test interpretations. *Journal of Psychology*, *123*, 477.

Guastello, S. J., & Rieke, M. L. (1990). The Barnum effect and validity of computer-based test interpretations: The Human Resource Development Report. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*, 186–190. doi:10.1037/1040-3590.2.2.186

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, *18*, 135–164. doi:10.1111/j.1744-6570.1965.tb00273.x

Harris, M. M., Dworkin, J. B., & Park, J. (1990). Preemployment screening procedures: How human resource managers perceive them. *Journal of Business and Psychology*, *4*, 279–292. doi:10.1007/BF01125240

Herrmann International. (n.d.). A history of innovation in business. Retrieved December 30, 2013, from http://www.herrmannsolutions.com/our-foundation/

Herrmann, N. (1989). *The creative brain*. Lake Lure, NC: Brain Books.

Hodgkinson, G. P., Daley, N., & Payne, R. L. (1995). Knowledge of, and attitudes towards, the demographic time bomb: A survey of its impact on graduate recruitment in the UK. *International Journal of Manpower*, *16*, 59–76. doi:http://www.emeraldinsight.com/ijm.htm

Hodgkinson, G. P., & Payne, R. L. (1998). Graduate selection in three European countries. *Journal of Occupational and Organizational Psychology, 71*, 359–365. doi:10.1111/j.2044-8325.1998.tb00682.x

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92,* 1270–1285. doi:10.1037/0021-9010.92.5.1270

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, *88*, 100–112. doi:10.1037/0021-9010.88.1.100

Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, *18*, 331–341. doi:10.1207/s15327043hup1804_1

Hough, L. M., & Oswald, F. L. (2005). They're right, well ... mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance*, *18*, 373–387. doi:10.1207/s15327043hup1804_4

Hülsheger, U. R., & Maier, G. W. (2008). Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf: Eine Bestandsaufnahme internationaler und nationaler Forschung. [Personality traits, intelligence, and success at work: A review of international and national research]. *Psychologische Rundschau*, *59*, 108–122. doi:10.1026/0033-3042.59.2.108

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, *85*, 869–879. doi:10.1037/0021-9010.85.6.869

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*, 765–780. doi:10.1037/0021-9010.87.4.765

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, *87*, 530–541. doi:10.1037/0021-9010.87.3.530

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the

life span. *Personnel Psychology*, *52*, 621–652. doi:10.1111/j.1744-6570.1999.tb00174.x

Jung, C. G. (1960). *Psychologische Typen [Psychological types]* (9th ed.). Zürich: Rascher.

Keil, F. C. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Sciences*, *7*, 368. doi:10.1016/S1364-6613(03)00158-X

Keil, F. C. (2006). Explanation and understanding. *Annual Review ofPsychology*, *57*, 227–254. doi:10.1146/annurev.psych.57.102904.190100

Kellett, D., McCahon, S., & James, J. (1991). Preliminary evaluation of five computer-generated narrative reports derived from four different personality questionnaires. *European Work and Organizational Psychologist*, *1*, 196–210.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*, 624–662. doi:10.1177/1094428112452760

König, C. J., Klehe, U., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, *18*, 17–27. doi:10.1111/j.1468-2389.2010.00485.x

Lievens, F., & Harris, M. M. (2003). Research on internet recruiting and testing: Curren status and future directions. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 18). New York, NY: Wiley.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*, 464–470. doi:16/j.tics.2006.08.004

MacLean, P. D. (1985). Evolutionary psychiatry and the triune brain. *Psychological Medicine, 15*, 219–221. doi:10.1017/S0033291700023485

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.

Marcus, B., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance, 20,* 275–285. doi:10.1080/08959280701333362

Marston, W. M. (1979). *Emotions of normal people*. Minneapolis: Persona Press.

McCrae, R. R., & Costa, P. T. J. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509–516. doi:10.1037/0003-066X.52.5.509

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599–616. doi:10.1037/0021-9010.79.4.599

Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods, 10,* 322–345. doi:10.1177/1094428106289393

Mintzberg, H. (1976). Planning on the left side and managing on the right. *Harvard Business Review, 54*, 49–58.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K. R., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683–729. doi:10.1111/j.1744-6570.2007.00089.x

Muñiz, J., & Fernández-Hermida, J. R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests [Spanish psychologists opinions on test use]. *Papeles Del Psicólogo*, *31*, 108–121.

Muñiz, J., Prieto, G., Almeida, L., & Bartram, D. (1999). Test use in Spain, Portugal and Latin American Countries. *European Journal of Psychological Assessment*, *15*, 151–157. doi:10.1027//1015-5759.15.2.151

Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, *18*, 343–357. doi:10.1207/s15327043hup1804_2

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, *17*, 609–626. doi:10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance*, *18*, 389–404. doi:10.1207/s15327043hup1804_5

Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of fortune 1000 firms. *North American Journal of Psychology*, *8*, 489–496.

Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56*, 733–752. doi:10.1111/j.1744-6570.2003.tb00757.x

Psychometric Success. (2013). Understanding the Personality Test Industry. *Psychometric Success*. Retrieved August 26, 2013, from http://www.psychometric-success.com/personality-tests/personality-tests-understanding-industry.htm

Robie, C., Tuzinski, K. A., & Bly, P. R. (2006). A survey of assessor beliefs and practices related to faking. *Journal of Managerial Psychology*, *21*, 669–681. doi:10.1108/02683940610690204

Robins, R. W., John, O. P., & Caspi, A. (1998). The typological approach to studying personality. In R. B. Cairns, L. R. Bergman, & J. Kagan (Eds.), *Methods and models for studying the individual* (pp. 135–160). Thousand Oaks, CA: Sage.

Robins, R. W., & Tracy, J. L. (2003). Setting an agenda for a person-centered approach to personality development. *Monographs of the Society for Research in Child Development*, *68*, 110–122. doi:10.1111/j.1540-5834.2003.06801012.x

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, *16*, 155–180. doi:10.1016/j.hrmr.2006.03.004

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562. doi:10.1207/s15516709cog2605_1

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, *52*, 359–391. doi:10.1111/j.1744-6570.1999.tb00165.x

Ryan, A. M., & Sackett, P. R. (1987). A survey of individual assessment pratices by I/O psychologists. *Personnel Psychology*, *40*, 455–488. doi:10.1111/j.1744-6570.1987.tb00610.x

Ryan, A. M., & Sackett, P. R. (1992). Relationships between graduate training, professional affiliation, and individual psychological assessment practices for personnel decisions. *Personnel Psychology*, *45*, 363–387. doi:10.1111/j.1744-6570.1992.tb00854.x

Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, *50*, 987–1008. doi:10.5465/AMJ.2007.27151939

Rynes, S. L., Orlitzky, M. O., & Bretz, R. D. (1997). Experienced hiring versus college recruiting: Practices and emerging trends. *Personnel Psychology*, *50*, 309–339. doi:10.1111/j.1744-6570.1997.tb00910.x

Salgado, J. F. (1997). The Five Factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, *82*, 30–43. doi:10.1037/0021-9010.82.1.30

Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment*, *11*, 194–205. doi:10.1111/1468-2389.00243

Scholarios, D., & Lockyer, C. (1999). Recruiting and selecting professionals: Context, qualities and methods. *International Journal of Selection and Assessment*, *7*, 142–156. doi:10.1111/1468-2389.00114

Schuler, H., Frier, D., & Kauffmann, M. (1993). *Personalauswahl im europäischen Vergleich. [Personnel selection in a European comparison.]*. Göttingen, Germany: Verlag für Angewandte Psychologie.

Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen: Ein Vergleich über 20 Jahre. [Use of personnel selection instruments in German organizations during the last 20 years.]. *Zeitschrift für Personalpsychologie*, *6*, 60–70. doi:10.1026/1617-6391.6.2.60

Shackleton, V., & Newell, S. (1994). European management selection methods: A comparison of five countries. *International Journal of Selection and Assessment*, *2*, 91–102. doi:10.1111/j.1468-2389.1994.tb00155.x

Smith, E. R., & Queller, S. (2008). Mental Representations. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 111–133). New York, NY: Wiley.

Sneath, F., Thakur, M., & Madjuck, B. (1976). *Testing people at work* (Information Report No. 24). London: Institute of Personnel Management.

Snyder, D. K. (2000). Computer-assisted judgment: Defining strengths and liabilities. *Psychological Assessment*, *12*, 52–60. doi:10.1037/1040-3590.12.1.52

Spieß, E., Eckstaller, C., & Woschée, R. (2004). Gutachten INSIGHTS MDI ® [expert report INISGHTS MDI ®].

Steck, P. (1997). Psychologische Testverfahren in der Praxis: Ergebnisse einer Umfrage unter Testanwendern. [Psychological tests in practice: A survey among test users.]. *Diagnostica*, *43*, 267–284.

Stewart, G. L., Darnold, T. C., Zimmerman, R. D., Parks, L., & Dustin, S. L. (2010). Exploring how response distortion of personality measures affects individuals. *Personality and Individual Differences*, *49*, 622–628. doi:10.1016/j.paid.2010.05.035

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality–job performance relationship. *Journal of Organizational Behavior*, *24*, 335–356. doi:10.1002/job.191

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124 –1131. doi:10.1126/science.185.4157.1124

Williams, R. S. (1992). Management selection in local government: A survey of practice in England and Wales. *Human Resource Management Journal*, *3*, 63–73. doi:10.1111/j.1748-8583.1992.tb00311.x

Wilson, R. A., & Keil, F. C. (1998). The shadows and shallows of explanation. *Minds and Machines: Journal for Artificial Intelligence, Philosophy, and Cognitive Science*, *8*, 137–159.

York, K. L., & John, O. P. (1992). The four faces of Eve: A typological analysis of women's personality at midlife. *Journal of Personality and Social Psychology*, *63*, 494–508. doi:10.1037/0022-3514.63.3.494

Zibarras, L. D., & Woods, S. A. (2010). A survey of UK selection practices across different organization sizes and industry sectors. *Journal of Occupational and Organizational Psychology*, *83*, 499–511. doi:10.1348/096317909X425203

TABLES AND FIGURES

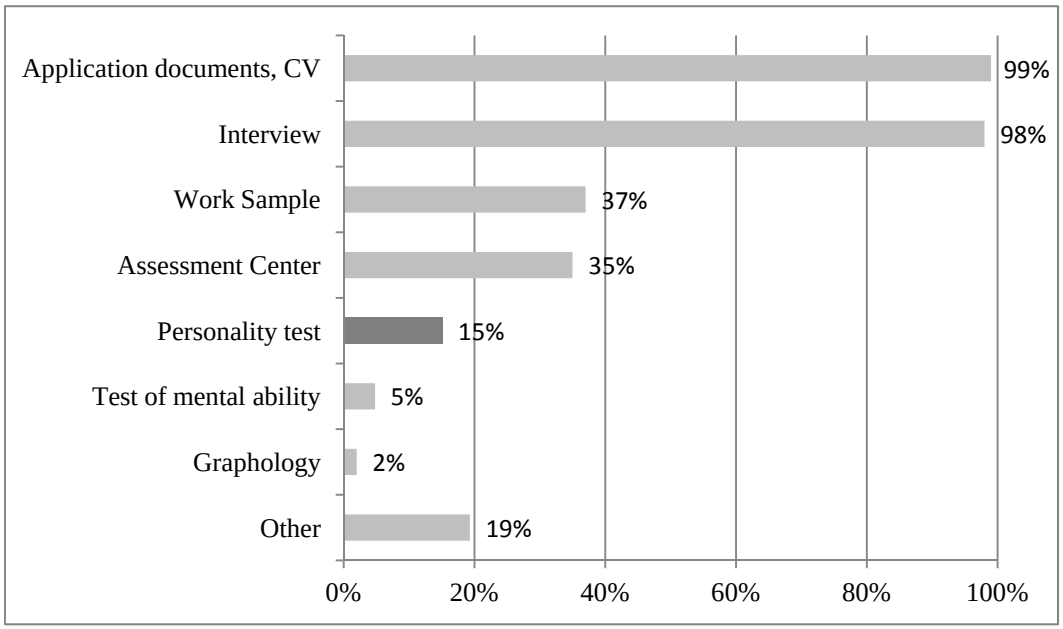

Figure 1. Frequency of selection methods used in percent (*N* = 166 German companies).

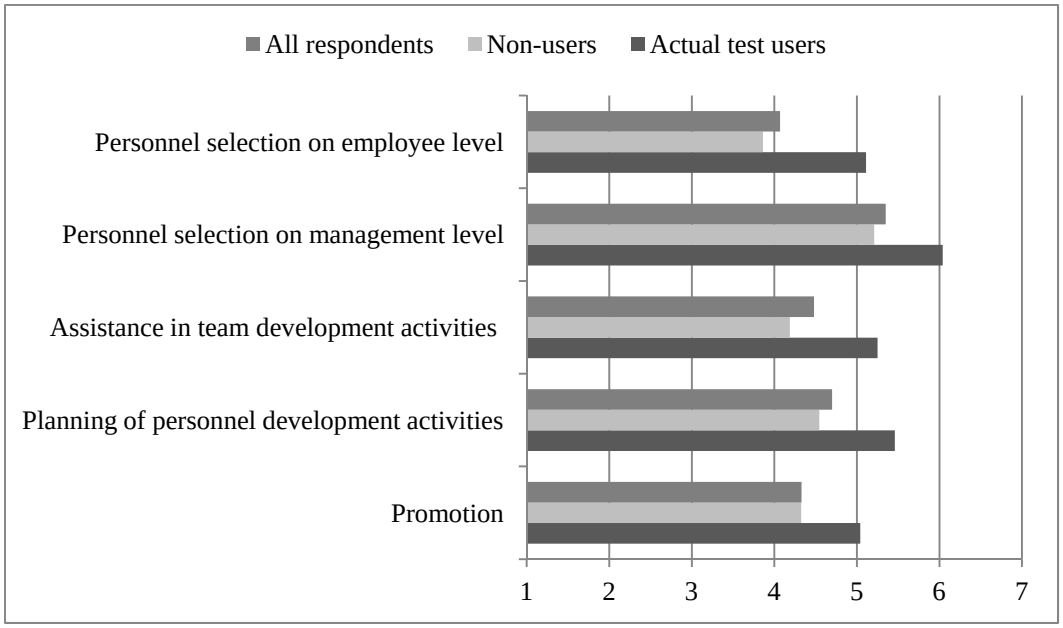

Figure 2. Evaluation of the benefit of personality tests for different purposes. Agreement regarding usefulness was given on a seven-point scale (1 = no agreement to 7 = full agreement). All differences between users and non-users were significant ($p < 0.01$, all t's < -2.6).

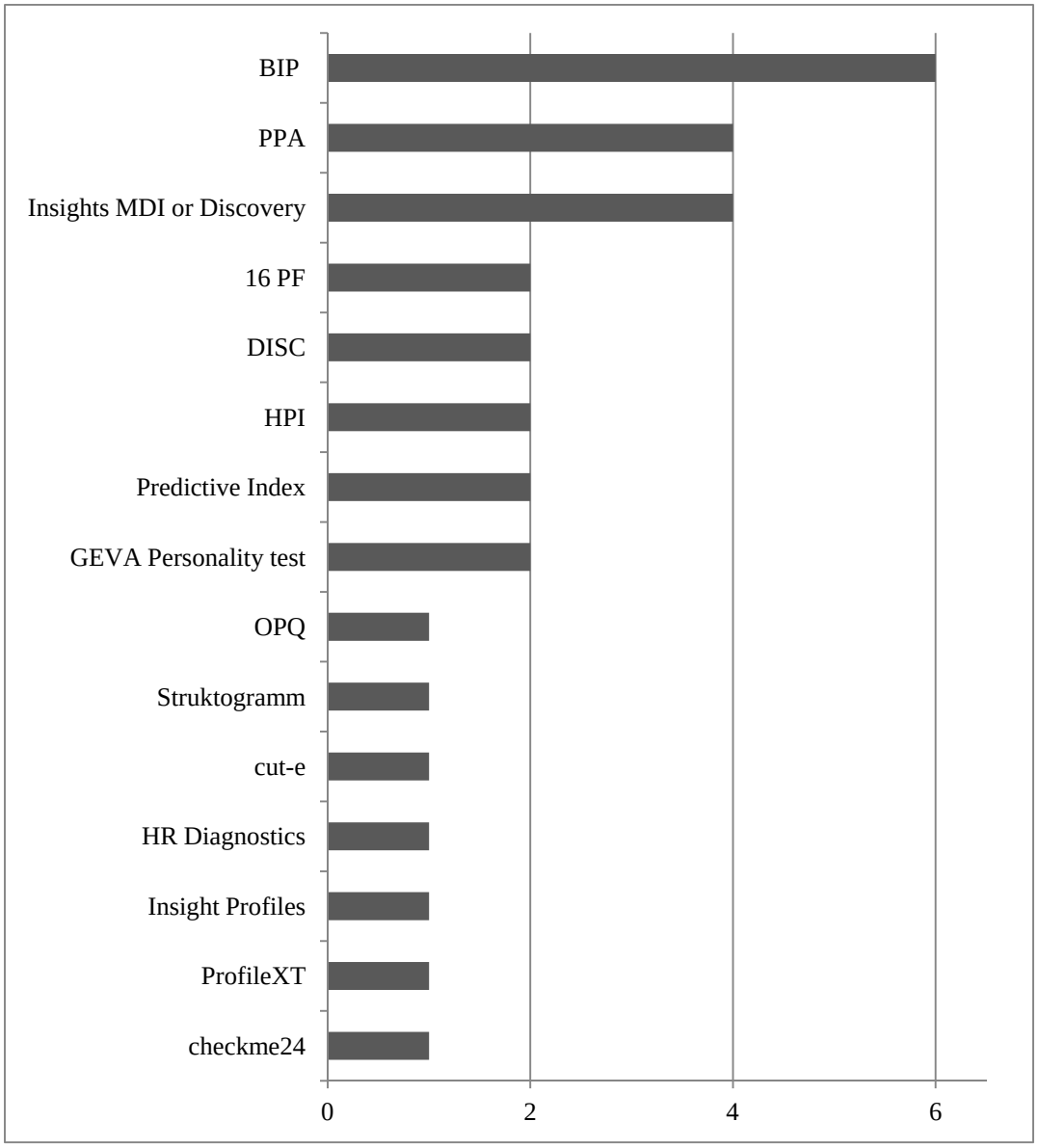

Figure 3. Personality tests used in Germany (in frequencies). (BIP = Bochum Inventory for profession-related personality description; PPA = Thomas Assessment / Personal Profile Analysis; HPI = Hogan Personality Inventory; GEVA = the GEVA institute is a German consulting company specialized in behavioral analysis and evaluation tools; OPQ = Occupational Personality Questionnaire).

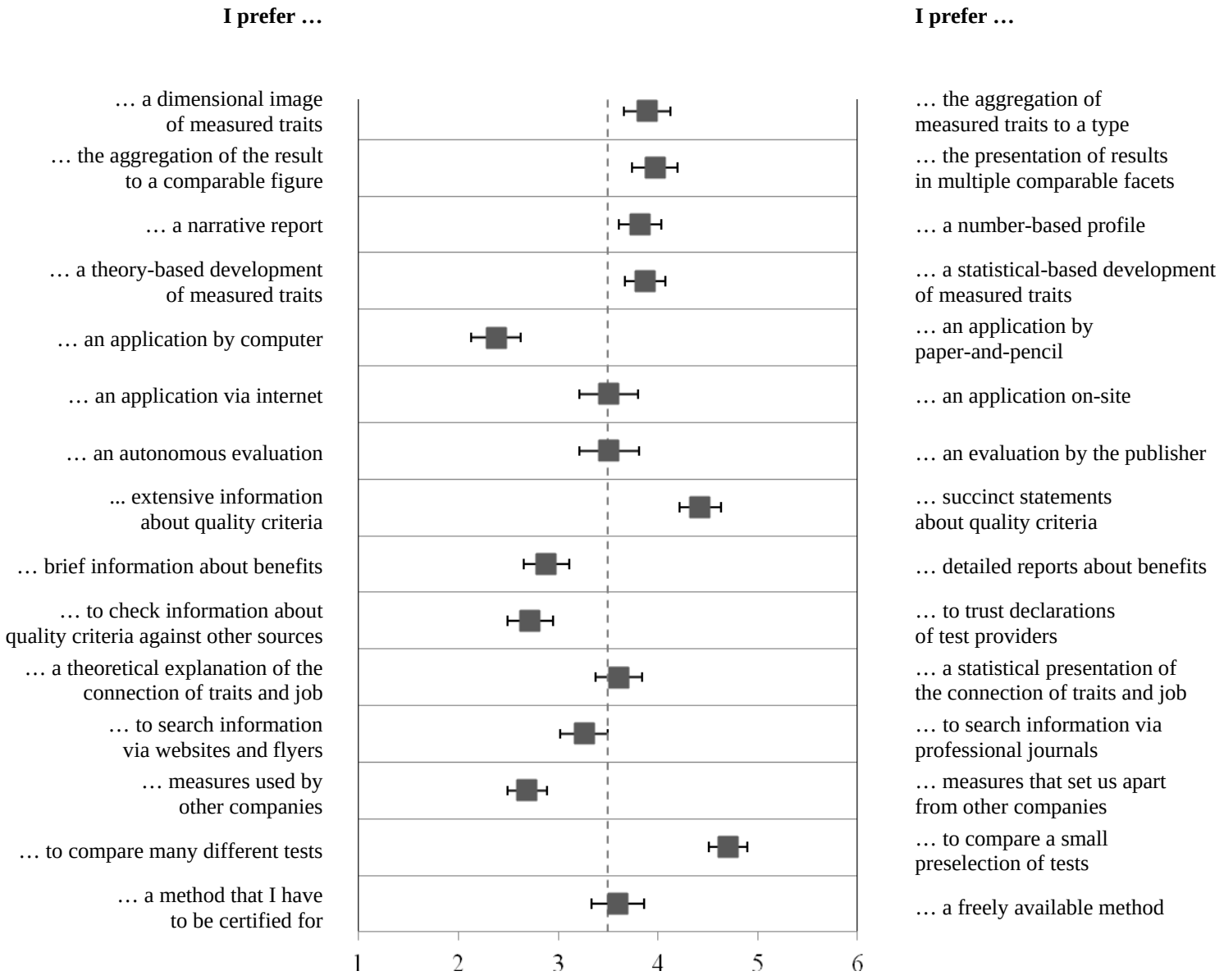

Figure 4. Preferences of different criteria that distinguish personality tests. Error bars indicate 95% confidence interval. Preferences were rated on a six-point semantic differential scale.