SOCIAL DESIRABILITY IN SELF-REPORT PERSONALITY QUESTIONNAIRES FOR PERSONNEL SELECTION: FRIEND OR FOE?

Dirk H.M. Pelt

Social Desirability in Self-report Personality Questionnaires for Personnel Selection: Friend or Foe?

Dirk H.M. Pelt

The studies in this thesis were financially supported by Ixly BV, Utrecht, The Netherlands.

Cover design: Daniëlle Balk | www.persoonlijkproefschrift.nl
Printing: Ridderprint | www.ridderprint.nl

ISBN: 978 94 6375 463 7

Social Desirability in Self-report Personality Questionnaires for Personnel Selection:

Friend or Foe?

Sociale wenselijkheid in zelf-rapportage persoonlijkheidsvragenlijsten voor personeelsselectie:

Vloek of zegen?

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de
rector magnificus

Prof. Dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 5 september 2019 om 15.30 uur

door
Henricus Maria Pelt
geboren te Utrecht

Erasmus University Rotterdam

Erasmus

Promotiecommissie

Promotoren:	Prof. Dr. M.Ph. Born Prof. Dr. D. van der Linden
Overige leden:	Prof. Dr. H.T. van der Molen Prof. Dr. C.J. König Prof. Dr. R.R. Meijer

Contents

General Introduction

1

When choosing a new career path or when running for a promotion, an assessment of one's qualities and competencies that includes a self-report personality inventory is a rather common procedure. Similarly, when applying for a job, chances are that applicants will be subjected to a personality assessment in one form or another (Kantrowitz, Tuzinski, & Raines, 2018). With the arrival of extensive and fast internet connections, it now has become even easier for companies to include computerized personality testing as part of the selection and assessment process. Therefore, personality testing may provide organizations with a cost-effective and quick way of selecting candidates who are most suitable for the available jobs or, alternatively, selecting out the "bad apples". This trend is reflected in data on personality test use; results from a large global yearly survey among large numbers of human resource professionals (between 500 and 3,000) show that over the past ten years, the use of personality questionnaire as a pre-hire tool by companies has fluctuated between 60% and 86% (Fallaw & Kantrowitz, 2013; Kantrowitz et al., 2018).

Apart from advantages in terms of costs and efficiency, organizations increasingly use self-report personality questionnaires as part of their selection procedures because they allow hiring better employees. Research has shown that personality test scores predict organizational citizenship behavior and counterproductive work behavior, and to a lesser degree job performance, even on top of cognitive abilities (Barrick & Mount, 1991; Cook, 2016; Schmidt & Hunter, 1998). Yet, the popularity of personality questionnaires among practitioners has also raised concerns about potential response distortion by applicants on personality inventories. Because it is very difficult to define whether an applicant is telling the truth, methods for estimating base rates of response distortion vary widely. Therefore, estimates of the prevalence of applicant response distortion also show large variability (e.g., Donovan, Dwight, & Hurtz, 2003; Dunnette, McCartney, Carlson, & Kirchner, 1962; Griffith, Chmielowski, & Yoshita, 2007; Griffith & Converse, 2011). Yet, keeping these measurement difficulties in mind, a cautious estimate is that around 30% ($SD = 10$) of applicants distort their responses to increase their chances of being hired (Griffith & Converse, 2011). This tendency by applicants to distort responses in order to maximize the chances of obtaining a desired job, rather than to answer honestly, is often referred to as socially desirable responding (Mesmer-Magnus & Viswesvaran, 2006; Ones, Viswesvaran, & Reiss, 1996).

The term socially desirable responding, or social desirability in short, is known among personality researchers under a large number of terms and concepts, such as faking, malingering, self-enhancement or impression management, and each concept has a slightly different definition (Ziegler, MacCann, & Roberts, 2011a). Yet, all the concepts revolve around responding in a way that provides a favorable image of oneself, rather than reflecting one's 'true' personality, or how one typically behaves (Edwards, 1957; Jackson & Messick, 1958; Paulhus, 2002). As such, social desirability is proposed to be an artefact of how personality is measured, namely through self-reports. In this so-called artefactual definition, social desirability is considered a *bias*, resulting from a person by situation interaction (e.g., Ziegler & Buehner, 2009), creating *spurious measurement error* (Schmidt, Le, & Ilies, 2003); in this view, individual differences in social desirability are seen as a function of the characteristics of the situation (in our case a selection procedure) in combination

with differences in individuals' reactions to these characteristics, rather than a stable trait *across* situations. Consequently, the valid measurement of personality traits may be obscured, with important consequences for personnel selection. For example, a job candidate who would engage in socially desirable responding during the selection process will appear to have a more favorable personality profile (in terms of work-related outcomes) compared to a candidate who will largely refrain from such a response style. As a result, the former candidate will have a higher probability of getting selected for the job even if his or her true personality profile is not more favorable than the latter candidate. An even worse situation would be when, due to socially desirable responding, someone with a less favorable true personality profile would get selected above someone with a more favorable profile.

At the same time, earlier research has questioned this notion of social desirability as a form of bias, arguing that it may largely reflect a substantive construct (e.g., Connelly & Chang, 2016; McCrae & Costa, 1983; Ones et al., 1996; Uziel, 2010a; Zettler, Hilbig, Moshagen, & De Vries, 2015). Different definitions of social desirability as a substantive construct exist, such as "interpersonally oriented self-control" (Uziel, 2010a) or as a culturally influenced "communication filters", that is, how an individual expresses oneself to others (He & Van de Vijver, 2013). What the different definitions have in common is that they all relate to how one deals with interactions with others, or phrased differently, the ability to successfully navigate the social world (Dunkel, 2013). In any case, regardless of the exact meaning attached to social desirability, in this substantive interpretation socially desirable responding is assumed to be more trait-like and hence less dependent on characteristics of specific situations. In this view, social desirability is seen as a substantive individual difference variable which is relatively stable over time and contexts and with implications for various work-related outcomes. Note that this substantive interpretation does not imply that the expression of social desirability is completely unaffected by the type of situation it is evoked by (Mischel & Shoda, 1995). For example, in selection situations we can expect people to put their best foot forward, yet some people may on the whole in reality possess more socially desirable traits than others, while others possess more undesirable traits (Oltmanns, Smith, Oltmanns, & Widiger, 2018). Alternatively, some people may have higher levels of social knowledge and skills enabling them to present themselves more favorably so that they are regarded as more socially desirable by others. According to this view, a person's social desirability score largely reflects a person's personality, rather than an indication of how much someone has 'faked'.

Recently, it has been proposed that a large part of social desirability as a trait-like construct can be captured by the so-called General Factor of Personality (GFP; Figueredo, Vásquez, Brumbach, & Schneider, 2004; Musek, 2007). The GFP represents the shared variance of personality domains such as the Big Five (Van der Linden, Dunkel, & Petrides, 2016). It typically captures the socially desirable ends of personality dimensions. Although the substance versus artefact debate also revolves around the GFP as will be discussed later, in its substantive interpretation individuals scoring high on the GFP, on average, would be characterized as being relatively open-minded, diligent, sociable, friendly, and emotionally stable.

The GFP has been recovered in a large number of personality inventories across primary studies and in multiple meta-analyses (e.g., Davies, Connelly, Ones, & Birkeland, 2015; Van der Linden, Te Nijenhuis, & Bakker, 2010a). It has been found to be associated with relevant outcomes such as self-esteem, mood, health, social relationships, leadership, and (negatively) with delinquent behavior (Dunkel, Van der Linden, Brown, & Mathes, 2016; Musek, 2007; Van der Linden, Dunkel, Beaver, & Louwen, 2015; Van der Linden, Scholte, Van Leeuwen, Te Nijenhuis, & Engels, 2010b). Recently, it has been suggested that the GFP represents social effectiveness (Van der Linden et al., 2016), that is a factor related to knowledge about what is considered socially desirable behavior and a tendency to act in that way, thereby optimizing the attainment of personal and social goals (e.g., getting a job or promotion).

The aim of this dissertation is to add to the growing body of evidence, across samples and methods of analyses (e.g., Dunkel et al., 2016; Fisher & Robie, 2019; Van der Linden et al., 2016) for the substantive interpretation of the general factor in self-report personality questionnaires. Four empirical studies are presented in which the substantive interpretation of this general social desirability factor as representing social effectiveness is put to the test. In broader terms, the aim of this dissertation is to provide novel and useful insights for the discussion about the extent to which social desirability can be considered substantive and to what extent an artefact. This dissertation approaches the topic of social desirability mainly from the domain of organizational psychology and personnel selection. In this field, given the contrasting interpretations of social desirability outlined previously, important questions remain unanswered about its practical utility and how to deal with it in selection situations. Adherents of the substantive interpretation of social desirability would advocate that it can be used as a selective trait, given that it would probably predict a wide range of behaviors, including job performance. Those favoring the artefactual interpretation would argue that social desirability introduces systematic error to the measurement of personality in selection procedures, with negative consequences such as hiring the 'wrong' while rejecting the 'right' candidates and subsequently leading to a decrease of the predictive power of personality measures.

In this introductory chapter, a short review of the literature as well as different theoretical perspectives and interpretations are provided of social desirability in general and the GFP more specifically. Subsequently, the main research question is presented followed by more specific research questions. This chapter ends with a summary of the specific research aims of the four empirical chapters of this dissertation.

Social Desirability: Bias, Substance, or Both?

Historically, social desirability has been viewed as a distorting influence on the valid measurement of personality traits. The issue of social desirability in self-report personality questionnaires became prominent in the psychological research literature in the 1950s, mostly through the development of the Minnesota Multiphasic Personality Inventory (MMPI) and in subsequent work by Edwards

(1953, 1957) and colleagues (Edwards & Walsh, 1963; Edwards, Diers, & Walker, 1962). Edwards posed that both individuals and personality test items may differ in their levels of social desirability. Consequently, Edwards advocated that socially desirable response tendencies obscured accurate measurement of traits through self-reports because people not only respond to the content of a given personality item, but also to the level of social desirability of the item (Bäckström, Björklund, & Larsson, 2009; Edwards et al., 1962; Hendriks, Hofstee, & De Raad, 1999; Peabody, 1967). Therefore, a trait score formed by the sum of a number of personality items will be a contamination of both trait content and social desirability (e.g., Biderman, McAbee, Job Chen, & Hendy, 2018). For decades to follow, the idea of social desirability as a nuisance factor which needed to be reduced or eliminated was the consensus among personality researchers and practitioners alike.

However, different streams of research, some dating back from centuries ago, have suggested that social desirability might possibly be more akin to a trait, that is, a stable characteristic of a person, related to knowing what to do and how to behave in social situations. Three such streams can be distinguished. The first stream relates to the general social desirability component found in personality questionnaires (i.e., the GFP). The second stream of research focuses on the interpretation of scores on social desirability scales. The third stream specifically focuses on knowing what is required and acting accordingly in selection situations. The three streams and their communalities are discussed in the following sections.

Social Desirability as a General Factor in Personality Measures

The first stream of research focuses on the general social desirability factor found in scores on personality questionnaires. Although the concept of a single personality dimensions ranging from socially undesirable to socially desirable is much older, as I will discuss below, the aforementioned studies by Edwards brought the concept of a general social desirability factor to the field of personality assessment. Specifically, among a sample of 151 U.S. students Edwards et al. (1962) showed that the item loadings on the first, most general factor of the MMPI correlated strongly (r = .90) with external judgments of the level of social desirability of these items. Consequently, it was concluded that the general factor in the MMPI, and presumably in other personality questionnaires, represented social desirability. However, as noted, Edwards explicitly regarded this social desirability factor to purely reflect misrepresentation of the self, thus reflecting a *bias* which contaminates the clear measurement of the 'true' personalities of individuals. Many other studies, including recent ones, have replicated the findings by Edwards and colleagues by showing a large correlation between first factor loadings of items and items' social desirability ratings (Anglim, Morse, De Vries, MacCann, & Marty, 2017; Bäckström & Björklund, 2013; Biderman et al., 2018).

Interestingly, the idea of a *true* social desirability factor - that is not a *bias* but a *trait* - had been around a long time before Edwards' influential work. The notion of a single overarching social desirability factor can be traced back all the way to Francis Galton (1887), a relative of Charles Darwin. Galton argued that people can be characterized by means of a single dimension ranging

from a "bad temper" to "good temper". Good tempers could be characterized by qualities such as "amiable", "calm", "gentle", "good", "self-controlled", and "sunny", while characteristics such as "aggressive", "contentious", "grumpy", "harsh", "quarrelsome", and "vicious" were ascribed to bad tempers. Ever since, this topic of a single personality continuum ranging from negative to positive has disappeared and resurfaced in the personality literature from time to time (e.g., Fiske, 1949; Hofstee, 2003; Peabody & Goldberg, 1989; Saucier, 1994; Webb, 1915).

However, in the course of the 20th century, a relative consensus arrived on the idea that the structure of personality was best described as multiple, presumably independent personality traits (John & Srivastava, 1999; McCrae & Costa, 1999). How many of such basic traits would best provide a comprehensive description of personality, to date, remains a topic of debate. For example, Eysenck (1970) advocated three broad and basic personality factors (Extraversion, Neuroticism and Psychoticism), while Goldberg (1990) and Costa and McCrae (1992) arrived at five factors. This Five Factor Model, also referred to as the Big Five, has become the standard personality taxonomy in the personnel selection literature. In fact, the widespread acceptance of the five factors – Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism forming the acronym OCEAN – was the catalyst for the use of personality measures in personnel selection (Barrick, Mount, & Judge, 2001). More recently, a six-factor model has been proposed (the HEXACO model; Ashton & Lee, 2001), which has met a considerable amount of attention in the selection and assessment literature (e.g., Lee, Ashton, & De Vries, 2005; Anglim, Lievens, Everton, Grant, & Marty, 2018; Anglim et al., 2017).

With the relative consensus that five or six factors constitute the most general level of the personality domain at which people can be differentiated, the possible existence of higher-level factors or maybe even one general, broad personality factor was pushed to the background. Yet, despite the assumption that the factors in the personality domain are relatively independent (e.g., Goldberg, 1990; Saucier, 2002), and despite efforts to develop instruments that measure them as independent factors, it has been consistently found that the Big Five factors show moderate intercorrelations (Block, 1995; Digman, 1997; DeYoung, Peterson, & Higgins, 2002). These intercorrelations imply that the five factors share a relevant proportion of variance; this phenomenon in turn implies that a factor might exist at a higher level in the personality hierarchy, explaining these intercorrelations.

Alternative explanations, other than a single underlying factor, for positive manifolds (i.e., all positive correlations between a set of traits) exist, for example the concept of *mutualism* or *network theory* (e.g., Cramer et al., 2012; Van Bork, Epskamp, Rhemtulla, Borsboom, & Van der Maas, 2017). Yet, these alternative explanations lie beyond the scope of the current dissertation. In the first place, because from an applied perspective, network theory requires relatively complicated models and statistical analyses (Cramer et al., 2012), which would be hard to explain to applicants in the selection context. Second, although network theory has been applied to personality models such as the Big Five, it is more accepted in the field of clinical psychology (Cramer et al., 2012). Furthermore, both network theory and the GFP provide alternative explanations for the same

phenomenon, namely the covariation among personality traits: in this dissertation, the feasibility of a general factor as one of the possible explanations for the positive manifold is investigated.

Currently, multiple primary studies and meta-analyses have made it clear that a general factor can be found in the Five Factor model as well as other models of personality, typically explaining about 20 to 60% of the variance among the lower order domain traits (Van der Linden et al., 2016). In fact, the existence of shared variance among personality traits actually seems to be highly consistent and has been replicated in many of the available personality datasets across the globe (Davies et al., 2015; Musek, 2017; Van der Linden, Bakker, & Te Nijenhuis, 2010a).

Based on these findings, most scholars would now probably agree that a general factor in self-report personality measures does *exist*. Despite these findings, however, diverging scientific views on the GFP exist in terms of its *interpretation*. Two opposing interpretations, substantive versus artefactual, are represented in the personality literature, while others take a more nuanced standpoint, acknowledging that any psychological construct measured through self-reports will capture both 'true' variance and variance related to the method of measurement (Davies et al., 2015; Dunkel et al., 2016).

Before turning to the discussion on the two opposing interpretations – artefactual vs. substantive – of the general factor in personality measures, it is interesting to note that similar general factor debates are found in different bodies of scientific literature. For example, one discussion revolves around whether "dark personality" is best represented by the Dark Triad (narcissism, psychopathy, and Machiavellianism; Paulhus & Williams, 2002) or by a single, unitary Dark Core (e.g., Bertl, Pietschnig, Tran, Stieger, & Voracek, 2017; Jonason, Li, Webster, & Schmitt, 2009). Other examples include the literature on vocational interests (Darcy & Tracey, 2003; Tracey, 2012; Prediger, 1998), psychopathology (where it is labeled the *p*-factor; e.g., Caspi et al., 2014), and the general factor of personality disorder (PD-factor; e.g., Jahng et al., 2011; Wright, Hopwood, Skodol, & Morey, 2016). The current dissertation is restricted to the general factor found in self-report measures of "normal" or "bright" personality, since these are most commonly used and studied in the field of selection and assessment (Moscoso & Salgado, 2004; Spain, Harms, & LeBreton, 2014). Yet, studies have shown that the general factors from "normal" personality, psychopathology and personality disorder largely seem to overlap (correlations ranging between .70 and .90; Oltmanns et al., 2018; Rosenström et al., 2018).

A substantive GFP

One view on the GFP is that it is a substantive construct which reflects *true* socially desirable behavior. According to this view, people with high scores on the GFP show a mix of socially desirable traits, being, on average, more friendly, hard-working, diligent, emotionally stable, and open than those with lower scores. The leading substantive interpretation is that the GFP reflects social effectiveness (see Van der Linden et al., 2016, for a review). All personality dimensions include interpersonal or social aspects, although to varying degrees (Wiggins, 1979). It therefore seems reasonable to assume that the common core of the Big Five relates to how one deals with others.

According to the social effectiveness interpretation, individuals high on the GFP may have a set of knowledge, skills, and motivation to act in socially desirable ways, hereby increasing their chances of achieving social goals. In line with the account of the GFP as a social desirability factor, it has been found that the GFP is influenced by culture (Dunkel, 2013; He & Van de Vijver, 2013). This finding is not surprising, given that what constitutes socially desirable behavior is defined by the culture at hand (Bou Malham & Saucier, 2016; Van der Linden et al., 2016). Yet, overall, the GFP appears to be relatively stable across cultures (Aghababaei, 2013; Musek, 2007, 2017; Van der Linden et al., 2018).

Explanations for the existence of the GFP have been sought in evolutionary processes (Dunkel, Nedelec, & Van der Linden, 2018; Figueredo et al., 2004; Van der Linden et al., 2016). That is, over the course of human history, those with more socially desirable personalities appeared to be more resistant to socially and sexually selective pressures, leaving them with a selective advantage and thus with more reproductive success (Van der Linden, Figueredo, De Leeuw, Scholte, & Engels, 2012a), a phenomenon described as the "survival of the friendliest" (Hare, 2017). Corroborating the idea of the GFP as general social effectiveness is the large overlap found between the GFP and emotional intelligence (EI), a construct that has previously been linked to being socially effective (e.g., Ferris, Perrewé, & Douglas, 2002). In a recent meta-analysis (Van der Linden et al., 2017), the GFP showed a large amount of overlap with so-called trait EI ($r = .86$). Trait EI relates to understanding one's own emotions and motivations and those of others, and the tendency to use this knowledge to be socially effective (e.g., Mavroveli, Petrides, Rieffe, & Bakker, 2007; Sevdalis, Petrides, & Harvey, 2007). These results on the overlap between the GFP and EI thus strengthen the notion of the GFP as a social effectiveness factor.

If the GFP indeed reflects a social effectiveness factor, associated with showing a wide range of socially desirable behaviors, then this should have a large impact on an individual's life; presumably, it would lead to better personal relationships, getting along with colleagues, and increasing one's chances of reaching personal or social goals such as getting a promotion or acquiring a leadership position. A large number of studies has now been devoted to testing the relations between the GFP and such criteria of social effectiveness. For example, GFP scores have been associated with peer-rated popularity and likeability (Van der Linden et al., 2010b), less delinquent behavior (Van der Linden et al., 2015), leadership, and ability tests of social abilities and skills (Van der Linden, Oostrom, Born, Van der Molen, & Serlie, 2014a). Taken together, this body of evidence seems to suggest that the GFP is a substantive construct with important consequences for a variety of life outcomes, yet not all researchers agree with this viewpoint.

An artefactual GFP

In contrast to the substantive view of the GFP the artefactual interpretation has been expressed by researchers. Over the past ten years, multiple scholars have argued that the GFP is a consequence of common-method bias (e.g., Anusic, Schimmack, Pinkus, & Lockwood, 2009; Chang, Connelly, & Geeza, 2012), general response styles such as acquiescence (Arias, Jenaro, & Ponce, 2018), or some

combination of such biases. Others focus on the relatively small and variable amount of variance explained by the GFP in personality measures across studies (Revelle & Wilt, 2013), the lack of convergent validity of GFP's extracted from different personality inventories (e.g., Hopwood, Wright, & Donnellan, 2011), or methodological issues with the method of extraction of the GFP (Ashton, Lee, Goldberg, & de Vries, 2009). And most important for the current discussion, scholars have argued that the GFP reflects social desirability or evaluative bias, rather than being a substantive construct in itself.

The explanation for the GFP as a social desirability *bias* is based on the argument that the evaluative content (i.e., social desirability level) of personality items may trigger the motivation to self-enhance in some people more than in others (Bäckström et al., 2009). In this way, because people respond both to the personality content and evaluativeness of items (Biderman et al., 2018), scales tend to become more correlated, leading to the emergence of a general factor. Evidence for this line of reasoning comes from studies in which the GFP diminished when personality items were reframed to be less socially desirable (Bäckström et al., 2009; Bäckström, Björklund, & Larsson, 2012). In another study, it was shown that trait adjectives with opposite meaning but with similar valence (e.g., *sluggish* and *manic*) loaded strongly on the general factor, indicating that this factor reflects evaluativeness and captures response bias rather than content (Petterson, Turkheimer, Horn, & Menatti, 2012). These findings would suggest that GFP scores reflect self-serving bias and relate to endorsing socially desirable personality items rather than a stable trait that can be observed and validated by others.

Although the focus of the present dissertation is on social desirability in self-report measures of personality, results from studies on the GFP using self- and other-ratings of personality are relevant for the current discussion. That is, a set of studies using multi-trait-multimethod (MTMM) designs based on self- and other-ratings of personality led to the conclusion that the GFP is a within-rater phenomenon, possibly due to rater biases including social desirability or halo effects (e.g., Chang et al., 2012; Gnambs, 2013). In these studies, no GFP emerged from self-other correlations, while it did emerge when self-ratings or other-ratings were analyzed separately. Based on these findings, it was concluded that higher-order personality factors are due to common method variance (the method being the person, rating several personality traits). Different sources of method variance can be present in self-ratings (e.g., self-serving bias), other-ratings (e.g., as halo effects, 'liking' the target person, or the use of implicit trait theories), or both (e.g., acquiescence). However, regardless of the source of common method variance, the result will be that presumably independent traits show more overlap when based on a single reporter rather than on multiple reporters, leading to spurious higher-order factors (Anusic et al., 2009; Biesanz & West, 2004; Chang et al., 2012; Gnambs, 2013; McCrae et al., 2008).

In sum, different scientific opinions and interpretations of the shared variance between dimensions in personality inventories exist. However, as noted earlier, the two conflicting interpretations (substantive and artefactual) need not be mutually exclusive; as any psychological construct measured through self-reports, part of the shared variance among traits may indeed be

due to biases associated with the method of measurement and it will probably also reflect a relevant substantive component (Davies et al., 2015; Dunkel et al., 2016). Interestingly, the contrast between substance versus artefact in the discussion on the general factor in the personality literature is mirrored in a different stream of research, namely on the nature of scores on social desirability scales. In the following section, this stream of research and its relation with the discussion on the general factor in self-report personality questionnaires will be discussed.

Social Desirability Scales as Measures of Substantive Individual Differences

Findings from a second stream of research focusing on the interpretation of social desirability scales, seem to converge with the research described above in the sense that these findings show that social desirability might be more substantive than artefactual. Inspired by the work of Edwards and the customary idea that social desirability represented a distortion of reality, many researchers created scales in order to detect this form of misrepresentation on personality questionnaires. Over the years, the number of social desirability scales have skyrocketed (e.g., Crowne & Marlowe, 1960; Eysenck & Eysenck, 1964; Jackson, 1984; Paulhus, 1984; Stöber, 1999). Some of these scales have been constructed by grouping items with the most extreme social desirability ratings (the *minimalist* approach; Paulhus, 2002). Other approaches were more theoretical in nature (the *elaborate* approach; Paulhus, 2002), examples of which include the Marlowe-Crowne scale (1960) and the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984). The BIDR distinguishes between self-deceptive enhancement (SDE) and impression management (IM), and is the most commonly used scale in the selection and assessment literature. SDE refers to an unconscious tendency to describe oneself in a positive way, while the respondent genuinely believes these self-descriptions to be true. IM refers to a conscious, intentional act of presenting oneself in a more favorable light, while the respondent is aware that the description is not true to the self. Given that IM concerns the deliberate attempt to create a more positive impression, this form of social desirability has received the most attention in the selection literature (MacCann, Ziegler, & Roberts, 2011).

Although originally designed to detect response biases, it has become increasingly clear that social desirability scales largely capture substantive trait variance, rather than error variance. This interpretation is supported by the consistent finding that self-report scores on social desirability scales correlate substantially with reports on the same scales provided by others (in the range of $r = .21$-$.35$ for SDE and $r = .35$-$.45$ for IM; De Vries, Zettler, & Hilbig, 2014; Konstabel, Aavik, & Allik, 2006; Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007; Paulhus, 1991). Although somewhat lower, these values are roughly in line with the self-other correlations for the Big Five dimensions as reported in a meta-analysis by Connelly and Ones (2010). Furthermore, self-report social desirability ratings show small to moderate correlations with other-ratings of personality (see also section 2.1), indicating that at least part of the variance captured by social desirability scales must be consensually valid (De Vries et al., 2014; Holden & Passey, 2010; Konstabel

et al., 2006; Kurtz, Tarquini, & Iobst, 2008; Lönnqvist et al., 2007; Roth & Altmann, 2019). A recent line of research (De Vries et al., 2014; Zettler et al, 2015) has indicated that scores on the impression management scale of the BIDR are positively related to (other-rated) Honesty-Humility scores of the HEXACO model (r = .56 for self-rated Honesty-Humility with self-rated IM and r = .32 for self-rated Honesty-Humility with other-rated IM). Thus, those scoring high on the IM scale are actually rated as being somewhat *more* integer and honest than those who scores low on the IM scale; this finding is hard to reconcile with an account of the IM scale measuring the deceitful claim of having socially desirable traits one does not possess in reality. Rather, the scale appears to capture aspects of personality that can be corroborated by reports of others.

This substantive interpretation of social desirability measures is further in line with results from studies showing that statistically controlling for said measures has a negligible, or perhaps even an adverse, effect on the criterion validity of personality tests (Li & Bagger, 2006; McCrae & Costa, 1983; Ones et al., 1996). If social desirability scales were to be measures of response *bias* in terms of self-enhancement, then controlling for it should increase the relation between personality scales and performance ratings, since invalid measurement error variance is removed from this relation. Yet, multiple studies, including meta-analyses, have shown this not to be the case. Taken together, the results from the studies described in this section seem to suggest that social desirability scales, at least partly, capture valid personality trait variance.

Relations between measures of social desirability and higher-order factors of personality

If social desirability measures tap into personality traits, then which traits are these? And where in the personality trait domain can they be located? These questions are relevant given that if the general factor in personality questionnaires and social desirability scales both measure 'true' social desirability, then they should theoretically and empirically be connected. Previous research has indeed shown moderate positive relations (average around r = .30) between the GFP and social desirability scale scores (Dunkel et al., 2016; Erdle & Rushton, 2010; Rushton & Erdle, 2010; Schermer, Carswell, & Jackson, 2012; Schermer & Goffin, 2018; Schermer, Holden, & Krammer, 2019a; Schermer & MacDougall, 2013; Schermer & Vernon, 2010).

Recently, Connelly and Chang (2016) showed with meta-analytic multi-trait-multi-method information that social desirability scales largely capture individual differences in Emotional Stability, Agreeableness and Conscientiousness. Interestingly, the three aforementioned Big Five factors together form the proposed higher-order personality factor Stability (DeYoung et al., 2002), also known as Alpha (Digman, 1997). The other higher-order factor in the models by DeYoung et al. (2002) and Digman (1997) is formed by Openness and Extraversion and is labeled Plasticity (or Beta). Stability represents the tendency to show prosocial, socially desirable behavior, while Plasticity reflects the tendency to seek new and pleasuring experiences, and are believed to reside

at an intermediate level in the personality trait hierarchy between the Big Five and the GFP (DeYoung et al., 2002; Van der Linden et al., 2010).

Stability and Plasticity are closely related to the two-dimensional (*egoistic* vs. *moralistic*) model of social desirability proposed by Paulhus and John (1998). Egoistic bias is the tendency of stressing one's exceptional qualities and social and intellectual status (claiming to have attributes of a "superhero"). Egoistic bias is driven by the need for agency, i.e., the need for status, personal growth and achievement (Paulhus & John, 1998; Wiggins, 1979). This is opposed to moralistic bias with tendencies related to claiming to have an overly large ability to control malicious impulses (i.e., to have "saint-like" attributes). Moralistic bias is driven by the need for communion, i.e., the need for affiliation, intimacy, belonging and social relationships (Paulhus & John, 1998). Egoistic bias is conceptually related to Openness and Extraversion, thus the Plasticity factor. Moralistic bias mostly relates to Conscientiousness and Agreeableness, and thus largely overlaps with Stability. Taken together, social desirability appears to be conceptually linked to higher-order traits, establishing its location at higher levels in the personality hierarchy.

Finally, a series of studies by Uziel (2010a, 2010b, 2014) reinforce the idea that social desirability scales measure content overlapping with higher-order factors such as the GFP. After reviewing the literature on impression management scales, Uziel (2010a) concluded and subsequently showed (e.g., Uziel, 2010b, 2014) that these should be reconsidered as measures of interpersonally oriented self-control. In his definition, those with high scores on IM scales possess the self-regulatory capacity to choose the appropriate (i.e., socially desirable) act, especially in social contexts where rewards (and punishments) for behavior are notably high. Clearly, this concept of self-control in social contexts is closely related to the assumed social skills and emotional intelligence underlying the substantive definition of the GFP as a social effectiveness factor.

In sum, it appears that social desirability scales measure substantive personality characteristics rather than response sets, located at higher levels in the personality trait domain. If social desirability can be considered a trait, then it should be relatively stable across situations and contexts, and also should play a role when applying for a job. The third and final research stream discussed here focuses specifically on identifying the desired responses and behaviors – and subsequently acting on this information – in selection contexts.

Knowing What to Do in Selection Situations

This third stream comes from the personnel selection and applicant faking literature, and provides further insights into the question whether social desirability is a concept related to response distortion, or rather a trait related to social skills and competences. Kleinmann and colleagues (Klehe et al., 2012; Kleinmann et al., 2011; König, Melchers, Kleinmann, Richter, & Klehe, 2007; Melchers et al., 2009) have outlined the concept of the *ability to identify criteria* (ATIC) in selection procedures. ATIC is defined as a person's ability to correctly perceive performance criteria in evaluative situations such as assessment center exercises or job interviews. ATIC is proposed to be a cognitive social

competence associated with the ability to perceive, interpret, and act on situational cues in evaluative situations. As such, the concept of ATIC is explicitly formulated as a social effectiveness construct (Klehe et al., 2012). In line with it being a social effectiveness construct, ATIC has been proposed and shown to be related to job performance (Kleinmann et al., 2011); because work situations are often ambiguous and thus also require skills and abilities for the interpretation of situational cues to find out what behavior is required, ATIC can be expected to positively influence both performance in selection procedures and on the job (Hogan & Shelton, 1998; Kleinmann et al., 2011; Marcus, 2009; Ones et al., 1996; Viswesvaran & Ones, 1999).

An important mechanism in the concept of ATIC and its criterion validity is behavioral consistency in both selection procedures and at work. This concept of consistency is an important feature of the socio-analytic theory by Hogan and colleagues (Hogan, 1982, 1991; Hogan & Shelton, 1998; Hogan & Holland, 2003), a theory with implications for social desirability and applicant faking. This theory states that individuals differ primarily in their strategies to *get along* and *get ahead* in life. At a fundamental and unconscious level, people are motivated by their needs for attention, approval, and acceptance (getting along) and status, power, and resource control (getting ahead). Note that getting along and getting ahead align with respectively the need for communion and need for agency discussed in the previous section. Unsurprisingly then, it has been argued that *getting along* is associated with the aforementioned higher-order Stability personality factor, and that the motive for *getting ahead* is captured by Plasticity (Hogan & Holland, 2003).

As humans are social by nature, attainment of both goals inevitably requires social interactions. Consequently, those who are better able to deal with social interactions will be those who are more likely to achieve their goals. Personality in the form of reputation is crucial in this regard. Successful or socially effective people know how to manage their reputations and manage it constantly during social interactions (Goffman (1959) as cited in Hogan & Blickle, 2018). Finally, people with higher levels of social skills are assumed to be better able to manage their reputation by reading emotional and social cues, and having the self-control to choose the appropriate behavioral responses (Hogan & Shelton, 1998).

Applied to the selection context, socio-analytic theory specifically describes the process of responding to a personality questionnaire in the selection context as any social interaction in which the respondent is always trying to portray a certain image or reputation of the self (Hogan & Blickle, 2018). This idea is in line with the interpretation that response styles, such as social desirability, function as culturally influenced "communication filters", that is how an individual expresses oneself to others (He & Van de Vijver, 2013; Smith, 2004). When faced with a personality questionnaire in a selection procedure, people present an image in line with the impressions or reputation held by others, and the success of these forms of self-presentation depends on one's social skills. That is, higher levels of socio-emotional knowledge and skills will allow people to show more socially desirable behavior – both in selection procedures and on the job – to maintain a good reputation, hereby increasing their chances of achieving socially valued goals (Hogan, Barrett, & Hogan, 2007; Hogan & Blickle, 2018; Kleinmann et al., 2011; Marcus, 2009).

Thus, the crucial argument in this third stream of research on social desirability is that individuals with higher levels of social skills or social effectiveness would engage and be more effective in impression management (i.e., act more socially desirable) not only during a selection procedure but also in their everyday (social) life and on the job (Blickle, Momm, Schneider, Gansen, & Kramer, 2009; Ingold, Kleinmann, König, & Melchers, 2015; Kleinmann et al., 2011; Marcus, 2009).

In sum, this line of thought states that the social skills and competences - of which the ability to identify criteria in selection procedures is one example - underlying social effectiveness positively influence performance in selection procedures and on the job. This reasoning fits with self-presentational theories of social desirability, which argue that people are always concerned with conveying a certain image of themselves (Hogan & Blickle, 2018). Taken together, these arguments and findings seem to suggest that being socially desirable requires having social and emotional skills, assets which we can expect to have positive effects on attaining personal and social goals, such as getting a job, and performing well on the job.

Aims and Overview of Research Questions

In the previous sections, different pieces of evidence for the substantive interpretation of social desirability and its conceptual and empirical links with higher-order personality factors have been brought to the fore. Yet, additional evidence for the substantive interpretation is needed, given that different opinions on and interpretations of social desirability still exist, with some scholars providing an artefactual explanation for the construct. Given such different opinions, the current dissertation aims to provide further tests of the nature of the socially desirable component in personality scales and hereby to contribute novel insights to the substance versus artefact debate. As noted, in the present dissertation the topic of social desirability will be viewed from the perspective of organizational psychology and personnel selection. The main research question of this dissertation thus states:

> *Research Question: Can social desirability in self-report personality questionnaires be regarded as a substantive factor in personnel selection?*

This general research question is applied to the context of selection and assessment in the specific research questions distinguished below. A description and discussion of these research questions, as well as how each chapter aims to answer them, is provided in the following sections.

Criterion Validity

From the perspective of selection and assessment, it is important to show whether social desirability relates to relevant outcomes in the work context and beyond. That is, for it to be useful construct in selection procedures, it would need to show relations to, for example, job performance

or organizational citizenship behavior. Therefore, Research Question 1 and 2 both focus on the criterion validity of the social desirability component in personality questionnaires (which will be labeled as the GFP). Specifically, Research Question 1 focuses on the criterion validity of the GFP in the work domain, while Research Question 2 relates to more general criteria in individuals' lives such as their social interactions and well-being.

Research Question 1: Does social desirability in self-report personality questionnaires predict work outcomes?

It can be expected that a general social desirability or social effectiveness factor will have a broad influence on work outcomes, because in virtually all jobs, interacting with people will be necessary to varying degrees. Thus, if the GFP represents the tendency to act in socially desirable ways, then it should facilitate cooperation with colleagues, handling customers and clients, and reacting in proper ways to supervisors, resulting in better performance ratings or in objective terms lead to, for example, more sales (Sitser, Van der Linden, & Born, 2013). In addition, previous studies have shown that the GFP is associated with higher levels of self-esteem and well-being; these higher levels of self-esteem and well-being will presumably also lead to reduced problems and increased performance at work (Judge & Bono, 2001).

In order to answer this first research question, the relations between the GFP and the several work-related outcomes are tested based on meta-analytic data on the relation between the Big Five personality traits on the one hand and job performance, leadership, organizational citizenship behavior, and counterproductive work on the other. This study is reported in **Chapter 2**. The added value of this study to the literature is that it provides a direct test of the relation between the GFP and work-related outcomes. Given that the GFP literature is relatively young (about 10 years), relatively little studies have directly investigated the role of the GFP in the workplace, even though a few primary studies allude to positive relations between the GFP and relevant work outcomes such as job performance (Van der Linden et al., 2010a) and leadership (Van der Linden et al., 2014a). At the same time, decades of literature and multiple meta-analyses have been devoted to the relation between the Big Five and work-related outcomes. As the GFP represents the shared variance among the Big Five domains, it is possible to extract the GFP from the Big Five and subsequently investigate the relation between the GFP and work-criteria in order to arrive at conclusions on the criterion validity of the GFP in work settings. **Chapter 2** presents a study that follows this procedure by fitting structural models on previously published meta-analytic data in order to test the relation between the GFP and job performance, leadership, organizational behavior and counterproductive work behavior. The strength of the relation between the GFP and job performance is compared across different job types, because previous studies have found that both personality and social desirability (as a trait) have differential predictive value in different types of jobs (Barrick & Mount, 1991; Ispas et al., 2014; Shaffer & Postlethwaite, 2013).

In addition, the relative importance of the GFP vis-à-vis the Big Five dimensions in the prediction of outcomes is investigated. Whenever a higher-order factor is proposed, it is relevant to test whether this higher-order factor (i.e., the GFP) has unique or added predictive value in comparison with its lower order constituents (Johnson, Rosen, & Chang, 2011). That is, in more practical terms, if the GFP does not increase the prediction of job performance beyond, the lower order Big Five dimensions, then the GFP will be less attractive for practitioners to use in selection procedures.

Finally, considering the previous discussion on the meaning of social desirability scale scores, the predictive power of the GFP relative to such scores is examined. Traditionally, as mentioned before, in the selection literature, social desirability scale scores have been taken as indicators of response bias, and thus partialled out from the relation between personality traits and the criterion to test whether this affects the strength of the associations found (e.g., Li & Bagger, 2006; McCrae & Costa, 1983; Ones et al., 1996). Although this notion of measures of bias is disputed in the literature, the same method is adopted in the study described in **Chapter 2**: it is tested whether the GFP-job performance relationship changes when social desirability as measured by social desirability scales is controlled for.

If the GFP indeed represents social effectiveness, then this should not only be reflected in higher levels of job performance or OCB, but also in the daily social experiences of people. More specifically, given their social-emotional skills and proclivity for showing socially desirable behaviors, we can expect people scoring high on the GFP to be more effective in their daily social interactions than their low-GFP counterparts (Van der Linden et al., 2010b; Dunkel et al., 2018). This notion is formulated in Research Question 2:

> *Research Question 2: Does social desirability in self-report personality questionnaires relate to (daily) criteria of social effectiveness?*

Research Question 2 is answered in **Chapter 3**, in which the relations between the GFP and indicators of social effectiveness at the daily level – that is: relationship quality, relationship quantity, interpersonal conflict, and impressions made on others – are investigated. Although previous studies have provided indirect evidence for such relations (e.g., Lopes et al., 2004; Van der Linden et al., 2010b), **Chapter 3** provides the first direct test of their existence. Important from a personnel selection point-of-view in **Chapter 3**, finding a positive relation between the GFP and (daily) impressions on others would suggest how higher GFP scores could lead to obtaining a desired job (see Van der Linden, Te Nijenhuis, Cremers, Van de Ven, and Van der Heijden-Lek, 2014b) given that leaving a good impression – either in the selection interview (e.g., Levashina, Hartwell, Morgeson, & Campion, 2014) or through responses on a personality questionnaire (Klehe et al., 2012; König, Melchers, Kleinmann, Richter, & Klehe, 2007; Roulin, Krings, & Binggeli, 2016) – on employers in the selection procedure is crucial for acquiring a desired job.

The study described in **Chapter 3** proposes and tests a theoretical mechanism for the relatively strong relationship (*r* typically between .40 and .50; Dunkel et al., 2016; Erdle, Irwing, Rushton, & Park, 2010; Musek, 2007; Şimşek, 2012) found previously between the GFP and subjective well-being such as self-esteem and positive mood. Social relationships are strongly related to subjective well-being (e.g., Argyle, 2001). Therefore, we can expect that the social skills associated with high-GFP individuals allow them to maintain better social relationships which in turn result in higher levels of well-being. In other words, we expect a mediation of the relation between the GFP, and well-being and mood by daily social interactions.

In addition to this mediation hypothesis, a hypothesis on moderation is proposed and tested in **Chapter 3**. Personality traits (Neuroticism and Extraversion specifically) have previously been associated with respectively increased and reduced sensitivity to negative daily social experiences such as interpersonal conflict. A similar moderating effect of the GFP on the relation between daily social experiences and daily well-being is formulated in **Chapter 3**. It is expected that, due to their social effectiveness, higher GFP scores may be positively related to the ability to deal with or react to negative interpersonal events (e.g., Dunkel & Van der Linden, 2014; Hengartner, Van der Linden, Bohleber, & Wyl, 2017). For example, during a conflict, higher scores on the GFP level may imply that one better regulates social and emotional behavior, and chooses the appropriate reaction, thereby providing a solution to the conflict. Therefore, we expect that higher GFP scores are associated with smaller declines in daily well-being after a conflict.

The study described in **Chapter 3** is based on secondary data from the Berlin Diary Study project by Denissen and colleagues (2005 – 2009), which contains a large sample of respondents who provided diary reports of their daily social experiences and daily well-being. It has been argued that data collection through diaries has several advantages over one-time, cross-sectional methods (Bolger, Davis, & Rafaeli, 2003), one of which is the reduction of social desirability bias (Barta, Tennen, & Litt, 2013). Naturally, this reduction is important because of the artefactual account of the GFP as a social desirability *bias* factor. By using diary data, relations found in **Chapter 3** are expected to be to a lesser extent influenced by inflated self-ratings.

Antecedents

If we assume that the socially desirable component in personality measures represents a substantive and stable trait, questions still remain about this component's antecedents. Previous research literature has argued that social desirability should be predicated on social knowledge and social skills (e.g., Argyle, 1969; Hogan & Shelton, 1998; Marcus, 2009; Roulin et al., 2016). Our third research question thus states:

> *Research Question 3: Does social desirability in self-report personality questionnaires in the selection context relate to social competences?*

In order to be able to show socially desirable behaviors, one needs to be able to read (emotional) cues of others, decode this information, and choose the appropriate type of response given the social context (Hogan & Shelton, 1998; Roulin et al., 2016). Social skills can thus be seen as a prerequisite for reading, understanding, and controlling social interactions in order to be socially effective (Ferris, Witt, & Hochwater, 2001). Although multiple operationalizations for social skills have been used in earlier research, for the current discussion the construct of trait emotional intelligence is most relevant. Social skills and competences are incorporated in trait EI as the latter also relates to understanding one's own emotions and those of others, and the tendency to use this knowledge act in socially effective ways. As noted previously, trait EI has been shown to largely overlap with the GFP conceptually and empirically (Van der Linden et al., 2017). In addition, studies have shown positive associations between trait EI and scores on social desirability scales (e.g., Kluemper, 2008; Mikolajczak, Luminet, Leroy, & Roy, 2007; Van der Linden, Tsaousis, & Petrides, 2012b). Finally, the interpretation of impression management scales as measures of interpersonally oriented self-control by Uziel (2010a) can be linked to emotional intelligence, since in virtually all models of EI, self-control is proposed to be an important component of EI.

One crucial aspect of emotional intelligence is the ability to read emotional cues, interpret them, and act or respond accordingly. In the current dissertation, we propose that these processes are all at work when responding to a personality questionnaire in a selection context (Hogan et al., 2007). More specifically, when applying for a job, it is not entirely clear what the hiring company is looking for exactly. Therefore, the applicant will need to interpret the situational cues at hand (e.g., personality test items) in order to leave a good impression (Roulin et al., 2016). It therefore seems reasonable to expect the aforementioned EI-related abilities and skills to increase one's chances of providing the desirable responses during a selection situation, which has previously been labeled the *ability to fake* in the literature (e.g., Raymark & Tafero, 2009). This expectation will be tested in **Chapter 4**. The aim of this chapter is thus to show that social desirability relates to behavior related to social skill associated with being able to detect which responses are desirable in selection situations.

Measures of social effectiveness in general (Ferris et al., 2002; Melchers et al., 2009), and emotional intelligence in particular, have previously been criticized for a lack of incremental validity over personality traits, such as the Big Five dimensions, and cognitive abilities in the prediction of a number of outcomes (MacCann, Matthews, Zeidner, & Roberts, 2003; Schulte, Ree, & Carretta, 2004). To address this concern, in **Chapter 4** we aim to show incremental validity of emotional intelligence in the prediction of faking ability over and above the Big Five personality traits and cognitive abilities.

A third and final aim is to provide an answer to the recurring question in the faking literature on the effect of response distortion on the criterion validity of personality questionnaires. Although the claim has been made that faking would lead to a reduction in criterion validity, evidence for this position has been mixed (Barrick & Mount, 1996; Dilchert, Ones, Viswesvaran, & Deller, 2006; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Morgeson et al., 2007; Ones et al., 1996). We aim

to provide a theoretical explanation for this mixed evidence: if the social skills that influence one's ability to fake on personality inventories during the selection process at the same time positively influence performance on the job, threats to validity by faking may be minimal (e.g., Hogan et al., 2007; Kleinmann et al., 2011). Indeed, empirical evidence suggests a positive link between faking ability and job performance (e.g., Blickle et al., 2009; Klehe et al., 2012; Viswesvaran & Ones, 1999), supporting the claim that applicant behavior in selection procedures actually reflects genuine social skills and abilities or social effectiveness. Note that this line of reasoning is consistent with the claim from socio-analytic theory that responding to a personality questionnaire is a form of self-presentation like any other form of social interaction in which one tries to convey a certain impression just as in one's everyday life (Hogan et al., 2007).

Construct validity

A crucial part in the debate on socially desirable responding of applicants in personality assessments is whether it is dependent on the context and circumstances under which the assessments are administered (e.g., Anglim et al., 2017; Ellingson, Smith, & Sackett, 2001; Schmit & Ryan, 1993; Smith & Ellingson, 2002). If for example the factor that captures social desirability in personality measures changes considerably under situations with higher pressures for social desirability (e.g., selection situations), then such a factor is of little significance or practical utility.

Put differently, and important question that still remains is whether social desirability is relatively stable across different situations and contexts. In the current dissertation, the influence of test-taking context and the item format of the personality questionnaire is investigated. The fourth and final research question thus states:

> *Research Question 4: Does social desirability in self-report personality questionnaires vary with the test-taking context and item format?*

In order to answer this question, we examined the characteristics of the general social desirability factor in a selection (high stakes situations) and a career advice (lower stake situation) context and compared traditional Likert type self-reports of personality with forced-choice personality surveys. It can be expected that selection situations provide job applicants with a motivation to distort their responses because there is a job at stake and levels are social desirability are thus higher than in other, less evaluative situations. In the applicant faking literature, some studies have found that under high-stakes settings, personality traits become more correlated, increasing their overlap and hence giving rise to a large general social desirability factor (the 'ideal employee' factor; Schmit & Ryan, 1993); these findings suggest that social desirability captures situation-induced response sets rather than a substantive trait. Yet, others have found that factor structures in general, and the common factor more specifically, are robust to differences in motivational pressures for response distortion of the context (e.g., Anglim et al., 2017; Ellingson et al., 2001; Marshall, De Fruyt, Rolland,

& Bagby, 2005; Smith & Ellingson, 2002). Studies from the GFP-literature have also yielded mixed results. Recent studies (MacCann, Pearce, & Jiang, 2017; Schermer et al., 2019a; Schermer, Krammer, & Goffin, 2019b) have found the general factor to be larger in groups instructed to provide socially desirable responses than groups responding honestly, while Van der Linden, Bakker, and Serlie (2011) found that the GFPs from real selection and assessment samples were highly similar.

There are three important caveats related to previous studies on this topic. First, the personality questionnaires employed in the studies predominantly use a response format (i.e., Likert scales) which facilitates response distortion if one is motivated to do so. Consequently, when Likert-type items are used, socially desirable responding may be relatively easy and general factor scores may be more easily inflated compared to when other response formats that limit response distortion are used. The forced-choice format is such a format, reducing the opportunity to respond in a socially desirable way (Christiansen, Burns, & Montgomery, 2005; Waters, 1965). As a result, it might be that the GFP is reduced or even disappears when forced-choice formats are used (Irwing, 2013). Second, the studies that have used this format did not always include samples in which the participants could be expected to be motivated to distort responses (Irwing, 2013). In other cases, the effects of test-taking context and response format on lower-level factor structures were investigated, without looking at the effects on the general factor present in the personality measure (e.g., Joubert, Inceoglu, Bartram, Dowdeswell, & Lin, 2015). Finally, there appear to be discrepancies between studies conducted in the lab, where participants are instructed to provide socially desirable responses, and real-world studies with actual applicants (MacCann et al., 2017; Schermer et al., 2019a, 2019b; Van der Linden et al., 2011).

In **Chapter 5** we aim to address these shortcomings by investigating the combined influence of test-taking context (development vs. selection) and response scale types (Likert vs. forced-choice) on social desirability (operationalized as the GFP), using data from real applicants and career development assessments. Differences between the development and selection group can be inferred to reflect motivational differences in intentional response distortion. At the same time, the forced-choice response format supposedly reduces the opportunity to distort responses when one is motivated to do so. As such, it can be tested how motivation and opportunity for response distortion simultaneously affect the size and nature of the general factor present in the personality tests. This chapter's contribution to the literature is a comprehensive test on the construct validity of the social desirability factor by investigating its robustness across variations in situational pressure in the form of motivation (test-taking context) and opportunity (item format) to distort responses. A second contribution is that a novel statistical method, the Thurstonian IRT model (Brown & Maydeu-Olivares, 2011, 2012), is used to extract personality scores from forced-choice questionnaires.

Summary

This dissertation aims to add new pieces of information to the accruing body of evidence for the substantial interpretation of the socially desirable component in self-report personality measures, mainly from the perspective of personnel selection and assessment. In four empirical studies, the criterion- and construct-validity as well as the antecedents of social desirability as a substantive construct are examined. First, two studies focus on the criterion validity of social desirability (operationalized as the GFP) both in the work context and in people's everyday lives and social interactions. Subsequently, it is tested whether providing the desirable answers in a selection procedure is related to social skills (operationalized by trait emotional intelligence). Finally, the construct validity of social desirability is investigated in a study on the robustness of the general social desirability factor in self-report personality questionnaires across test-taking contexts and item formats. A schematic overview of the topics covered in this dissertation is presented in Figure 1.

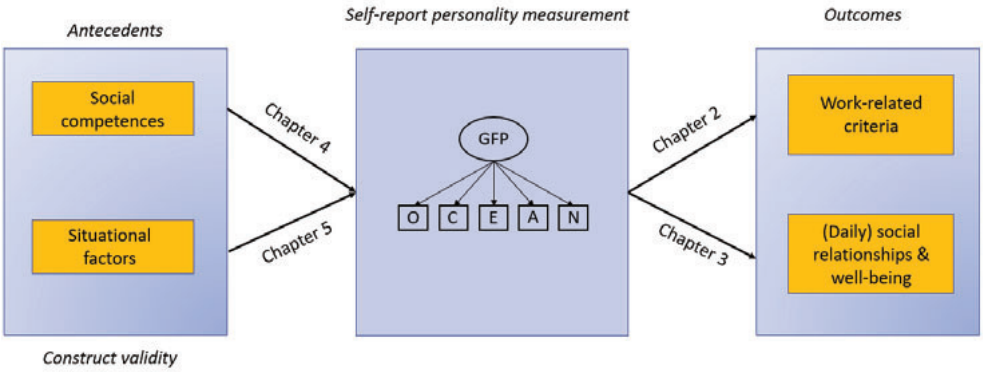

Figure 1. Overview of the empirical chapters in this dissertation. O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism.

The General Factor of Personality and Job Performance: Revisiting Previous Meta-Analyses

This chapter has been published as:
Pelt, D. H. M., Van der Linden, D., Dunkel, C. S., & Born, M. Ph. (2017). The general factor of personality and job performance: Revisiting previous meta-analyses. *International Journal of Selection and Assessment, 25,* 333-346.

This chapter has been nominated for the Best Paper Award at the 2015 conference of the Dutch Association of Work & Organizational Psychology (WAOP), 27 November 2015.

2

Abstract

The relationship between the General Factor of Personality (GFP) and several work-related outcomes such as job performance and organizational citizenship behavior was examined using meta-analytic data. Confirmatory factor analyses showed sizeable relationships between the GFP and various performance indicators ($\bar{r}$ = .34), larger than for any of the Big Five dimensions. Controlling for social desirability did not change the relationship between the GFP and job performance. Moreover, regression analyses showed that the GFP accounted for a larger part of the explained variance in the outcome measures than the unique variances of the Big Five. The results add to the evidence for the GFP as a social effectiveness factor and highlight the validity of the GFP in organizational contexts.

Introduction

A recent stream of articles has emphasized that personality traits tend to correlate with each other, leading to the emergence of a so-called general factor of personality (GFP) that typically explains somewhere between 20 to 60% of the variance in the underlying traits (Figueredo, Vásquez, Brumbach, & Schneider, 2004; Musek, 2007; Rushton, Bons, & Hur, 2008; Van der Linden et al., 2010a). This GFP captures the socially desirable ends of traits. Thus, in terms of the well-known Big Five, high-GFP individuals, on average, score relatively high on Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and are Emotionally stable (i.e., score low on Neuroticism). High scores on the GFP have been related to high levels of well-being and self-esteem (e.g., Musek, 2007). The existence of the GFP has been confirmed in two large meta-analyses (e.g., Rushton & Irwing, 2008, Van der Linden et al., 2010a). In addition, its presence is not confined to Big Five measures, but it has also been extracted from over 20 other measures of normal and abnormal personality (e.g., Rushton & Irwing, 2011). Based on these previous studies that have confirmed the presence of a general factor of socially desirable behaviors in a wide range of personality measures using various statistical methods, the GFP may now be considered a rather systematic finding (with an occasional exception, for example De Vries, 2011). Nevertheless, the literature also makes clear that there is an ongoing debate about how to measure and interpret this general factor (Anusic et al., 2009; Irwing, 2013). Some scholars have suggested that the GFP represents a substantive factor and may play an important role in understanding the structure of personality (e.g., Loehlin, 2012; Veselka, Just, Jang, Johnson, & Vernon, 2012). Others, however, adhere to the view that the GFP represents not much more than a statistical or methodological artefact and thus may not have strong theoretical or practical implications for personality theory (Ashton et al., 2009). The detailed arguments regarding the different points of views are well-documented in several previous articles (Ferguson, Chamorro-Premuzic, Pickering, & Weiss, 2011; Just, 2011; Irwing, 2013), showing that currently roughly two main interpretations – i.e. substantive versus artefact – can be differentiated.

The main substantive interpretation is that the GFP reflects general social effectiveness (Dunkel & Van der Linden, 2014; Loehlin, 2012; Van der Linden et al., 2016). This notion is supported by various types of empirical findings showing that the GFP relates to a wide range of relevant and objective or other-rated social outcomes such as popularity and likeability (Van der Linden et al., 2010b) and fewer problematic life events (Van der Linden et al., 2015). Moreover, in a recent meta-analysis, Van der Linden et al. (2017) showed a large overlap between trait emotional intelligence and the GFP (corrected r=.86). Given that high-EI individuals can utilize their emotional knowledge and skills to achieve personal and social goals, it allows them to behave in a socially desirable or effective way. As such, we can expect such knowledge and skills to have a broad influence on actions and behavior, hereby driving all personality traits into a socially desirable direction (Van der Linden et al., 2016).

The main artefact interpretation is that the GFP merely reflects response biases that are confined within a specific method of measurement. For example, when using self-reports of

personality, individual differences in the tendency to provide socially desirable answers may drive the intercorrelations between traits on which the GFP is based (Edwards et al., 1962; Paulhus, 2002). Findings that are in line with this interpretation suggest that the GFP decreases when using items that are socially neutral (Bäckström et al., 2009; Petterson et al., 2012), or when comparing self- and other-ratings of personality (Chang et al., 2012; Gnambs, 2013). A different argument has been made by Ashton et al. (2009), who suggest that the GFP is due to so-called blended facets. For example, the facet "enthusiasm" may relate to both Extraversion and Openness, thus leading to a correlation between those dimensions. Subsequently, they argue that higher-order factors above the level of the Big Five are spurious and would disappear when one controls for such blended facets.

The various measurement issues involved in the GFP have been extensively discussed in a recent review (Van der Linden et al., 2016) and meta-analysis (Van der Linden et al., 2017) on the topic, which concluded that the GFP is robust with regard to measurement or extraction method.

Given the various questions about the GFP currently addressed in the literature, such as whether the GFP represents substance or artefact, and whether the GFP has theoretical and practical value, we considered it useful and timely to test whether the construct is related to job performance, reanalyzing previous meta-analytic data. An important step in the theoretical development of a higher order construct and proving its usefulness (apart from a simple parsimony perspective) is to provide evidence for criterion-related validity (Johnson et al., 2011): in the present study we aim to do so. Based on the idea that the GFP is a substantive factor indicating individuals' general social effectiveness it can be expected that a relatively strong relationship with job performance would exists. In contrast, if the GFP would merely reflect statistical or methodological artefact confined to the method adopted to measure personality, then it would be very unlikely to have a broad and meaningful relationship to job performance.

Regarding the former interpretation, if the GFP indeed reflects social effectiveness it may relate to a broad range of outcome measures. Specifically, as humans are social by nature, the way people interact with others can be expected to influence the outcome of life areas such as who will be appointed as leader, the quality of close (e.g., romantic, family) relationships, and how to maximize the chances of reaching desired goals (e.g., getting the job one wants) in general. Being socially effective implies that one has the knowledge, motivation, and competencies in order to display adequate social behavior in different contexts (Ferris et al., 2002). For example, even though one may feel nervous when facing a stressful situation, one might be able to regulate or control behavior and keep one's cool. Another typical example is that when being angry at one's manager, one calmly expresses dissatisfaction instead of acting in rage. The relation between the GFP and measures of EI (Pérez-González & Sanchez-Ruiz, 2014; Van der Linden et al., 2012b) previously referred to above fit well with this idea.

Thus, based on the ideas described above it can be expected that the GFP would also relate to job performance. Obviously, job performance depends on various non-social factors such as the amount of job knowledge and experience (Schmidt & Hunter, 1998). Yet, it is widely acknowledged

that in almost every job, social aspects play a relevant role in performance. Hogan and Holland (2003) argued that in the vast majority of jobs, the ability to *get along* with others is imperative for achieving positive professional outcomes, such as obtaining favorable ratings from supervisors or peers, or establishing better connections to customers. This is what one would expect from an individual that acts in socially effective ways and is able to obtain desired social goals. For example, an individual who knows how to best approach others, or what type of behavior to reveal in specific situations may, on average, have a social advantage that can facilitate work performance (Ferris et al., 2001).

To the best of our knowledge, there are currently only two published studies that have *directly* examined the relationship between the GFP and other-rated or objective indicators of job performance. In a sample of 144 employees with various professional backgrounds, Van der Linden et al. (2010a; Study 2) found that the GFP was correlated $r = .23$ (observed correlation, not corrected for artefacts) with supervisor-rated performance. Moreover, they found that the unique variance of the Big Five dimensions did not significantly contribute to predicting performance beyond the effect of the general factor. In a sample of 433 sales employees, Sitser et al. (2013) found that, compared to the Big Five dimensions, the GFP was among the highest and most consistent predictors of supervisor-rated and objective sales performance (uncorrected $r = .20$ for the supervisor rated performance as well as for an objective performance measure). Averaged over these two studies, the relation between the GFP and job performance thus seems to be around $r = .20$.

Two other studies on the criterion validity of the GFP using other-ratings are worth mentioning here. Van der Linden et al. (2014b) showed that high-GFP individuals received higher performance ratings by interviewers in selection interviews ($r = .23$) and higher integrity ratings by supervisors ($r = .21$) than low-GFP individuals. This is a relevant finding given that integrity is a well-known predictor of job performance (Ones, Viswesvaran, & Schmidt, 1993). Finally, Van der Linden et al. (2014a) showed that the GFP was associated with supervisor/peer-rated leadership skills ($r = .22$) and having (versus not having) leadership experience ($d = .74$ in Study 1 and $d = .71$ in Study 2).

Although these studies suggest that the GFP indeed relates to job performance and related outcomes, they currently provide only a very limited set of empirical data compared to, for example, the much larger amount of studies on the Big Five and job performance (see for an overview, Barrick and Mount, 1991 and Barrick et al., 2001). Therefore, conclusions about the relationship between the GFP and job performance would be strengthened if they would be based on a much larger collection of studies and outcome measures. Subsequently, we decided to test the GFP-job performance relationship by reanalyzing relevant meta-analytic data from the literature. Specifically, several leading meta-analyses have appeared in the literature testing the relationship between the Big Five and various job performance measures. Based on these meta-analyses combined with meta-analytic information about the Big Five intercorrelations, it is possible to examine the personality-performance relationship from a GFP perspective. Accordingly, we used

the empirical literature to extract the relationships between the Big Five and various aspects of job performance and use them for novel tests on the relationship between the GFP and job performance (see also the Method section below).

First, we included a wide range of job performance studies as described in the seminal meta-analysis of Barrick et al. (2001), who collected all available studies on the Big Five and various categories of job performance measures. In addition to these direct performance measures we also decided to include meta-analytic data on Organizational Citizenship Behavior (OCB), which reflects behaviors not formally included in job descriptions that support the social context of an organization (e.g. helping a colleague). By definition, OCB is about helping colleagues or showing involvement in the company in which one works. As these OCB aspects of performance can be assumed to contain a relatively large social component, they may be useful to test in relation to the presumed socially effective behavior indicative of the GFP.

A negative side to vocational behavior that has received a relatively large amount of attention in the literature recently is counterproductive work behavior, i.e., deviant behavior in the workplace (see for example Rotundo and Spector (2010) for an overview). The relationship between counterproductive work behavior and social effectiveness may be less obvious than the previous two aspects of job performance. Yet, in as far as the GFP is also associated with socially desirable or acceptable behavior a negative relationship with counterproductive behavior can be expected, although testing this is admittedly more explorative compared to job performance and OCB. Berry, Ones, and Sackett (2007) used a meta-analytic approach to test the relationship between the Big Five and counterproductive work behavior.

We selected the three types of job performance behaviors and their corresponding meta-analyses as they provide a fairly comprehensive set of measures from the job performance domain (see also Ones, Dilchert, Viswesvaran, and Judge, 2007).

In addition, we decided to include information about leadership. Although leadership is not a direct measure of job performance, it is included in several meta-analyses on the relationship between job outcome measures and personality (Bono & Judge, 2004; Judge, Bono, Ilies & Gerhardt, 2002; Ones et al., 2007). With regard to the GFP, leadership is a relevant outcome variable, as due to the presumed higher levels of social effectiveness of high-GFP individuals, they may have a higher probability of obtaining, or being selected for, a leadership position (Figueredo & Rushton, 2009; Van der Linden et al., 2014a). Moreover, previous studies have predominantly linked social effectiveness to managerial job performance, acknowledging the interpersonal nature of the job and the fact that a good leader or manager requires a set of social skills (Ferris et al., 2002; Semadar, Robins, & Ferris, 2006). In their meta-analysis, Judge et al. (2002) examined the relationship between the Big Five and an overall measure of leadership, a composite based on leader emergence and leadership effectiveness. Leadership effectiveness refers to a leader's performance in influencing and guiding his or her followers toward achievement of the set goals. Leader emergence is about whether (or how much) an individual is viewed as a leader by others. Bono and Judge (2004) reported and examined meta-analytic correlations between the Big Five and transformational

leadership, a term used to describe charismatic and inspirational leaders who lead their followers with a clear vision.

When these previous meta-analytic findings on the Big Five and performance (or related criteria) are combined with meta-analytic findings on the Big Five intercorrelations (Van der Linden et al., 2010a) they allow tests of the relationship between the GFP and performance using confirmatory factor analysis and regression techniques (Viswesvaran & Ones, 1995; see also the Method section). Combining meta-analyses with such methods allows researchers to ask and answer questions and test models not addressed in the primary studies (Landis, 2013; Viswesvaran & Ones, 1995). Due to this advantage, combining data from several meta-analyses is becoming a commonly used strategy to thoroughly test relationships between (higher-order) constructs (e.g. Chang et al., 2012; Chiaburu, Oh, Berry, Li, & Gardner, 2011; Gnambs, 2013; Van der Linden et al., 2017).

All in all, testing the GFP-performance relationship with meta-analytic data provides useful insight into the validity and the practical implications of the GFP and in addition may provide useful empirical information in the current GFP debate. Specifically, if the GFP would be entirely artefactual and does not reflect how people would genuinely and consistently behave, it can be expected to show no meaningful relationships to other-rated or objective measures of behavior. Method artefacts, confined to specific circumstances or contexts, such as filling out personality questionnaires, generally do not show such type of relations. In contrast, a substantive personality factor with a large impact on how one interacts with others is likely to show clear relations with broad measures of performance.

Method

Meta-analytic Procedure

In order to test GFP-performance associations, personality-performance relationships as reported in previous meta-analyses were used, combined with meta-analytic findings on the Big Five intercorrelations. Big Five intercorrelations were obtained from the meta-analysis of Van der Linden et al. (2010a). They collected the Big Five intercorrelation matrices published in scientific peer-reviewed journals between 2000 and 2008, leading to 212 matrices representing a total N of 144,117 participants (sample sizes varied from 39 to 21,105, with a mean sample size of N = 679.8, median N = 233.5). The Big Five intercorrelations from this meta-analysis (corrected and uncorrected) are reported in Table A of the supplemental materials.

In their study, Van der Linden et al. (2010a) found that the model with the best fit was a hierarchical one in which the Big Five loaded on two higher-order factors namely Stability and Plasticity (DeYoung et al., 2002; also referred to as Alpha and Beta in the literature; Digman, 1997). Stability represents the tendency to show prosocial, socially desirable behavior with Conscientiousness, Agreeableness, and Emotional Stability loading positively on this factor. Plasticity reflects the tendency to seek new and pleasuring experiences and is formed by loadings

of Extraversion and Openness to Experience. The two higher-order factors then loaded on the GFP (see Van der Linden et al., 2010a, Figure 1, p. 319).

Using the Big Five intercorrelations from this previous meta-analysis has the advantage that it is based on a very large number of studies, larger than if we would only consider the Big five intercorrelations that are reported in the studies on each of the performance criteria. Thus, the Van der Linden et al. (2010a) study provides us with better and more stable estimates of those intercorrelations. It is therefore reasonable to assume that these correlations represent the 'true' correlations between the variables of interest.

For the correlations between the Big Five and job performance indicators, we used the meta-analysis from Barrick et al. (2001) for measures of overall job performance, supervisor rated performance, objective performance, team performance and training performance. Overall job performance included other-ratings as well as productivity data. Objective performance is based on productivity data, turnover, promotions and salary measures. Team performance includes measures such as ratings of cooperativeness, the quality of interpersonal relations and the ability to work with others (Hough, 1992; Mount, Barrick, & Stewart, 1998). Training performance consisted mostly of training performance ratings (Barrick & Mount, 1991).

We further used the Barrick et al. (2001) meta-analysis to examine performance in specific occupations of sales, managers (ranging from low to high level), police officers, professional (e.g., engineers, architects, accountants) and skilled/semi-skilled workers (e.g. clerical workers, flight attendants, medical assistants, assemblers, grocery clerks, truck drivers) to investigate the validity of the GFP across different jobs.

The second-order meta-analysis by Barrick et al. (2001) was an extensive study by all means; it included a large number of samples resulting in large overall *Ns*: e.g. for overall job performance the number of samples *k* ranged from 143 with *N* = 23,225 (for Openness to Experience) to *k* = 239 with *N* = 48,100 (Conscientiousness). The study dates from 2001; however, to our knowledge it is, to date, the most extensive meta-analysis on the topic. In addition, there is no reason to assume that the relationship between basic constructs such as personality and job performance has changed since then. In fact, the study was a second-order meta-analysis (i.e., a meta-analysis on previous meta-analyses) in which Barrick and colleagues already concluded that there were little differences in terms of effect sizes between the different primary meta-analyses. This notion is further supported by the more recent study of Judge, Rodell, Klinger, Simon, and Crawford (2013). They also conducted a meta-analysis on the personality and job performance relationships starting from the lower facet level of personality, yet they also considered the higher level Big Five factors. The Big Five-job performance associations reported in that study were highly similar to the ones reported in the 2001 meta-analysis by Barrick et al. (average corrected correlations of .15 (2001) versus .16 (2013), with a maximum difference of only .05 for the Extraversion-performance relation, i.e. .15 in 2001 versus .20 in 2013). Nevertheless, as the meta-analyses from 2013 contained a more restricted set of performance measures we considered it more informative to use the elaborate range of the 2001 meta-analytic results instead of the smaller set in the 2013 study (showing similar

values anyway). Yet, to be sure that the choice of meta-analytic estimates did not affect our results, we ran parallel analyses using values from several different meta-analyses as input for our models. These analyses, described in the Appendix, showed that our results and substantive conclusions remained the same regardless of which meta-analysis was used.

The meta-analysis by Chiaburu et al. (2011) was used for measures of organizational citizenship behavior (OCB). Besides an overall composite of OCB (*k*s in relation to personality measures ranged between 36 and 71, *N*s between 6,700 and 14,355), they distinguished between organization-directed (OCB-O; *k*s ranging between 7 and 20, *N*s between 1,311 and 4,598), individual-directed (OCB-I; *k*s ranging between 10 and 28, *N*s between 2,049 and 6,347) and change-oriented (OCB-CH; *k*s ranging between 6 and 19, *N*s between 1,144 and 3,761) forms of citizenship behavior. OCB-I and OCB-O concern prosocial behavior contributing to the social work environment, while OCB-CH concerns proactive behavior directed at making valuable changes and improvements in the organization (Chiaburu et al., 2011). They only included studies in which citizenship behavior was based on other-ratings (thus, no self-report OCB measures were included).

Correlations between the Big Five and counterproductive work behavior were taken from the meta-analysis by Berry et al. (2007). These authors distinguished between interpersonal deviance (*k*s ranging between 8 and 11, *N*s between 2,360 and 3,458), which is targeted towards individuals (e.g. workplace bullying, gossip) and organizational deviance, targeted at the organization (e.g. bullying, bad-mouthing the organization, stealing from the company). They reported correlations between the Big Five on the one hand, and self-report and non-self-report CWB measures on the other hand. No significant differences were found between the correlations based on self-reports and non-self-reports. Therefore, we used the correlations based on the full sample of studies.

The meta-analyses by Judge et al. (2002; leadership composite, leader emergence and leader effectiveness, *k*s ranging between 35 and 60, *N*s between 7,221 and 11,705) and Bono and Judge (2004; transformational leadership, *k*s ranging between 18 and 20, *N*s between 3,338 and 3,916) were used for the leadership criteria. Both meta-analyses excluded studies that used self-report measures of leadership: ratings of leadership behavior had to be provided by observers.

We also included meta-analytic information about the Big Five and academic performance (*k*s ranging between 109 and 138, *N*s between 58,522 and 70,926) from Poropat (2009) as a test for discriminant validity. As academic performance relatively strongly depends on cognitive abilities and is less strongly influenced by personality, we expected the GFP-academic performance association to be weaker than the GFP-job performance associations.

Finally, the meta-analysis by Li and Bagger (2006) was used for the correlations between the Big Five and socially desirable responding; they reported correlations between the Big Five and both scales of the *Balanced Inventory for Desirable Responding* (BIDR; Paulhus, 1991), self-deceptive enhancement (SDE, *k*s: 16-26, *N*s 2,881-4,361) and impression management (IM, *k*s: 18-27, *N*s 3,223-4,978). These correlations were used to examine whether the relationship between the GFP and the performance criteria were altered after controlling for social desirability. Given the

alternative explanation of the GFP as a social desirability bias factor (e.g., Bäckström et al., 2009), we considered it relevant to include such tests. Given that Li and Bagger (2006) only considered a single and broad job performance variable (no distinction between subtypes of jobs or performance) in their study, we decided to conduct the social desirability analyses only on the overall job performance measure.

Statistical Analyses

The meta-analytic correlation matrices extracted from the studies described above (see Table A in the supplemental materials), were used as input for Structural Equation Modeling (SEM). Based on previous studies (e.g., DeYoung, 2006; Digman, 1997; Rushton & Irwing, 2011; Van der Linden et al., 2010a) we compared several alternative models, namely:

1. A hierarchical model, with the Big Five loading on the intermediate Stability and Plasticity factors, which then load on the GFP which correlates with the performance criterion.
2. A single higher-order factor (GFP) that is directly extracted from the Big Five and correlates with the performance criterion;
3. Two independent higher-order factors, Stability and Plasticity, directly correlating with the performance criterion;
4. Independent Big Five dimensions relating to the specific performance criterion; this model is equivalent to a simple linear regression model in which the performance criterion is regressed on all the Big Five personality dimensions simultaneously;

Noteworthy here is that in testing these four models, we follow the ideal "strictly confirmatory" strategy as outlined by Jöreskog (1993), basing the models on prior theory and research and testing them in a different sample to assess whether they can be confirmed and thus show generalizability.

For each of the performance criteria, the fit of the four models were compared in terms of their values for the comparative fit index (*CFI*; Bentler, 1990), Tucker-Lewis index (*TLI*; Tucker & Lewis, 1973), root mean square error of approximation (*RMSEA*; Steiger, 1990) and standardized root mean square residual (*SRMR*) to examine which model described the data best. The guidelines on thresholds for adequate fit were followed, as reported by Hu and Bentler (1999). The 90% confidence intervals of the *RMSEA* values are also reported, providing information on the accuracy of the estimate. The lower limit should be close to 0 while the upper limit should be less than .08 in order to indicate good fit (Hooper, Coughlan, & Mullen, 2008). The values of the correlations between the GFP and the criteria are reported for the best fitting model.

To substantiate the usefulness of a higher-order construct, the extent to which it adds to the criterion validity of lower-order traits should be compared to the criterion validity of the unique variance of the lower-order traits (e.g., Jensen, 1998). Therefore, we examined the relative contribution of the Big Five beyond the GFP in two ways. First, regression analyses were conducted to examine the incremental criterion validity of the individual Big Five factors above and beyond the GFP. Second, we calculated partial correlations in which social desirability and impression

management measures were controlled for when testing the GFP and overall job performance relation.

For the *N*s of the correlation matrices, we used the lowest reported *N* value of the sample sizes for the correlations in the meta-analytic correlation matrix. Some authors have advocated using the harmonic mean of the sample sizes (e.g., Landis, 2013); however, given that the harmonic mean will always be larger than the smallest sample size, we chose to use the lowest reported *N* being the most conservative value. The main analyses were conducted on these matrices and their results are reported here. We analyzed both the uncorrected and the corrected correlation matrices. The Big Five intercorrelations were corrected for unreliability, range restriction, and sampling error (see Van der Linden et al., 2010a). The meta-analyses on the relations between the Big Five and the performance criteria adopted different approaches for correction of the correlations, i.e. either only for predictor unreliability or criterion reliability or both. The sample weighed mean correlations were used for the uncorrected correlations.

Results

Model Fits

Table 1 shows the results from the SEMs. The analyses revealed the same general pattern for the majority of the performance criteria. Generally, and in line with the results of Van der Linden et al. (2010a), the hierarchical model (Figure 1) showed the best fit to the data, with the fit indices indicating adequate to good fit. The other models showed lower fit to the data and generally ranged from adequate (i.e., the direct GFP model) to poor fit indices (i.e., the orthogonal two-factor model, and the orthogonal Big Five model). Due to the large number of models and fit indices we tested, we only report the results of the best fitting hierarchical model in Table 1. Specific results of the other models can be obtained upon request from the first author.[1]

[1] Mean *SRMR*, *CFI*, *TLI*, *RMSEA* values were respectively: Direct GFP = .069, .83, .71, .14; Orthogonal Stability and Plasticity = .143, .78, .63, .15; Uncorrelated Big Five = .234, .11, -.33, .29.

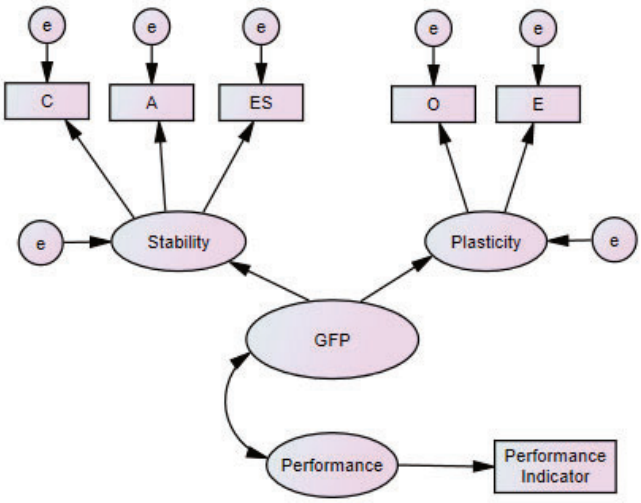

Figure 1. Hierarchical model for the relationship between the GFP and performance criteria.

The mean *SRMR* value (averaged over each of the performance criteria) for the hierarchical model was .045, the mean *CFI*, *TLI* and *RMSEA* values were .93, .88 and .09 respectively, indicating that the hierarchical model described the data rather well.[2] More specifically, Table 1 shows that the hierarchical model showed good fit for overall job performance, objective performance, team performance, training performance, the composite OCB and individual OCB measures, leadership effectiveness, transformational leadership, and for police and semi-skilled/skilled jobs. It did not show good fit for the following performance criteria: training performance, organizational OCB, counterproductive work behavior (composite and interpersonal deviance), leadership, leader emergence and professionals. Note that in those cases, the indices did not reach the thresholds for good fit, but the hierarchical model still had a much better fit relative to the alternative models. The model for academic performance showed lower fit as well. Mediocre fit was found for the OCB Change and organizational deviance as well as the sales and management criteria.

[2] Values were nearly identical when using higher level performance criteria only (i.e. excluding the specific jobs and the lower level CWB and OCB variables). For the five specific job types, the mean values were .042, .93, .87 and .09, respectively.

Table 1. Model fit indices of hierarchical model.

Indicator	χ^2	*df*	*SRMR*	*CFI*	*TLI*	*RMSEA*	90% *CI*
Basic model (without performance criteria)[a]	2818.5	4	.020	.98	.94	.071	-
Performance criteria							
Overall job performance	949.87	8	.029	.96	.92	.071	.067 - .075
Supervisor rated performance	1031.77	8	.036	.94	.89	.083	.079 - .087
Objective performance	173.00	8	.030	.96	.92	.068	.060 - .078
Team performance	88.39	8	.036	.96	.92	.074	.061 - .089
Training performance	311.78	8	.059	.91	.83	.109	.099 - .120
Contextual performance							
OCB - Composite	240.44	8	.028	.96	.93	.066	.059 - .073
OCB - Organizational	94.20	8	.042	.93	.87	.091	.075 - .108
OCB - Individual	95.34	8	.032	.96	.92	.073	.060 - .087
OCB - Change	68.51	8	.043	.94	.89	.081	.064 - .100
CWB[b]	293.50	9	.077	.88	.79	.125	.113 - .137
Interpersonal deviance[b]	481.44	9	.085	.82	.71	.149	.138 - .161
Organizational deviance[b]	156.34	9	.061	.92	.87	.096	.083 - .110
Leadership criteria							
Leadership	524.25	8	.049	.93	.87	.095	.088 - .101
Leader emergence[c]	29.89	8	.057	.92	.86	.100	.063 - .139
Leadership effectiveness[c]	4.38	8	.033	1.00	1.00	.000	.000 - .069
Transformational leadership	109.36	8	.031	.97	.94	.061	.051 - .072
Specific job types							
Sales	196.07	8	.047	.91	.83	.104	.092 - .117
Management	464.15	8	.034	.94	.90	.081	.075 - .087
Police	73.62	8	.032	.96	.92	.070	.056 - .085
Professional	56.36	8	.066	.89	.79	.113	.086 - .141
Skilled/semi-skilled	238.98	8	.032	.96	.92	.069	.062 - .077
Academic performance	4670.80	8	.044	.91	.84	.100	.097 - .102

[a] Values as reported in Van der Linden et al. (2010a).

[b] The error variance of the Stability factor showed a small negative value. The error variance was fixed at zero, resulting in a gain of a single degree of freedom.

[c] For leader emergence and leadership effectiveness, only the mean sample sizes were reported by Judge et al. (2002). The total mean sample size including the Big Five intercorrelations was used as *N* for the CFAs.

Van der Linden et al. (2010a) showed that for the Big Five intercorrelations, the hierarchical model showed better fit when the uncorrected correlations instead of when the corrected correlations were used. Predictably, this was also the case in the current study for the hierarchical models including the performance criteria, as indicated by the mean fit index values of .028, .96, .93 and .05 for *SRMR*, *CFI*, *TLI* and *RMSEA* respectively.[3] The same general patterns in terms of fit were found for the uncorrected correlations. These findings are consistent with Michel, Viswesvaran, and Thomas (2011), who compared SEMs on observed and corrected meta-analytic correlation matrices and found that although the substantive conclusions remained the same, models based on the observed correlations showed better fit.

GFP and Performance Criteria Relationships

Table 2 shows the corrected as well as the uncorrected correlations between the GFP and the performance criteria in the hierarchical model. In this model, the absolute corrected correlations between the GFP and the criteria ranged from .13 (professionals) to .49 (leader emergence), with a mean of .34. For comparison, the mean absolute corrected correlations between Conscientiousness and Emotional Stability, the dimensions marked as the most important personality predictors of job performance (Barrick et al., 2001), and the criteria were .26 (range .12 - .42) and .15 (.02 - .26), respectively.

The corrected correlations for the separate job types were somewhat lower, ranging from .13 (professionals) to .31 (management), with a mean of .24. The relation between the GFP and academic performance was among the lowest of the associations under investigation ($r_{corrected}$ = .16).

The uncorrected absolute correlations between the GFP and performance ranged from .07 (professionals) to .46 (counterproductive work behavior), with a mean of .23. These correlations remain substantial, especially when compared again with the mean absolute uncorrected correlations of Conscientiousness (.15, ranging from .08 to .34) and Emotional Stability (.10, ranging from .01 to .22) with performance.

Controlling for Social Desirability

Li and Bagger (2006) reported correlations with job performance of .10 and .12 for self-deceptive enhancement (SDE) and impression management (IM), respectively. We found that in the hierarchical model, the correlations between the GFP and SDE and IM were .66 and .55, respectively. These correlations and the correlation between the GFP and overall job performance (r = .31) were used to calculate partial correlations between the GFP and overall job performance while controlling for IM and SDE. Controlling for SDE, the partial correlation between the GFP and

[3] Excluding leader emergence and leader effectiveness, for which Judge et al. (2002) did not provide uncorrected correlations.

performance became .33. When IM was controlled for, the partial correlation was .29, which is a negligible attenuation. Thus, although the direct correlations between the GFP and the social desirability measures were relatively high, they did not affect the validity of the GFP in terms of job performance.

Table 2. Correlations between the GFP and criteria in the hierarchical model.

Indicator	Corrected *r*	Uncorrected *r*
Performance criteria		
Overall job performance	.31	.16
Supervisor rated performance	.33	.19
Objective performance	.28	.14
Team performance	.44	.28
Training performance	.47	.23
Contextual performance		
OCB - Composite	.30	.23
OCB - Organizational	.26	.20
OCB - Individual	.34	.26
OCB - Change	.18	.14
CWB	-.47	-.46
Interpersonal deviance	-.40	-.37
Organizational deviance	-.48	-.44
Leadership criteria		
Leadership	.46	.31
Leader emergence[a]	.49	-
Leadership effectiveness[a]	.40	-
Transformational leadership	.32	.30
Specific job types		
Sales	.20	.12
Management	.31	.17
Police	.29	.17
Professional	.13	.07
Skilled/semi-skilled	.25	.15
Academic performance	.16	.17

[a] No uncorrected correlations provided by Judge et al. (2002).

Regression Analyses

As noted before, for a higher order construct to be a useful contribution to the literature in addition to its indicators, its relative importance in terms of criterion validity should be demonstrated (Johnson et al., 2011). In order to assess the relative importance of the combined unique variance of traits at a lower hierarchical level, i.e. the Big Five or Stability and Plasticity, above and beyond the GFP in predicting job performance, regression analyses were conducted for each of the criteria. We adopted the same strategy as Van der Linden et al. (2010a); We first examined how the relationship between the lower-order personality traits (i.e., Big Five or Stability/Plasticity) and performance changed after controlling for the GFP (see also the *Discussion* for a justification of this procedure). Second, we conducted hierarchical regression analyses in which the GFP was entered in the first step, and either all the Big Five, or Stability/Plasticity were entered in the second step. By examining the variance that is explained by the predictors in both steps, we can assess the predictive power of the lower-order personality traits *over and beyond* the GFP.

Table 3 shows the corrected zero-order and partial correlations between each of the Big Five dimensions and the performance indicators. We found that the personality-performance correlations were considerably attenuated and sometimes even reversed sign when the GFP was controlled for. The mean partial correlations were .03, .05, .04, -.07 and -.04 for Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability, respectively. These changes in correlations were substantial. For Conscientiousness, for example, the mean correlation with the performance indicators was reduced by 69% (from .16 to .05). We conducted similar analyses for Stability and Plasticity (see Table 3). The pattern of results in these analyses was similar as for the Big Five in the sense that in the vast majority of cases, the correlations were strongly reduced or even reversed sign.

Table 4 shows the R^2 values of our hierarchical regression analyses. On average, the GFP explained 11% of the total variance in the performance indicators. Adding the Big Five dimensions in the second step increased the explained variance, on average, to 16%. This indicates that about two-third of the explained variance in the performance indicators was attributable to the GFP. Although the individual Big Five dimensions added unique variance to the prediction of job performance and other work-related outcomes, the GFP appeared to be relatively important. In contrast, compared to the unique variance of the Big Five, the GFP accounted for a smaller part of the total explained variance in academic performance ($R^2_{GFP} = 2.6\%$ and $R^2_{GFP+B5} = 8.3\%$). The analyses involving the two meta-factors showed that, on average, the total amount of explained variance in the criteria was 18.5%, of which 68% could be attributed to the GFP and 32% to the unique variance (above and beyond the GFP) of Stability and Plasticity (see Table 4).

Table 3. Relationships between the Big Five dimensions and performance indicators after controlling for the GFP.

Indicator	O	C	E	A	ES	Stability	Plasticity
Performance criteria							
Overall job performance	-.05 (.07)	.12 (.27)	-.01 (.15)	-.04 (.13)	-.06 (.13)	.06 (.30)	-.05 (.16)
Supervisor rated performance	-.06 (.07)	.16 (.31)	-.04 (.13)	-.06 (.13)	-.08 (.13)	.11 (.34)	-.08 (.15)
Objective performance	-.08 (.03)	.09 (.23)	-.01 (.13)	.03 (.17)	-.08 (.10)	.05 (.27)	-.08 (.12)
Team performance	-.01 (.33)	.00 (.27)	-.09 (.27)	.13 (.14)	-.06 (.09)	.12 (.44)	-.06 (.24)
Training performance	.21 (.16)	.04 (.27)	.09 (.27)	-.11 (.34)	-.21 (.22)	-.32 (.29)	.24 (.46)
Contextual performance							
OCB - Composite	.06 (.17)	.05 (.22)	-.06 (.11)	.00 (.17)	-.04 (.15)	.05 (.29)	.03 (.21)
OCB - Organizational	.10 (.19)	.05 (.20)	-.15 (.02)	.06 (.19)	-.05 (.12)	.09 (.27)	-.01 (.16)
OCB - Individual	.08 (.20)	.05 (.25)	-.09 (.11)	.01 (.20)	-.05 (.17)	-.08 (.27)	.16 (.33)
OCB - Change	.12 (.17)	.02 (.12)	.08 (.15)	-.15 (-.03)	-.01 (.09)	-.11 (.11)	.17 (.24)
CWB - Composite	.14 (-.08)	-.06 (-.35)	.32 (-.03)	-.23 (-.44)	.06 (-.26)	-.35 (-.56)	.34 (-.07)
Interpersonal deviance	.09 (-.09)	.06 (-.23)	.33 (.02)	-.29 (-.46)	.04 (-.24)	-.42 (-.53)	.39 (.02)
Organizational deviance	.18 (-.04)	-.18 (-.42)	.22 (-.09)	-.07 (-.32)	.09 (-.23)	-.23 (-.52)	.32 (-.09)
Leadership criteria							
Leadership	.09 (.24)	.03 (.28)	.11 (.31)	-.21 (.08)	-.03 (.24)	-.20 (.33)	.18 (.42)
Leader emergence	.08 (.24)	.08 (.33)	.12 (.33)	-.27 (.05)	-.04 (.24)	-.17 (.37)	.17 (.43)
Leadership effectiveness	.11 (.24)	-.10 (.16)	.05 (.24)	.00 (.21)	-.01 (.22)	-.13 (.30)	.16 (.37)
Transformational leadership	.04 (.15)	-.08 (.13)	.10 (.24)	-.04 (.14)	-.02 (.17)	-.15 (.22	.12 (.29)

(continued)

Table 3 (continued).

Indicator	O	C	E	A	ES	Stability	Plasticity
Specific job types							
Sales	-.11 (-.03)	.19 (.25)	.02 (.11)	-.11 (.01)	-.08 (.05)	.12 (.23)	-.02 (.11)
Management	-.01 (.10)	.10 (.25)	.08 (.21)	-.07 (.10)	-.11 (.09)	-.04 (.26)	.03 (.22)
Police	-.08 (-.11)	.13 (.24)	-.03 (-.11)	-.03 (.06)	-.06 (.06)	.07 (.29)	-.09 (.12)
Professional	-.17 (.03)	.22 (.26)	-.21 (.12)	-.01 (.13)	-.01 (.12)	.25 (.23)	-.33 (-.17)
Skilled/semi-skilled	-.05 (.05)	.11 (.23)	-.08 (.06)	-.04 (.10)	.01 (.15)	.11 (.27)	-.11 (.08)
Academic performance	.06 (.12)	.16 (.22)	-.12 (-.01)	-.03 (.07)	-.11 (.02)	.13 (.20)	.02 (.12)

Note. Zero order correlation in brackets.

Interestingly, considerable differences were found in the relative importance of the GFP compared to the Big Five between the different performance indicators. The lowest effect of the GFP on the total explained variance was found for performance of professionals, in which the unique characteristics of the Big Five (above and beyond the GFP) accounted for a substantially larger share of explained variance than the GFP (R^2_{GFP} = 1.2% and R^2_{GFP+B5} = 10.8%). For the composite score of OCB, the relative contribution of the GFP was the strongest as it accounted for almost all explained variance whereas the unique contribution of the Big Five was small (R^2_{GFP} = 8.4% and R^2_{GFP+B5} = 8.6%).

Discussion

Using meta-analytic data, the present study clearly shows that the GFP, extracted from the Big Five dimensions, is associated with a broad range of job performance measures. This is true for supervisor-rated performance as well as for objective performance indicators. The GFP did not only positively relate to overall performance indicators, but also to many of the specific performance outcomes such as team performance, training performance, and OCB. The GFP was negatively related to counterproductive work behavior, although the models including these outcomes showed somewhat lower fit. Noteworthy is that the GFP was relatively strongly related to leadership outcomes – the highest correlation was found for leader emergence (r = .49) which relates to factors associated with being perceived as leader-like. This is fully in line with previous theorizing on the GFP, stating that individuals who are socially effective and tend to display socially desirable behavior are chosen more often as leaders by others (Figueredo & Rushton, 2009; Rushton & Irwing, 2011; Van der Linden et al., 2014b). Our findings are also in line with previous research showing a positive relation between social effectiveness and managerial performance (Semadar et al., 2006).

The correlation between the GFP and teamwork was also quite high ($r_{corrected}$ = .44). Given the social aspect of this specific type of performance, this is in accordance with the interpretation of the GFP as a factor that fosters social effectiveness. Noteworthy is also the relatively high association with training performance ($r_{corrected}$ = .47). Trainings often include social-evaluative settings (Uziel, 2010a), in which interpersonal competences can be expected to play an important role (Viswesvaran, Ones, & Hough, 2001), which would explain the high association with the GFP.

Table 4. Summary of hierarchical regression analyses for variables predicting criteria.

Indicator	Step 1	Step 2		Step 1	Step 2	
	R^2_{GFP}	R^2_{GFP+B5}	ΔR^2	R^2_{GFP}	$R^2_{GFP+Stability/Plasticity}$	ΔR^2
Performance criteria						
Overall job performance	.078	.102	.024	.096	.102	.006
Supervisor rated performance	.090	.129	.039	.109	.126	.017
Objective performance	.063	.084	.022	.078	.087	.009
Team performance	.176	.199	.023	.194	.208	.014
Training performance	.144	.226	.082	.221	.347	.126
Contextual performance						
OCB - Composite	.084	.086	.002	.090	.093	.003
OCB - Organizational	.063	.107	.045	.068	.075	.007
OCB - Individual	.109	.130	.021	.116	.143	.027
OCB - Change	.026	.064	.038	.032	.072	.040
CWB	.230	.345	.115	.221	.413	.192
Interpersonal deviance	.176	.335	.159	.160	.443	.283
Organizational deviance	.212	.291	.079	.230	.355	.125
Leadership criteria						
Leadership	.168	.219	.051	.212	.270	.058
Leader emergence	.185	.268	.083	.240	.285	.045
Leadership effectiveness	.130	.148	.018	.160	.198	.038
Transformational leadership	.084	.098	.014	.102	.136	.034
Specific job types						
Sales	.029	.101	.072	.040	.054	.014
Management	.073	.109	.036	.096	.098	.002
Police	.068	.094	.026	.084	.097	.013
Professional	.012	.108	.096	.017	.193	.176
Skilled/semi-skilled	.053	.073	.020	.063	.085	.022
Academic performance	.026	.083	.057	.026	.042	.016

Note. GFP = General Factor of Personality; B5 = Big Five personality traits.

With regard to specific job types, the GFP showed the weakest mean association with performance in professional jobs. In fact, the $r = .13$ correlation was the weakest correlation in the study. The professionals in the current study consisted of engineers, architects, attorneys, accountants, teachers, doctors and ministers (Barrick & Mount, 1991). Some of those jobs are characterized by employees who have a rather specific set of abilities and knowledge. Thus, one could argue that, although the social component will also play a role in some of those jobs (e.g. teachers and doctors), it may have a weaker influence in others (e.g. engineers, architects, accountants). Subsequently, the presumed social effectiveness indicative of the GFP may also have less of an influence on performance in professional jobs on the whole, especially when compared with other types of job categories such as sales or management (e.g. Joseph & Newman, 2010; Mount et al., 1998). The GFP-academic performance was found to be relatively small. Getting good grades is mostly dependent on knowledge and cognitive skills. Naturally, motivational and dispositional factors (i.e. personality) will also play a role, but academic achievement is known to be fairly strongly influenced by cognitive capacity and less so by personality traits: Poropat (2009) reported a correlation of .25 between intelligence and academic performance, higher than any of the correlations between the Big Five and academic performance. Others have reported a correlation as high as .56 between intelligence and academic performance (see for example Strenze, 2007).

We found that the average correlation (across all performance indicators) between the GFP and performance was higher than the average correlation between each of the individual Big Five dimensions and performance (e.g. .34 for the GFP and .26 for Conscientiousness). Moreover, after controlling for the GFP, the associations between the individual Big Five dimensions and overall job performance were strongly diminished and sometimes even became negative, suggesting that the shared variance component in the Big Five (i.e., the GFP) was largely responsible for the personality-performance link. From the perspective of the GFP, testing the contribution of the unique variance of the Big Five beyond the shared component makes sense because the GFP is assumed to be present in each of the individual dimensions. This can be compared to the *g*-factor of general intelligence in the cognitive domain. This *g*-factor is generally assumed to be present in each of the specific cognitive abilities such as verbal, numerical, or spatial ability (Jensen, 1998). Thus, any measure of cognitive ability partly reflects the *g*-factor and partly reflects the unique variance of the specific ability. Similarly, from a GFP-perspective, any specific personality measure such as extraversion partly reflects variance that can be attributed to the general personality factor and partly reflects variance that is unique to extraversion.

Note that because of this, testing whether the GFP contributes *beyond* the Big Five would make less sense. By first controlling for the Big Five, one would then already control for the true variance of the general factor that is present in the individual dimensions. This can also be compared to the cognitive domain: it would generally not be considered useful to test how the *g*-factor contributes beyond the specific cognitive ability tests (e.g., verbal, numerical) from which it is extracted.

It should be noted here that the reversed sign of the relations between the Big Five and the outcomes when controlling for the GFP is well accounted for by the social effectiveness interpretation. For example, once the social effective component (i.e. the GFP) is controlled for, high Extraversion is negatively related to team performance. Perhaps once social effectiveness is taken out of the equation, extraverts can be *too* dominant in teams or groups, trying to outvoice others, without listening to other team members or taking input by others into account, resulting in lower team performance. However, such explanations are rather tentative at this point, and warrant further investigation in future studies.

In general, the present patterns of findings contribute to new insights into the nature and validity of the GFP. First, it provides support for the GFP as a substantive factor with a broad influence on behavior. Such findings are in line with previous studies suggesting that the GFP represents general social effectiveness (Dunkel & Van der Linden, 2014; Loehlin, 2012; Van der Linden et al., 2014a). Being socially effective implies that one behaves in such a way that the odds of reaching desired social goals are maximized. Thus, in a work context this may imply that one is able to establish relationships with colleagues, supervisors, or customers that contribute to getting good performance appraisals, making promotions, or simply selling more products (in the case of sales jobs). Subsequently, high GFP individuals, on average, would obtain higher performance outcomes on a wide range of indicators.

It is imperative to note that the present findings are *not* in accordance with the GFP being wholly artefactual, i.e. merely reflecting response bias or a statistical by-product. This is corroborated by the finding that, despite relatively strong correlations between indices of social desirability and the GFP, controlling for these measures of social desirability did not attenuate the GFP-performance relationship – echoing findings on the lower order Big Five (Barrick & Mount, 1996; Ones et al., 1996). In addition, the majority of the performance variables in the present study were based on supervisor ratings or objective outcomes, thus showing that the associations we found cannot be attributed to common-method bias.

A construct that only emerges due to how participants respond to questionnaires (e.g., acquiescence, inflated self) but has little relation to how they would genuinely behave cannot be expected to have such a clear and strong association with job performance. Moreover, it is unlikely that the GFP merely reflects how individuals can present themselves (impression management) in short-term and high-stake situations. In contrast, indicators of job performance often reflect the behavior of individuals over an extended period of time (e.g., several months to years).

If the GFP indeed can be considered a meaningful and stable personality trait, then this has implications for employee selection and HR-practices. Although it is relatively common practice to include a Big Five measure or some other sort of personality questionnaire in the selection process, interpretations of scores are confined to the facet or factor levels. Based on our findings that the GFP relates to such a wide range of work-related outcomes, it might be advisable to calculate a GFP-score based on the Big Five dimensions. This of course will not be much of an effort, given that the Big Five scores are readily available. Naturally, for some jobs – i.e., those in which emotional

competencies or interpersonal skills play a larger role – this might make more sense than for other jobs; when hiring a computer programmer, obtaining applicants' GFP-scores may be less valuable than when hiring a customer sales agent. Yet, given that in virtually all jobs people will have to deal with co-workers or other people, having a measure of social effectiveness would never be superfluous. Skeptics who are afraid that assessing the GFP in selection procedures will not be useful because it will be inflated due to 'faking' may be comforted by the finding that the GFPs extracted from selection and assessment samples are highly similar (Van der Linden et al., 2011).

Limitations

A limitation that should be taken into account when interpreting the results relates to the combination of meta-analytic correlations from different sources. Although combining estimates from several meta-analyses in order to test structural models is becoming a more common method in the literature (e.g., Chang et al., 2012; Connelly & Chang, 2016; Heller, Watson, & Ilies, 2004; Marcus, Taylor, Hastings, Sturm, & Wiegelt, 2013; Van der Linden et al., 2017), some authors have called for caution in its use because of the risk of second-order sampling error (Cheung & Chan, 2005; Landis, 2013). Differences in the meta-analytic procedures may introduce additional uncontrolled error variance in the estimation of structural models: consequently, this argument is less pertinent if the different meta-analyses from which the correlations are drawn largely adopted the same procedures and methods in arriving at their meta-analytic estimates (Heller et al., 2004).

We believe there to be relatively little cause for concern in the present study. In terms of inclusion rules and classification of personality measures in Big Five factors, the coding scheme by Barrick and Mount (1991) was largely used or similar procedures were reported. All studies only included measures of the Big Five factors, not facets. In addition, all meta-analyses exclusively included independent samples in their study. As noted earlier, outcomes were based on other-ratings or objective measures, thus limiting the influence of common-method bias. And finally, in terms of computations, all meta-analyses used psychometric meta-analytic procedures as outlined by Hunter and Schmidt (2004). Thus, the meta-analyses combined in the present study were highly similar in terms of their procedures hereby largely reducing the influence of respective differences on our results.

Strength of our chosen procedure is that it is in line with the recommendations by Landis (2013), who provided guidelines on successfully combining meta-analyses and structural models. As he advised, we used meta-analytic correlations reported in other sources in order to fill blank cells in the correlation matrix in order to test relations not tested in the primary studies (i.e., GFP-outcome associations). Yet, Landis (2013) does argue for caution in drawing causal inferences from studies combining meta-analyses and SEM. However, as he points out, this argument is not unique to the current study but pertains to the psychological literature at large and thus also to the primary meta-analyses combined in the current study.

Concluding Remarks

The present study contributes to research on the GFP. In the literature there has been a debate about the validity and the practical relevance of this construct (Chang et al., 2012; Ferguson et al., 2011; Irwing, 2013). The present study supports the practical relevance of the GFP as it apparently can be used to predict job performance. Theoretically, the GFP reflects a substantive factor with meaningful relations to many work-related outcomes. Such a conclusion is in line with previous studies such as the study by Ones et al. (1996) who also concluded that the shared variance among personality traits is substantive and that it may be considered a 'red herring' to label this shared variance as social desirability bias.

When interpreting the present findings two considerations need to be taken into account. First, even though the study provides useful information regarding the relationship between the GFP and job performance, additional research is necessary to further delineate the nature of the construct. For example, it would be useful to conduct micro-level (e.g., diary, observational) studies in which it is directly tested whether the GFP-performance associations may be mediated by socially effective behavior. In line with this, it would be useful to further examine what types of (work) behavior are specific for high-GFP individuals. In addition, it would be fruitful to further investigate the relationship between the GFP and different job types, and mostly in terms of their levels of emotional labor (Joseph & Newman, 2010). Some jobs require more emotional labor in terms of interpersonal relations, e.g. dealing with customers. We can expect social effectiveness to come to the fore in jobs in which interpersonal relations lie at the core of the job resulting in a larger effect of the GFP on performance. Although the present study provides some evidence for this, more research is needed to draw a more solid conclusion on this issue.

Second, the fact that their shared variance is a good predictor of performance does not imply lower-order personality factors to become obsolete. For example, in the general population, GFP scores may be a good predictor of performance, but for individuals who score in a similar range on the GFP, specific patterns of traits may become more important. Related to this, very specific forms of job performance may be better predicted by lower level traits that are more aligned with this specific type of performance than the GFP. Criterion-related validity can often be improved when the level of personality measurement is aligned with that of the criterion (Jenkins & Griffith, 2004). Sitser et al. (2013), for example, showed that handling customer complaints was predicted by Agreeableness and the Consideration facet of this trait, but not by the GFP. Thus, for some narrow performance measures, a more narrow scope might be needed. Future studies may want to focus on more specific performance aspects and how they relate to broad or narrower personality measures. Regardless, the present study showed that the GFP is a good predictor when it comes to broad measures of performance in a wide range of different occupations. Such information may be considered a relevant piece of the puzzle in delineating the nature of this general factor.

Appendix

In the current study, the meta-analysis on the Big Five intercorrelations from Van der Linden et al. (2010a) and the meta-analysis on the Big Five-performance relations from Barrick et al. (2001) were used because they are the most comprehensive studies in the field. Over the years, several other (smaller) meta-analyses have been published on the relations under scrutiny in the current study. To investigate whether the choice of the meta-analytic estimates used as input for our analyses affected the results, parallel analyses were conducted using correlations from different meta-analyses. First, we replaced the Big Five-performance relations from 2001 by those from the Judge et al. (2013) meta-analysis (*k*s between 40 and 74, *N*s between 14,321 and 41,939). Using overall job performance as the criterion, and fitting the hierarchical model to the corrected correlation matrix, resulted in a model with nearly identical model fit (*CFI*: .95, *TLI*: .91, *RMSEA*: .076, *SRMR* .031 compared to .96, .92, .071 and .029 respectively) and GFP-performance estimate (r = .34, compared to .31, Table 2).

Additionally, we ran parallel analyses by replacing the Big Five intercorrelations by the meta-analytic estimates from Mount, Barrick, Scullen, and Rounds (2005). These values have often been used to fill in the empty cells of meta-analytic correlation matrices in order to test new path models (e.g. Chang et al., 2012; Connelly & Hülsheger, 2012). Compared to the meta-analysis by Van der Linden et al. (2010a), this was a relatively small meta-analysis (based on only four studies with a total *N* of 4000). Still, the fit of the model remained unchanged (*CFI*: .95, *TLI*: .90, *RMSEA*: .079, *SRMR* .037), just as the GFP-performance estimate (r = .35).

To further support the robustness of the GFP-performance relation in terms of the used meta-analytical estimates, we adopted the most recent meta-analytic estimates of the Big Five intercorrelations (*k*s between 39 and 89, *N*s between 9,886 and 18,405) from Davies et al. (2015). They made a distinction between within and between-inventory relations. Parallel analysis on the between-inventory relations showed worse fit for the hierarchical model (*CFI*: .88, *TLI*: .77, *RMSEA*: .086, *SRMR* .041), but again a highly similar GFP-performance relation (r = .38). Thus, the choice of the meta-analytic correlations did not appear to affect our results.

Supplemental material

Table A. Big Five intercorrelations, correlations with criteria and social desirability measures.

	Corrected correlations					Uncorrected correlations				
Variable	O	C	E	A	ES	O	C	E	A	ES
Big Five intercorrelations[a]										
Openness	1					1				
Conscientiousness	.20	1				.14	1			
Extraversion	.43	.29	1			.31	.21	1		
Agreeableness	.21	.43	.26	1		.14	.31	.18	1	
Emotional Stability	.17	.43	.36	.36	1	.12	.32	.26	.26	1
Performance criteria[b]										
Overall job performance	.07	.27	.15	.13	.13	.03	.12	.06	.06	.06
Supervisor rated performance	.07	.31	.13	.13	.13	.03	.15	.07	.06	.07
Objective performance	.03	.23	.13	.17	.10	.02	.10	.06	.07	.05
Training performance	.33	.27	.28	.14	.09	.14	.13	.13	.07	.05
Team performance	.16	.27	.16	.34	.22	.08	.15	.08	.17	.13
Contextual performance[c]										
OCB - Composite	.17	.22	.11	.17	.15	.11	.14	.07	.11	.10
OCB - Organizational	.19	.20	.02	.19	.12	.13	.13	.01	.12	.08
OCB - Individual	.20	.25	.11	.20	.17	.13	.16	.07	.13	.11
OCB - Change	.17	.12	.15	-.03	.09	.11	.08	.10	-.02	.06
CWB - Composite	-.08	-.35	-.03	-.44	-.26	-.06	-.30	-.03	-.35	-.22
CWB - individual deviance	-.09	-.23	.02	-.46	-.24	-.07	-.19	.02	-.36	-.20
CWB - organizational deviance	-.04	-.42	-.09	-.32	-.23	-.03	-.34	-.07	-.25	-.19

(continued)

Table A (continued).

Variable	Corrected correlations					Uncorrected correlations				
	O	C	E	A	ES	O	C	E	A	ES
Leadership criteria[d]										
Leadership	.24	.28	.31	.08	.24	.16	.20	.22	.06	.17
Leader emergence[†]	.24	.33	.33	.05	.24	-	-	-	-	-
Leadership effectiveness[†]	.24	.16	.24	.21	.22	-	-	-	-	-
Transformational leadership	.15	.13	.24	.14	.17	.11	.10	.19	.10	.15
Specific job types[b]										
Sales	-.03	.25	.11	.01	.05	-.01	.11	.07	.01	.03
Management	.10	.25	.21	.10	.09	.05	.12	.10	.04	.05
Professional	-.11	.24	-.11	.06	.06	-.05	.11	-.05	.03	.04
Police	.03	.26	.12	.13	.12	.02	.13	.06	.06	.07
Skilled/semi-skilled	.05	.23	.06	.10	.15	.03	.12	.03	.05	.08

(continued)

Table A (continued).

Variable	Corrected correlations					Uncorrected correlations				
	O	C	E	A	ES	O	C	E	A	ES
Social desirability[e]										
Self-deceptive enhancement	.19	.42	.31	.19	.54	.14	.32	.23	.14	.41
Impression management	.10	.42	.03	.42	.35	.07	.33	.02	.33	.27
Academic performance[f]	.12	.22	-.01	.07	.02	.10	.19	-.01	.07	.01

[a] From Van der Linden et al. (2010a); *k* = 212, *N*s between 39 and 21,105.

[b] From Barrick et al. (2001); *k*s in relation to Big Five and Overall job performance ranged between 143 and 239, *N*s between 23,225 and 48,100), Supervisor (*k*s: 116-185, *N*s: 18,535-33,312), Objective (*k*s: 25-37, *N*s: 4,401-7,101), Training (*k*s: 18-25, *N*s: 3,177-4,100), Team (*k*s: 10-48, *N*s: 1,820-3,719).
Sales (*k*s: 17-36, *N*s: 2,168-4,141), Management (*k*s: 44-67, *N*s: 8,678-12,602), Professional (*k*s: 4-10, *N*s: 476-965), Police (*k*s: 16-22, *N*s: 1,688-2,369), Skilled/semi-skilled (*k*s: 32-47, *N*s: 6,055-7,682).

[c] OCB criteria from Chiaburu et al. (2011); OCB - Composite (*k*s: 36-71, *N*s: 6,700-14,355), OCB-O (*k*s: 7-20, *N*s: 1,311-4,598), OCB-I (*k*s: 10-28, *N*s: 2,049-6,347), OCB-CH (*k*s: 6-19, *N*s: 1,144-3,761)
CWB criteria from Berry, Ones and Sackett (2007); CWB-ID (*k*s: 8-11, *N*s: 2,360-3,458), CWB-OD (*k*s: 5-8, *N*s: 1,772-2,934).

[d] Leadership (*k*s: 35-60, *N*s: 7,221-11,705), leader emergence (*k*s: 17-37) and leadership effectiveness (*k*s: 17-23) from Judge et al. (2002), transformational leadership (*k*s: 18-20, *N*s: 3,338-3,916) from Bono and Judge (2004).

[e] From Li and Bagger (2006); SDE (*k*s: 16-26, *N*s 2,881-4,361), IM (*k*s: 18-27, *N*s 3,223-4,978).

[f] From Poropat (2009); *k*s ranging between 109-138, *N*s between 58,522-70,926.

[†] No uncorrected correlations were provided in Judge et al. (2002).

The General Factor of Personality and Daily Social Experiences: Evidence for the Social Effectiveness Hypothesis

3

Abstract

Using data from the Berlin diary study (N = 1,223), the associations between the General Factor of Personality (GFP) and daily social experiences, self-esteem and mood (positive affect and negative affect) were examined. As predicted, high-GFP (vs. low) individuals reported fewer interpersonal conflicts, higher levels of relationship quality, and better impressions made on others. Also in line with our expectations, the relations between the GFP and negative affect, positive affect and self-esteem were significantly mediated by relationship quality and daily impressions. Furthermore, multilevel analyses showed that the GFP significantly moderated the associations between daily social experiences and daily levels of self-esteem and negative affect so that the effects were stronger when GFP-levels were low. In sum, the study results were in line with the notion of the GFP as a social effectiveness factor, with important consequences for people's daily social life and well-being.

Introduction

In the personality literature, there is an increasing number of studies on the General Factor of Personality (GFP; Figueredo et al., 2004) which emerges due to the intercorrelations among more specific personality dimensions, such as the well-known Big Five, or Giant Three models. The GFP constitutes the socially desirable ends of those dimensions. and has now been extensively replicated (e.g., Musek, 2007; Van der Linden et al., 2010a). In terms of the Big Five, high-GFP individuals can, on average, be described as relatively open-minded, diligent, sociable, friendly and stable. Moreover, the GFP has shown criterion-related validity and is associated with various important life outcomes such as job performance and leadership (Pelt, Van der Linden, Dunkel, & Born, 2017).

Despite such consistent findings, however, diverging scientific views on the *interpretation* of the GFP exist. One such view is that the GFP is substantive and mainly represents social effectiveness (see Van der Linden et al. (2016) for an overview). In this view, high-GFP individuals have the knowledge, abilities, and motivation to generally behave in socially desirable ways, which helps them to get ahead and along with others (Rushton & Irwing, 2011). In contrast are the views that the GFP represents not much more than a methodological artefact, due to, for example, socially desirable response bias (Bäckström et al., 2009), common method variance (e.g., Chang et al., 2012) or other statistical artefacts (Ashton et al., 2009; Revelle & Wilt, 2013).

The different arguments for the substantive and artefact views of the GFP have been discussed extensively in several review articles (Ferguson et al., 2011; Irwing, 2013; Revelle & Wilt, 2013; Van der Linden et al., 2016), and will therefore not be repeated here. The main point, however, is that there appears to be evidence for each of the different views. This is not surprising, given that the different explanations of the GFP need not necessarily be mutually exclusive. That is, GFP scores appear to represent a blend of measurement error, and self-evaluative tendencies, but the largest share of variance appears to be attributable to social effectiveness (Davies et al., 2015; Dunkel et al., 2016).

In the present paper, we aim to contribute to the literature on this topic by further testing the nature of the GFP in several ways. First, we test the validity of the GFP as a social effectiveness factor through its relations with daily social experiences. Second, we test to what extent the well-documented relation between the GFP and well-being and mood (Erdle & Rushton, 2011; Musek, 2007) is mediated by the presumed effective daily social experiences. And finally, we test how the GFP moderates the relation between daily social experiences and daily mood. Together, the present study forms a comprehensive diary study on the GFP as representing socially effectiveness.

The GFP and Daily Social Experiences

Research on the Big Five traits already suggests a positive relation between a mix of scores on the socially desirable ends of these traits ("a positive manifold") and social behaviors or outcomes. For example, Neuroticism has been linked to lower levels of relationship satisfaction and relationship

quality (Heller et al., 2004; Henderson, Byrne, & Duncan-Jones, 1981; Holland & Roisman, 2008; Karney & Bradbury, 1995), while the reverse is found for Extraversion (e.g., Asendorpf & Wilpers, 1998). Agreeableness has been linked to less (intense) interpersonal conflicts (Asendorpf & Wilpers, 1998; Graziano, Jensen-Campbell, & Hair, 1996) and Conscientiousness to relationship quality (Jensen-Campbell & Malcolm, 2007), stability and length (Karney & Bradbury, 1995; Shaver & Brennan, 1992). Results for Openness are more mixed, although relations with relationship length and reduced interpersonal problems have been found (Gurtman, 1995; Shaver & Brennan, 1992).

Although the combined evidence on the Big Five dimensions points in the expected direction, the GFP is not simply the sum of its parts (Van der Linden, 2011), emphasizing the need to test the relation between the GFP and social experiences more directly. From a theoretical point of view, if the GFP reflects social effectiveness by means of displaying socially desirable behavior, then it should also relate to daily social experiences and outcomes. That is, high-GFP individuals should, on average, be more effective in their daily social interactions than their low-GFP peers. There is initial evidence for this claim; for example, the GFP has been found to be related to likeability and popularity reported by peers (Van der Linden et al., 2010b), as well as to scores on tests of social knowledge and social ability (Van der Linden et al., 2014a). Presumably, these higher levels of social skill will lead to the enjoyment of better social relations. In addition, compared to individuals with lower GFP scores, people with higher GFP-scores are rated more favorably in selection situations, where leaving a good impression is essential because a job is at stake (Van der Linden et al., 2014b).

Given their ability to better deal with social demands and proclivity to show prosocial behavior, it also seems reasonable to assume that high-GFP individuals are less likely to get in conflict with others. Similarly, we can expect those with higher GFP scores to report higher levels of relationship quality. Indirect evidence for this idea comes from studies on the relationship between emotional intelligence, a construct that strongly overlaps with the GFP (corrected $r = .86$; Van der Linden et al., 2017), and social relationships. Studies have consistently shown that EI is positively related to social outcomes such as maintaining good relations with others and high quality of (daily) interactions with friends (e.g. Lopes et al., 2004; Schutte et al., 2001). In addition, EI-scores are found to be related to leaving a favorable impression on others during a given day (Lopes et al., 2004).

In sum, we can expect that the presumed social effectiveness indicative of the GFP is reflected in daily indicators of one's social life. Being more socially effective, high-GFP individuals should be able to navigate more easily through social encounters on a daily basis and consequently report higher levels of relationship quality and lower levels of interpersonal conflict. Additionally, they should report relatively high ratings of the impressions they make on others, given their tendency for acting in socially desirable ways. Based on the previous discussion, our first hypothesis states:

> *Hypothesis 1. The GFP is negatively associated with (a) interpersonal conflict and positively associated with (b) relationship quality, and (c) the impressions made on others.*

Mediation of the GFP - Well-being/Mood Relation by Daily Experiences

If high-GFP individuals show highly socially effective behavior on a daily basis, then this may partly be the reason why, on average, they also tend to report enhanced self-esteem and mood (e.g., Musek, 2007). On days when people feel socially included, they also tend to report higher levels of well-being than on days when they feel more socially isolated. This is known as the Sociometer theory (Leary, Tambor, Terdal, & Downs, 1995), which links the GFP to both self-esteem and social inclusion (e.g., relationship quality): higher GFP levels may be associated with higher levels of social inclusion, which in turn should result in higher levels of self-esteem and mood.

Previous studies have indeed confirmed that, as humans are social by nature, interpersonal events are among the most potent ones to influence (fluctuations in) self-esteem and mood (e.g., Diener, 1984; Gable, Reis, & Elliot, 2000). Thus, given that daily social events and well-being tend to covary, it is reasonable to assume that at least part of the GFP-well-being relation is mediated by daily interpersonal events.

> *Hypothesis 2. The positive relations between the GFP, and self-esteem, positive affect and negative relation with negative affect are at least partially mediated by (a) less daily interpersonal conflict, (b) better daily relationship quality, and (c) the enhanced daily impressions made on others.*

Daily Social Experiences, and Daily Well-being and Mood: Moderation by the GFP

Even though high-GFP individuals may, on average, experience more positive and fewer negative interpersonal events, it is obvious that sometimes, experiencing a negative social event such as a conflict will be inevitable. Yet, when interpersonal conflicts *do* occur, part of the social effectiveness that is attributed to high-GFP individuals may consist of their ability to adequately deal with or react to such negative interpersonal events (e.g., Dunkel & Van der Linden, 2014; Hengartner et al., 2017). For example, during a conflict, a high GFP level may allow one to better regulate social and emotional behavior and choose the appropriate reaction, thereby providing a solution to the conflict or at least preventing further escalation. Another possibility is that high-GFP individuals are better able to choose the correct coping strategies (e.g., self-regulation tactics; Hengartner et al., 2017; Van der Linden et al., 2017) when faced with negative events such as interpersonal conflicts. Similarly, those with higher GFP scores may also be more likely to choose the appropriate behavioral responses on days when relationship quality is lower (e.g., trying to find a solution rather than doing nothing) to make things better (Dunkel & Van der Linden, 2014). This notion that high-GFP individual may also be more adept to deal with negative social events is in line with the recent meta-analytic finding that the GFP highly overlaps with trait emotional intelligence (Van der Linden et al., 2017). Therefore, it can be expected that at high (versus low) levels of the GFP, daily interpersonal conflicts

and days with lower levels of relationship quality or worse impressions made on others will have less effect on one's daily reported mood and well-being. In sum, it is likely that the GFP will also moderate the relation between daily indicators of social behavior and daily well-being.

In a broader perspective, the more general moderating role of personality traits in the relation between daily events and daily well-being is well-documented. For example, the increased sensitivity to stressors associated with neuroticism (Bolger & Zuckerman, 1995; Zautra, Affleck, Tennen, Reich, & Davis, 2005) is assumed to result in relatively larger dips in mood on days when stressful events occur for neurotic individuals compared to their less neurotic peers (Bolger & Schilling, 1991). Poor coping strategies (Bolger & Zuckerman, 1995; DeLongis & Holtzman, 2005; O'Brien & DeLongis, 1996) and ineffective emotion regulation (Kokkonen & Pulkkinen, 2001) may enhance these effects. Denissen and Penke (2008a) – using largely the same data as the present study – showed, for example, that higher neuroticism was indeed related to increased reactivity to daily interpersonal conflict and daily perceptions of relationship quality in terms of self-esteem. Their findings do not stand in isolation (e.g., Bolger & Schilling, 1991; Bolger & Zuckerman, 1995; Mroczek & Almeida, 2004). Similar moderating effects of personality traits on the relation between both positive and negative events and daily well-being have been found for Extraversion (David, Green, Martin, & Suls, 1997; Larsen & Ketelaar, 1991; Longua, DeHart, Tennen, Armeli, 2009; Nezlek & Allen, 2006), depression (Nezlek & Gable, 2001; Nezlek, Hampton, & Shean, 2000) and the behavioral inhibition system (BIS) and behavioral activation system (BAS) from Gray's (1987) typology (Gable et al., 2000).

Personal dispositions thus appear to influence one's *reactivity* to either positive or negative events, that is, the extent to which events or experiences affect one's daily level of well-being and mood (Bolger & Zuckerman, 1995). For reasons outlined above, a similar role for the GFP can be expected. More generally speaking, because high-GFP individuals are more adapted to their social environment on the basis of their more stable social relationships (see section *The GFP and daily social experiences*) and higher levels of self-esteem (e.g., Musek, 2007), they can be expected to show less fluctuations in mood states caused by specific – even positive – social events. In sum, we expect that for high-GFP individuals the impact of social events and experiences (negative or positive) on daily well-being is weaker compared to low-GFP individuals. This notion is summarized in Hypothesis 3:

> *Hypothesis 3. The relations between (a) daily interpersonal conflict (b) daily levels of relationship quality, (c) daily impressions made on others, and daily levels of self-esteem, positive affect and negative affect are moderated by the GFP such that the relations are stronger for those with lower (compared to higher) GFP scores.*

The Present Study: Using Diary Data

The majority of previous studies on the GFP used cross-sectional designs. Although informative in their own right, such designs are limited because they provide a snapshot of ongoing psychological states and processes. In addition, they rely on people's imperfect ability (and perhaps willingness) to correctly recollect events or behaviors, which can lead to significant biases and inaccuracies. Because of these drawbacks, scholars have argued for the use of diary methods (Bolger et al., 2003) that have the advantage of assessing events and processes as they are naturally occurring, thereby increasing the ecological validity of the results.

Moreover, diary methods may be less susceptible to socially desirable response bias than cross-sectional designs (Barta et al., 2013). This notion is especially important because of the previously discussed interpretation of the GFP as being purely artefactual. Intuitively, it may be equally possible to over-report desirable events or traits on a daily basis as in a single measurement. However, evidence suggests this is not likely to be the case, i.e., daily reports are often found to be more accurate than single, one-time measurements (e.g., Presser & Stinson, 1998). Considering the above, our hypotheses are best tested with daily level data. To this end, we use data from the Berlin Diary Study (https://www.psychologie.hu-berlin.de/de/prof/perdev/downloadentwper/diarystudy) by Denissen and colleagues (2005 – 2008), one of the largest diary studies in the world.

Method

Data files, analysis scripts, and supplemental analyses can be accessed at https://osf.io/kywdf/.

Sample and Procedure

The Berlin Diary Study (2005 – 2008) consisted of multiple phases starting with a general questionnaire, including personality surveys. At the end of this phase, participants were requested to name the friend and family member with whom they had most contact with, and their partner (if present). The second phase was the diary phase. For 30 days, participants filled out a daily questionnaire including randomly presented questions on daily well-being and daily interactions with the two or three identified others in the previous phase. To minimize recall bias, the diaries were only accessible between 9 pm and 4 am. For additional information on the procedures and study design we refer to Denissen and Penke (2008a) and Denissen, Penke, Schmitt, and Van Aken (2008), who used earlier versions of these data.

In the present study, we decided to only include participants who had completed at least 7 diary entries in order to minimize the influence of idiosyncratic days and assure commitment to the study, the latter being known to affect study results (Bolger et al., 2003). Our final sample therefore consisted of 1,223 German participants (1,055 women, 86%), with an average number of 19.28 (*SD*

= 6.81) daily reports. The average age was 29.47 (*SD* = 10.49). Most people were either single (39%) or in a steady relationship without being married or engaged (40%), without children (79% of the total sample). About 50% of the sample was relatively highly educated.

Measures

General questionnaire

Personality/GFP

Two measures of personality were used to extract a GFP. First, the Big Five Inventory (BFI; John & Srivastava, 1999) was used to measure Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N). Sample coefficient alphas ranged from .72 to .90 (see Table 1). The second Big Five measure was the 50-item Five Individual Reaction Norms Inventory (FIRNI; Denissen & Penke, 2008b), which is based on the idea that the five factor model represents stable individual differences in people's motivational reactions under different environmental pressures. Sample reliabilities ranged from .78 to .89 (see Table 1). Both instruments use a 5-point Likert-scale format.

Principal axis factoring was used to extract GFPs from both measures. For the BFI, the first unrotated factor explained 26.10% of the variance in the Big Five, while for the FIRNI this was 20.6%. The GFP factor loadings of O, C, E, A, and N were .36, .42, .66, .47 and -.58 (BFI) and .32, .48, .48, .10, and -.67 (FIRNI), respectively. Notwithstanding the smaller loading of A on the GFP in the FIRNI, these results and the Big Five intercorrelations from Table 1 show that a general factor was present in both personality measures. Scores on the first unrotated factor were saved using the regression method as a measure of the GFP.[4] The GFPs extracted from the two measures (using the regression method) were highly similar (r = .81), and more similar than any of the Big Five; the maximum correlation between the two measures in terms of the Big Five was found for Extraversion (r = .73).

Daily questionnaire

Reliabilities of the daily constructs as reported in Table 1 represent group mean (intercept) reliability estimates, which take the *intraclass correlation* (ICC) and the number of diary entries into account (Bryk and Raudenbush, 1992). All reliabilities were good to excellent.

Daily well-being and mood

Self-esteem. State self-esteem was measured by taking the mean of four items from the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965): previous studies have shown that these four items form a

[4] Confirmatory factor analysis (CFA) also supported the existence of a GFP in both measures. Results on the CFAs and convergence across different GFP extraction methods are reported in Appendix A1).

good and reliable measure of the construct at the daily level (Denissen & Penke, 2008a; Denissen et al., 2008).

Positive affect. Positive affect (PA) was measured by the average score over 9 mood adjectives – active, attentive, determined, enthusiastic, excited, inspired, interested, proud, and strong – from the PANAS (Watson, Clark, & Tellegen, 1988).

Negative affect. Negative affect (NA) was measured by 9 adjectives from the PANAS survey: afraid, ashamed, distressed, guilty, hostile, irritable, jittery, nervous, scared, and upset. A NA-score was calculated by averaging over the scores on these 9 mood adjectives.

Daily social experiences

Relationship quality. Participants were asked to rate (on a five-point scale) nine commonly used facets of interaction quality (Denissen and Penke, 2008a), i.e. their level of enjoyment, interest, intimacy, power, important, calm, safe, wanted and respected in the interactions with the identified persons (i.e. friend, family member and, if present, partner). Following the procedure of Denissen and Penke (2008a), an overall index of relationship quality was created by averaging over all indicators across the two or three identified others.

Interpersonal conflict. Participants were asked whether they experienced (0 = not present, 1 = present) a conflict with the identified others on eight different topics (financial resources, communication problems, activities, life plans, encouragement, opinions, third persons, and other topics). Scores were summed over the topics each day and the three relationship types (i.e., friend, family member and partner): a zero thus indicated no conflict on that particular day. This variable was characterized by a highly skewed distribution (a large number of zero's). Therefore, in addition to the count variable, a dichotomized version with 0 indicating no conflict and 1 indicating any conflict was also created.

Impressions made on others. A subsample of the participants (N = 970) was asked to indicate the impressions they made on others during that day on eight different dimensions (competence, civility, ethical, artistic, sympathetic, orderly, psychical attractiveness, and tolerant). All questions were in 7-point Likert format. A total *impression on others* score (i.e., the average) was calculated.

Statistical Analyses

Given that the relations between the GFP, social outcomes and well-being constituted between-person effects, we aggregated daily reports of these latter two variables across days.

The mediation hypotheses were tested using the PROCESS Macro (Hayes, 2013) for SPSS. Due to the large sample size, we focus on the ratio (i.e., the effect size) of the standardized indirect effect to the total effect, rather than on significance levels.

To test the moderation hypotheses, we used multilevel regression analyses, as the data follow a hierarchical structure because days (Level 1) are nested within individuals (Level 2). Multilevel analysis or hierarchical linear modeling (HLM) provides more accurate parameter

estimates and significance tests than comparable ordinary least squares regression techniques by accounting for variance at each level of analysis. Moreover, HLM can adequately deal with differences in the number of observations (in this study; number of days) per unit of analysis (see Nezlek, 2001). In the present study, the intraclass correlations (ICC; Bryk & Raudenbush, 1992) ranged between .32 and .54, indicating that there was a significant amount of variance at both levels to justify multilevel analyses.

. In each multilevel model, the main effects of the GFP and the daily experience were included, as well as their cross-level interaction. The daily predictors were person-mean centered to prevent individual differences in daily social experiences from influencing parameter estimates (Nezlek, 2001); a participant's coefficient therefore reflects daily fluctuations from his/her average level. All models were fitted using the *nlme* package in R (Pinheiro, Bates, DebRoy, Sarkar, 2016; R Core Team, 2016). A detailed description of the multilevel procedure is included in Appendix A2.

Finally, the relative contribution of the GFP as compared with the Big Five in explaining the outcomes was assessed. When a higher order factor such as the GFP is proposed, it is relevant to also test its relative importance compared to its lower order constituents in terms of predictive power (Johnson et al., 2011). To this end, hierarchical regression analyses were conducted in which the GFP was entered in the first step, and all the Big Five factors simultaneously in the second. The additional amount of explained variance in the second step indicates the predictive power of the Big Five factors over and beyond the GFP.

Results

Descriptive Statistics

Descriptive statistics and reliabilities (on the diagonal) of all variables are presented in Table 1. Participants reported relatively few conflicts, on average only .37 conflict per day. Conflicts and relationship quality, although related (r = -.33), appeared to assess different aspects of interpersonal relationships.

Table 1. Means, standard deviations, reliability coefficients and intercorrelations between variables.

		M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	GFP – BFI	.00	1.00	-													
2.	GFP – FIRNI	.00	1.00	.81	-												
3.	Openness - BFI	3.83	.60	.45	.35	.84											
4.	Conscientiousness - BFI	3.48	.64	.52	.47	.14	.83										
5.	Extraversion - BFI	3.33	.83	.81	.64	.35	.25	.90									
6.	Agreeableness - BFI	3.54	.57	.58	.39	.12	.26	.24	.72								
7.	Neuroticism - BFI	3.17	.77	-.72	-.63	-.13	-.24	-.39	-.33	.85							
8.	Openness - FIRNI	3.79	.60	.33	.41	.60	.16	.22	.08	-.18	.80						
9.	Conscientiousness - FIRNI	3.29	.72	.50	.62	.25	.72	.30	.13	-.33	.27	.89					
10.	Extraversion - FIRNI	3.26	.71	.64	.61	.18	.16	.73	.35	-.34	.09	.18	.86				
11.	Agreeableness - FIRNI	3.63	.55	.18	.13	.06	.16	**.00**	.49	**-.03**	.09	**.02**	.09	.78			
12.	Neuroticism - FIRNI	2.96	.71	-.62	-.86	-.16	-.24	-.46	-.32	.62	-.17	-.31	-.38	**-.03**	.83		
13.	Self-esteem	3.89	.60	.52	.56	.15	.33	.39	.22	-.48	.14	.35	.34	**-.04**	-.51	.93	
14.	Positive affect (PA)	2.87	.53	.43	.41	.25	.26	.37	.14	-.31	.22	.29	.32	**-.01**	-.29	.51	.92
15.	Negative affect (NA)	1.82	.53	-.36	-.39	**-.02**	-.24	-.20	-.23	.40	-.08	-.22	-.18	-.06	.40	-.66	**-.05**
16.	Relationship quality	3.93	.51	.33	.33	.10	.23	.26	.23	-.21	.07	.19	.26	.09	-.28	.45	.34
17.	Interpersonal conflict	.45	.50	-.08	-.08	.05	-.06	**-.01**	-.12	.09	**.03**	**-.04**	**-.03**	**-.03**	.10	-.18	**.02**
18.	Interpersonal conflict (no/yes)	.37	.24	-.08	-.07	.02	-.06	**.01**	-.11	.11	**.00**	**-.05**	**.00**	**-.03**	.08	-.15	**.04**
19.	Impressions made on others	4.68	.67	.45	.38	.26	.32	.34	.23	-.28	.10	.24	.31	**.02**	-.30	.53	.54

(continued)

Table 1 (continued).

		15	16	17	18	19
1.	GFP – BFI					
2.	GFP – FIRNI					
3.	Openness - BFI					
4.	Conscientiousness - BFI					
5.	Extraversion - BFI					
6.	Agreeableness - BFI					
7.	Neuroticism - BFI					
8.	Openness - FIRNI					
9.	Conscientiousness - FIRNI					
10.	Extraversion - FIRNI					
11.	Agreeableness - FIRNI					
12.	Neuroticism - FIRNI					
13.	Self-esteem					
14.	Positive affect (PA)					
15.	Negative affect (NA)	.94				
16.	Relationship quality	-.40	.93			
17.	Interpersonal conflict	.29	-.33	.88		
18.	Interpersonal conflict (no/yes)	.31	-.30	.77	.80	
19.	Impressions made on others	-.30	.48	**-.06**	-.07	.95

Note. Variables 13 to 19 are daily measures; means and standard deviations (by taking the square root of the between-person variance, u_{0j}) for these variables are taken from intercept only models. *N*s between 971-972 for the (daily) impression variables, *N*s between 1,218 and 1,223 for all other variables. All correlations are significant ($p < .05$) except for those in bold. Same constructs measured by different methods are highlighted. Reliabilities are on the diagonal.
BFI = Big Five Inventory; FIRNI = Five Individual Reaction Norms Inventory.

The GFP, Self-esteem and Mood

At the between-person level, self-esteem, PA and NA appeared to capture relatively different but overlapping aspects of subjective well-being. Specifically, self-esteem showed notable associations with PA (r = .51) and NA (r = -.66), while the relation between PA and NA was small (r = -.05, p = .07). In line with previous studies (e.g., Musek, 2007), we found sizeable correlations between both GFP measures and averaged self-esteem (r_{BFI} = .52, r_{FIRNI} = .56), PA (r_{BFI} = .43, r_{FIRNI} = .41) and NA (r_{BFI} = -.36, r_{FIRNI} = -.39).

Relations between the GFP and Daily Social Experiences

As predicted, both GFP scores positively related to ratings of relationship quality and daily impressions on others, and negatively related to the number of conflicts (Table 1). These results support H1a-H1c. Interestingly, the relations between the GFP scores and the daily indicators of social effectiveness were roughly equal in size or larger than those involving the Big Five and these outcomes. Thus, based on zero-order correlations, the GFP appeared to be at least as important as the lower order Big Five in terms of its relations with the social indicators.

To test this notion more thoroughly, we compared the relative contribution of the Big Five traits beyond the contribution of the GFP in terms of their explained variance. Table 2 shows that in most cases, the individual Big Five factors added little explained variance in the predictions of the outcomes beyond the GFP. The exception was the prediction of interpersonal conflict (especially for the BFI). Yet the majority of variance in the indicators of social effectiveness was attributable to the GFP.

Table 2. Results from relative importance analyses.

	BFI				*FIRNI*			
	R^2 GFP	R^2 Unique Big Five variances	Total R^2	% GFP	R^2 GFP	R^2 Unique Big Five variances	Total R^2	% GFP
Self-esteem	.273***	.022***	.295***	93	.309***	.010***	.319***	97
Positive affect	.184***	.000	.184***	100	.167***	.010***	.177***	94
Negative affect	.132***	.037***	.169***	78	.151***	.014***	.165***	92
Interpersonal conflict	.006**	.003	.009***	67	.007**	.003	.010***	70
Interpersonal conflict (no/yes)	.006**	.004*	.010***	60	.005*	.001	.006***	83
Relationship quality	.108***	.001	.109***	99	.109***	.000	.109***	100
Impressions made on others	.198***	.006**	.204***	97	.142***	.001	.143***	99

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

GFP = General Factor of Personality; *BFI* = Big Five Inventory; *FIRNI* = Five Individual Reaction Norms Inventory.

Results from Mediation Analyses

Given that the GFP related to both the mediators and outcomes, mediation analyses (Table 3) were permissible (Hayes, 2013). For illustrational purposes, Figure 1 summarizes the results for the mediation of the GFP_{BFI} – self-esteem relationship through impressions made on others.

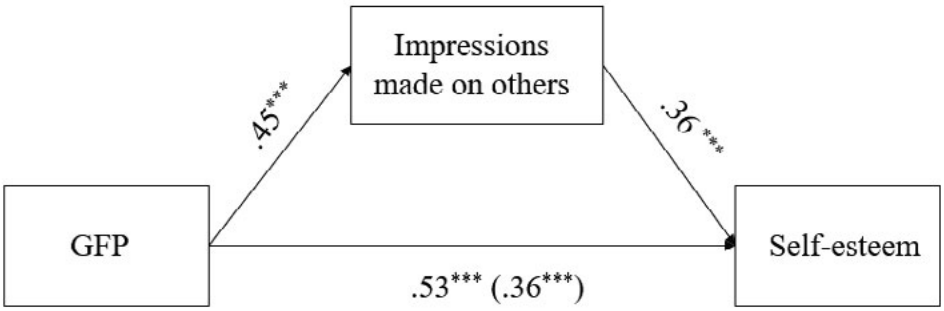

Figure 1. Standardized regression coefficients for the relationship between the General Factor of Personality (GFP) and self-esteem mediated by the impressions made on others. The standardized regression coefficient between the GFP and self-esteem controlling for impressions made on others (c') is in parentheses. The indirect effect of the GFP on self-esteem through impressions (a x b) was .16 ($p < .001$).

Focusing on the direct/total effect ratio, the most important mediators of the relations between the GFP and all three outcomes were relationship quality and daily impressions made on others. For self-esteem and PA, daily impressions were the most important mediators of the effect of the GFP. For PA, as much as about half of the total effect of the GFP was mediated by the daily impressions made on others. Relationship quality was the most important mediator for NA. These results support the predictions from H2b and H2c. The effect of the GFP on self-esteem and mood did not appear to be substantively mediated by the number of conflicts with the identified peers. Thus, only limited support was found for H2a.

Table 3. Results from mediation analyses.

	BFI																	
	Self-esteem						Positive affect						Negative affect					
	a	b	Indirect (a x b)	Direct c'	Total	%	a	b	Indirect (a x b)	Direct c'	Total	%	a	b	Indirect (a x b)	Direct c'	Total	%
Interpersonal conflict	-.07	-.13	.01	.51	.52	2	-.07	.07	.00	.43	.42	-1	-.07	.26	-.02	-.34	-.36	5
Interpersonal conflict (no/yes)	-.08	-.11	.01	.51	.52	2	-.08	.07	-.01	.43	.42	-1	-.08	.29	-.02	-.34	-.36	6
Relationship quality	.33	.30	.10	.42	.52	19	.33	.21	.07	.36	.42	17	.33	-.31	-.10	-.26	-.36	28
Impressions made on others	.45	.36	.16	.36	.53	31	.45	.44	.20	.21	.40	49	.45	-.17	-.07	-.31	-.38	19
	FIRNI																	
	Self-esteem						Positive affect						Negative affect					
	a	b	Indirect (a x b)	Direct c'	Total	%	a	b	Indirect (a x b)	Direct c'	Total	%	a	b	Indirect (a x b)	Direct c'	Total	%
Interpersonal conflict	-.08	-.13	.01	.54	.56	2	-.08	.06	.00	.41	.41	-1	-.08	.26	-.02	-.37	-.39	5
Interpersonal conflict (no/yes)	-.07	-.12	.01	.55	.56	1	-.07	.06	.00	.41	.41	-1	-.07	.29	-.02	-.37	-.39	5
Relationship quality	.33	.29	.10	.46	.56	17	.33	.23	.07	.33	.41	18	.33	-.30	-.10	-.29	-.39	26
Impressions made on others	.37	.37	.14	.41	.55	25	.37	.46	.17	.20	.38	46	.37	-.17	-.06	-.33	-.40	16

Note. All values represent standardized coefficients. Interpretation of a, b, and c' are provided in Figure 1. All indirect effects were significantly different from zero (at $\alpha = .05$). Total = direct + indirect effect. % = indirect / total effect.
BFI = Big Five Inventory, *FIRNI* = Five Individual Reaction Norms Inventory.

Results from Moderation Analyses

Before proceeding to the results from the moderation analyses, a note on statistical power should be made here. Because of our large number of observations (> 20,000), we run the risk of finding significant effects even when, in reality, they are trivial (i.e., inflated Type I error rate). To address this issue, auxiliary power analyses were conducted which are detailed in Appendix A3. These analyses showed that the effects reported below are likely to reflect true population values rather than being artefacts of the large number of observations.

Results for the HLM-analyses are presented in Table 4 (BFI) and Table 5 (FIRNI). Note that, for daily self-esteem and PA, we expected a negative interaction effect between the GFP and both daily relationship quality and daily impressions, and a positive interaction effect with daily interpersonal conflict. For NA, a reverse pattern was expected (H3a-H3c).

Tables 4 and 5 show that the hypothesized effects were largely found for self-esteem and NA. Although small in absolute terms, the cross-level interaction effects between the GFP and the various daily measures resulted in non-trivial decreases of random slope variance (between 0.10% and 5.51%, interpretable as R^2-values). Taking the moderating effect of the GFP_{BFI} and NA as an example, we found that one unit decrease in daily impressions – or having a comparatively bad day compared to an average day – results in an increase in negative affect of about .30 for low-GFP individuals and .24 for high-GFP individuals (Figure 2A).

Table 4. Results from multilevel regression analyses examining the interaction between the GFP and daily social experiences on self-esteem, positive affect, and negative affect – Big Five Inventory.

	Self-esteem				Positive affect				Negative affect			
	1	2	3	4	1	2	3	4	1	2	3	4
Intercept	3.89***	3.97***	3.89***	3.88***	2.87***	2.91***	2.87***	2.87***	1.82***	1.72***	1.82***	1.83***
GFP	.31***	.29***	.31***	.32***	.22***	.23***	.22***	.21***	-.19***	-.18***	-.19***	-.21***
1. Interpersonal conflict	-.17***				-.09***				.22***			
2. Interpersonal conflict (no/yes)		-.21***				-.10***				.28***		
3. Relationship quality			.33***				.27***				-.28***	
4. Impressions made on others				.48***				.46***				-.27***
GFP x Interpersonal conflict	.01†				-.01				.00			
GFP x Interpersonal conflict (no/yes)		.03**				-.01				-.02†		
GFP x Relationship quality			-.05***				.00				.03**	
GFP x Impressions made on others				-.06***				.03**				.03***
% slope variance explained	.10	3.13	4.37	4.04				2.06		.48	1.47	2.34
φ	.22***	.22***	.22***	.22***	.21***	.21***	.21***	.21***	.26***	.26***	.26***	.26***
r_{ij}^2	.22***	.21***	.23***	.23***	.20***	.21***	.20***	.20***	.22***	.20***	.22***	.22***
u_{0j}^2	.03***	.03***	.06***	.06***	.01***	.01***	.03***	.03***	.03***	.04***	.04***	.04***
u_{1j}^2	.40***	.41***	.38***	.33***	.36***	.36***	.34***	.29***	.27***	.27***	.26***	.27***

Note. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$; Number of persons: 1219 (M1-M3), and 968 (M4). Number of observations (days): 23.518 (M1-M2), 23.491 (M3), and 19.996 (M4).

GFP = General Factor of Personality; φ = autocorrelation; r_{ij}^2 = within-person variance; u_{0j}^2 = between-person variance; u_{1j}^2 = random slope variance (see Appendix A2 for more information).

Table 5. Results from multilevel regression analyses examining the interaction between the GFP and daily social experiences on self-esteem, positive affect, and negative affect – Five Individual Reaction Norms Inventory.

	Self-esteem				Positive affect				Negative affect			
	1	2	3	4	1	2	3	4	1	2	3	4
Intercept	3.89***	3.97***	3.89***	3.89***	2.87***	2.91***	2.87***	2.87***	1.82***	1.71***	1.82***	1.82***
GFP	.33***	.31***	.33***	.33***	.21***	.21***	.21***	.20***	-.21***	-.19***	-.21***	-.21***
1. Interpersonal conflict	-.17***				-.09***				.22***			
2. Interpersonal conflict (no/yes)		-.21***				-.10***				.28***		
3. Relationship quality			.33***				.26***				-.28***	
4. Impressions made on others				.48***				.46***				-.27***
GFP x Interpersonal conflict	.02**				.00				.00			
GFP x Interpersonal conflict (no/yes)		.04***				.00				-.02†		
GFP x Relationship quality			-.05***				.00				.02*	
GFP x Impressions made on others				-.06***				.03**				.04***
% slope variance explained	1.29	4.95	3.08	5.51				2.21		.65	1.28	3.10
φ	.22***	.22***	.22***	.22***	.21***	.21***	.21***	.21***	.26***	.26***	.26***	.26***
r_{ij}^2	.21***	.20***	.21***	.22***	.20***	.21***	.21***	.21***	.22***	.29***	.22***	.22***
u_{0j}^2	.02***	.03***	.06***	.06***	.01***	.01***	.03***	.03***	.03***	.04***	.04***	.04***
u_{1j}^2	.40***	.41***	.38***	.33***	.36***	.36***	.34***	.29***	.27***	.27***	.26***	.27***

Note. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$; Number of persons: 1221 (M1-M3), and 970 (M4). Number of observations (days): 23.543 (M1-M2), 23.516 (M3), and 20.041-20.042 (M4).

GFP = General Factor of Personality; φ = autocorrelation; r_{ij}^2 = within-person variance; u_{0j}^2 = between-person variance; u_{1j}^2 = random slope variance (see Appendix A2 for more information).

At first sight, Figure 2A may give the impression of a floor effect because, compared to low-GFP individuals, high-GFP individuals have low scores on NA and thus less opportunity to move down the response scale. However, this methodological explanation is at odds with GFP moderation of the relation between conflict and both NA and self-esteem. For example, for high-GFP individuals, there is enough leeway for interpersonal conflicts to negatively affect one's daily self-esteem (Figure 2B). Yet, as expected, the negative effect of interpersonal conflicts on daily self-esteem is stronger (i.e., steeper) for those with lower GFP scores.

For both of the GFP measures, none of the hypothesized moderating effects were found when positive affect was used as the criterion. Interestingly, for both GFP measures and opposite to expectations, the interaction effect between the GFP and daily impressions on PA was found to be positive: a day with comparatively bad impressions resulted in a larger decrease of PA for high-GFP individuals. In conclusion, for daily self-esteem and NA, H3a through H3c were largely supported, while no support for H3 was found when daily PA was the outcome.

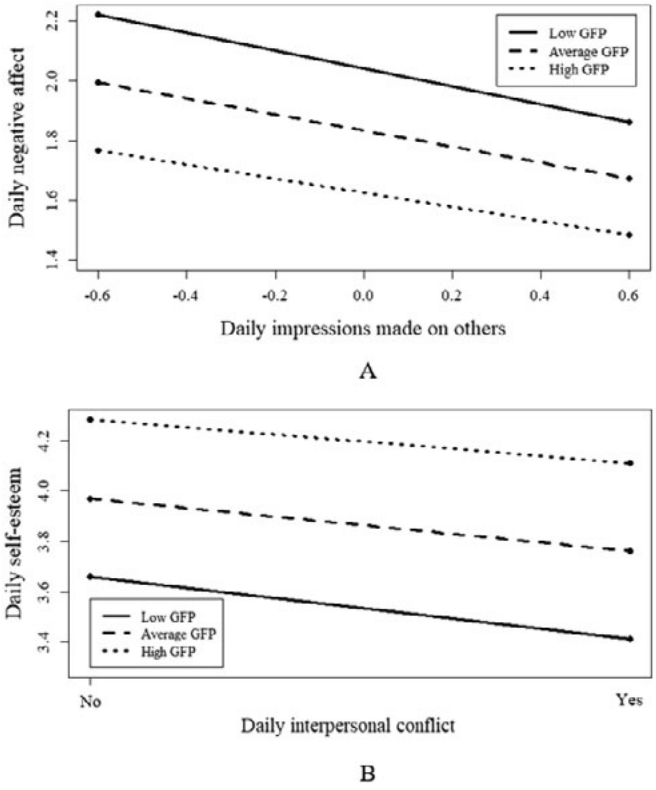

Figure 2. Cross-level interaction between the General Factor of Personality (GFP) and daily impressions made on others on daily negative affect (Figure 2A) and daily interpersonal conflict on daily self-esteem (Figure 2B).

Validity Check 1: The Role of Self-esteem

To examine the robustness and validity of our results, we inspected whether alternative explanations of our findings could be ruled out. Specifically, we found a considerable association between the GFP and a trait measure of self-esteem included in the general pre-diary questionnaire ($r_{BFI} = .68$ and $r_{FIRNI} = .69$), raising the question whether the GFP is not merely self-esteem (e.g., Şimşek, 2012) and thus whether the results hold when trait self-esteem is controlled for.

We tested this by repeating the analyses reported in Table 4 and Table 5, now controlling for the trait self-esteem variable; this did not change the outcomes as reported above. In fact, in a model where daily self-esteem was only predicted by trait (pre-diary) self-esteem and the GFP, the effect of the GFP was still found to be significant (BFI: $b = .04$, $p = .03$; FIRNI: $b = .05$, $p = .01$). Thus, GFP scores predicted self-esteem measured at the daily level, even when controlling for baseline levels of this trait.

Validity Check 2: The Role of Neuroticism and Extraversion

Previous studies have predominantly linked Extraversion and Neuroticism to social relationship variables as well as subjective well-being (e.g., Asendorpf & Wilpers, 1998; Watson, Clark, McIntyre, & Hamaker, 1992). In addition, the loadings of E and N on both GFP measures were relatively high. Together, this might suggest that the relations found in the current study are mainly driven by E and N, rather than by the shared variance of all Big Five domains (i.e., the GFP).

Additional analyses (described in more detail in Appendix A4) revealed that the GFP predicted a significant amount of variance in the outcomes even on top of the unique variance of Extraversion and Neuroticism. We did find that E and N often had significant effects on the outcomes when controlling for the effect of the GFP; thus, in line with previous studies we found E and N to be relevant predictors of social relations and well-being. Yet, it could be concluded that the findings on the GFP reported above were not due to a few specific Big Five dimensions (such as E and N) but seem to truly reflect the effects of the shared variance of all Big Five dimensions.

Discussion

The present study showed that 1) GFP scores were related to daily social experiences and well-being/mood, 2) that daily social experiences partly mediated the relation between the GFP and well-being/mood, and 3) that the GFP related to how individuals react to daily social events. To the best of our knowledge, this is the first time that, in a large sample, the GFP is studied using daily reports of social experiences and well-being and mood, allowing us to test – and largely support – the social effectiveness hypothesis (Van der Linden et al., 2016).

This study may contribute to the personality literature in a number of ways. First, we found that the GFP was related to enjoying better social relationships, as indicated by higher relationship quality and less interpersonal conflicts. These outcomes can be viewed as relatively strong indicators of social effectiveness (Denissen et al., 2008). Combining this notion with the

results from previous studies showing associations between the GFP and popularity (Van der Linden et al., 2010b), job performance and other social organizational outcomes such as leadership (Pelt et al., 2017), the current study corroborates the perspective of the GFP as a broad social effectiveness factor that facilitates prosocial behavior influencing a wide range of life domains (Van der Linden et al., 2016).

Second, the present study showed how GFP scores were associated with leaving better impressions on others on a daily basis. Recent studies have confirmed that impression management is best seen as a stable, substantive trait related to self-control in social contexts (Uziel, 2010a). This definition is not far from the substantive interpretation of the GFP. Accordingly, it can be argued that (successful) impression management is inseparable from personality (cf. Danay & Ziegler, 2011). This argument fits with the large overlap found between the GFP and trait emotional intelligence, which influences virtually all of personality by being able to correctly interpret and act on social cues (Van der Linden et al., 2016).

A third contribution is our provision of a potentially relevant mechanism for the strong relationship between the GFP and subjective well-being (e.g., Musek, 2007). Because social relationships have been proposed to be "the greatest single cause" of well-being (Argyle, 2001), it may not come as a surprise that the social skills associated with the GFP allows for maintaining better social relationships in turn resulting in higher levels of well-being.

Another relevant finding was that, compared to low-GFP individuals, high-GFP individuals' daily mood was less strongly influenced by daily fluctuations in social interactions and events. These results support the notion of the GFP as an *adaptive* trait that not only reflects social aptness, but that also cushions the impact of adversities. More evidence for this claim comes from Hengartner et al. (2017), who showed that higher GFP scores were associated with more socially adaptive coping strategies following a stressful event.

A final contribution of this study relates to the finding that the GFP appeared to account for the lion's share of the overlap between the Big Five and well-being, mood, and the daily indicators of social effectiveness. Such results do not imply that the Big Five become obsolete. Prediction of a criterion tends to be optimal when the predictor and criterion are conceptually aligned at the same level: This is known as the 'bandwidth-fidelity' debate (Ones & Viswesvaran, 1996) and a similar pattern has also been demonstrated in the GFP literature (Sitser et al., 2013).

Counter to our hypotheses, we did not find a moderating effect of the GFP on the relation between daily conflict and relationship quality on the one hand, and PA on the other. One possible explanation is that the participants' PA-levels resided around the midpoint of the scale, and daily social experiences may not be salient enough to warrant a reaction at such levels. In contrast, average self-esteem scores were relatively high and NA scores relatively low; a deviation from such higher levels will perhaps trigger a more direct reaction (thus allowing for a moderation of this reaction by the GFP). At this point, however, this explanation is rather speculative and should be tested in the future.

Salience of a daily experience may also be responsible for the unexpected finding that fluctuations in daily impressions on others had *stronger* effects on the daily PA of high (vs. low) GFP individuals. Note that by far the largest mediation effect was found for the GFP – daily impressions – PA link. Thus, given that leaving a good impression on others is especially

important when PA is concerned, daily successes or failures in achieving this can be assumed to be more pleasing or disturbing, respectively, at higher GFP levels.

Limitations

Some limitations of this study are worth mentioning here. First, participants were requested to identify the interaction partners (i.e., friend, family member, and partner) they had most contact with. This presumably resulted in identification of the most important people in one's life, which was reflected in relatively high reports of relationship quality ($M = 3.94$, possible range 1-5) and few interpersonal conflicts. One could argue that this is not a good reflection of one's general social life. On the other hand, interactions with more proximal people (e.g., partners) will presumably have larger effects on well-being than those with more peripheral others (e.g., acquaintances).

Although the current study benefits from using a large community sample, participants were relatively young, childless, and the majority consisted of women. Therefore, generalizability of the results in more heterogeneous samples would be desirable in the future.

A final limitation was that all analyses were based on self-reports, introducing possible influences of common method bias on the results. However, the use of diary data can be assumed to reduce some biases associated with self-reports (e.g., recall bias and social desirability). In addition, by looking at within-person fluctuations over the days, individual differences in response tendencies are eliminated, reducing the influence of common method variance (Beal, 2015). Furthermore, it is unlikely to find cross-level interactions as found in the present study when large amounts of common method variance are present (Lai, Li, & Leung, 2013). Still, being self-reports, it is unclear to what extend an event (e.g., interpersonal conflict) really occurred during the day. It could be that lower GFP scores are associated with quicker interpretation of a given social situation as a conflict or even with selecting oneself into conflicts (e.g., Bolger & Schilling, 1991). Future studies should include other-reports to remedy these drawbacks.

Concluding Remarks

The present study revealed how the GFP, as a presumed social effectiveness factor, translates to day-to-day social experiences. Using an extensive diary design, it was found that individuals scoring high on the GFP tend to experience fewer interpersonal conflicts, and are less negatively influenced by potentially disruptive social events. It is not difficult to imagine how the effects of being socially adaptive and knowing how to adequately respond to social situations on a daily basis will accumulate and eventually would affect broader life outcomes such as job performance and better social relations.

Appendix A1: GFP extraction methods

We supplemented the analysis and extraction of the GFP by means of EFA with analysis and extraction by confirmatory factor analyses, all conducted in AMOS 21.0. For the BFI, a model in which the Big Five directly loaded onto the GFP showed just below acceptable fit (*CFI*: .89, *RMSEA*: .11 (90% CI including .09) and *SRMR*: .05; see Table A1.1, $GFP_{CFA\_direct\_factor}$). The loadings for O, C, E, A, and N were .38, .42, .66, .46, and -.59, respectively. Modification indices implied that substantial residual relations remained between E and O on the one hand, and C, A and N. This pattern reflects the meta-order traits Alpha and Beta, or Stability and Plasticity (DeYoung et al., 2002; Digman, 1997). A hierarchical model was fitted in which O and E loaded on Beta and C, A and N on Alpha. These two meta-level traits in turn loaded onto the GFP. This model showed excellent fit to the data (*CFI*: .98, *RMSEA:* .06, *SRMR*: .02; $GFP_{CFA\_hier\_factor}$ in Table A1.1).

For the FIRNI, we also found the direct model to show below acceptable fit (*CFI*: .86, *RMSEA*: .10 (90% CI including .09) and *SRMR*: .05). In this model, the loadings for O, C, E, A, and N were .29, .45, .50, .08, and -.70, respectively. In the design of the FIRNI, Extraversion and Neuroticism are both assumed to be outcomes of the punishment/reward system and can therefore be expected to show a residual correlation. In fact, Denissen & Penke (2008b) indeed showed the relation between E and N to be the highest among all of the FIRNI intercorrelations. Adding the residual correlation between E and N resulted in a well-fitting model (*CFI*: .97, *RMSEA*: .05 and *SRMR*: .03; $GFP_{CFA\_direct\_factor\_res}$ in Table A1.1). Note that the decision to free the residual correlation between E and N was based on theoretical considerations, rather than on modification indices. From the analyses described above we can conclude that clear and robust general factors were found in both personality measures.

Because previous studies have criticized popular methods of GFP extraction in the literature (e.g., Ashton et al., 2009; Revelle & Wilt, 2013), we decided to test the robustness of the GFP across methods of extraction. Therefore, GFP scores were also computed based on item-level data with similar models as described above ($GFP_{CFA\_direct\_item}$ and $GFP_{CFA\_hierarchical\_item}$). In addition, a bifactor model, a model that has gained popularity in the personality literature in recent years (e.g., Biderman et al., 2018), was fitted on the item-level data ($GFP_{CFA\_bifactor\_item}$).

After fitting the CFA models, GFP scores were saved using the regression method, which we then correlated with each other, as well as with the GFP scores based on the EFAs (Table A1.2). Overall, the results of the analyses indicated that the GFPs based on very different models and extraction methods largely converged. Within the BFI, the correlation between the different GFP measures was .92 on average (*Med.* = .95, *SD* = .07, *Min.* = .73, *Max.* = 1.00), while for the FIRNI the average correlation was .90 (*Med.* = .94, *SD* = .09, *Min.* = .73, *Max.* = 1.00). Across instruments, the average correlation was lower but substantial (*r* = .76; *Med.* = .77, *SD* = .05, *Min.* = .59, *Max.* = .84). These results indicate that GFP scores based on different methods largely converge.

Finally, to address the critique that saturation by the GFP is too low to be truly considered a 'general' factor (Revelle & Wilt, 2013), the *psych* package in R was used to calculate the hierarchical omega coefficient (ω_h) which is a measure of explained variance by a higher-order factor (e.g., McDonald, 1999). For the BFI, ω_h was .45 both when calculated at the factor and item level, while for the FIRNI the values respectively were .37 and .49. Although not as high as for

example in the intelligence domain, these levels of explained variance are relatively high, considering the fairly limited amount of items/factors, and the amount of effort put into the development of the personality instruments to measure the Big Five as independent factors. In addition, the values are in line with those found in previous studies (see Chen, Watson, Biderman, & Ghorbani, 2016).

In sum, it can be concluded that there appeared to be a general factor in both personality measures, which largely converged, regardless of the method of extraction.

Table A1.1. Confirmatory factor analysis model fit indices.

	χ^2	*df*	*CFI*	*TLI*	*RMSEA*	*RMSEA 90% CI*
BFI						
$GFP_{CFA\_direct\_factor}$	79.09	5	.89	.78	.110	.090-.132
$GFP_{CFA\_direct\_item}$	7450.11	814	.71	.68	.082	.080-.083
$GFP_{CFA\_hier\_factor}$	19.40	4	.98	.94	.056	.033-.082
$GFP_{CFA\_hierarchical\_item}$	7459.91	815	.71	.68	.082	.080-.083
$GFP_{CFA\_bifactor\_item}$	5872.17	777	.78	.74	.073	.072-.075
FIRNI						
$GFP_{CFA\_direct\_factor}$	64.55	5	.86	.73	.099	.078-.121
$GFP_{CFA\_direct\_factor\_res}$	17.36	4	.97	.92	.052	.029-.079
$GFP_{CFA\_direct\_item}$	5397.77	1172	.81	.79	.054	.053-.056
$GFP_{CFA\_hier\_factor}$	86.13	5	.82	.63	.115	.095-.137
$GFP_{CFA\_hierarchical\_item}$	5444.63	1173	.80	.79	.055	.053-.056
$GFP_{CFA\_bifactor\_item}$	4506.80	1127	.84	.82	.050	.048-.051

Note. χ^2 = Chi-square statistic; *df* = degrees of freedom; *CFI* = Comparative Fit Index; *TLI* = Tucker-Lewis index; *RMSEA* = Root Mean Square Error of Approximation; *CI* = confidence interval; GFP = General Factor of Personality; BFI = Big Five Inventory; FIRNI = Five Individual Reaction Norm Inventory.

Table A1.2. Correlations between General Factor of Personality (GFP) scores based on different extraction methods.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	BFI															
1.	GFP_{EFA_factor}															
2.	GFP_{EFA_item}	.98														
3.	$GFP_{CFA\_direct\_factor}$	1.00	.98													
4.	$GFP_{CFA\_direct\_item}$	.99	.98	.99												
5.	$GFP_{CFA\_hier\_factor}$	.95	.95	.95	.95											
6.	$GFP_{CFA\_hierarchical\_item}$	.95	.90	.95	.94	.80										
7.	$GFP_{CFA\_bifactor\_item}$	.85	.91	.86	.84	.89	.73									
	FIRNI															
8.	GFP_{EFA_factor}	.81	.80	.81	.78	.76	.76	.75								
9.	GFP_{EFA_item}	.82	.84	.82	.80	.78	.77	.80	.96							
10.	$GFP_{CFA\_direct\_factor}$	.80	.79	.80	.78	.76	.75	.74	1.00	.94						
11.	$GFP_{CFA\_direct\_factor\_res}$	.69	.72	.69	.66	.59	.69	.68	.86	.91	.82					
12.	$GFP_{CFA\_direct\_item}$	.82	.82	.82	.79	.77	.76	.77	.99	.97	.99	.84				
13.	$GFP_{CFA\_hier\_factor}$	.80	.79	.80	.78	.78	.73	.74	.98	.93	.99	.76	.98			
14.	$GFP_{CFA\_hierarchical\_item}$	.80	.81	.80	.78	.75	.75	.75	.98	.96	.97	.87	.99	.97		
15.	$GFP_{CFA\_bifactor\_item}$	.71	.75	.71	.70	.70	.63	.74	.78	.84	.77	.73	.81	.76	.83	

Note. All correlations are significantly different from zero (at $\alpha = .001$). GFP = General Factor of Personality; BFI = Big Five Inventory; FIRNI = Five Individual Reaction Norm.

Appendix A2: Multilevel Analyses

The following multilevel models were estimated for both GFP measures. Each model is estimated by the following equations;

$$y_{ij} = \beta_{0j} + \beta_{1j}(DE) + r_{ij} \qquad (\text{Level 1})$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(GFP) + u_{0j} \qquad (\text{Level 2})$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(GFP) + u_{1j} \qquad (\text{Level 2})$$

where DE stands for daily experience. Taking self-esteem as the dependent variable as an example, the first equation states that daily self-esteem is a function of one's mean level of self-esteem (β_{0j}) over the days and the effect of one's daily report of the social indicator (β_{1j}). The last term, r_{ij}, indicates random error and its variance constitutes the within-person residual variance. The second equation states that each person's intercept is a function of the overall intercept (γ_{00}, the grand mean of person-level means, the β_{0j}s) and the additive effect of one's GFP-level. The error term, u_{0j}, indicates that the intercept is allowed to vary across persons and its variance is the Level 2 (between-person) residual variance. And finally, the last equation indicates that for each individual *j*, the slope of DE on the outcome is a function of the overall slope for DE (γ_{10}, the grand mean of the person level slopes, the β_{1j}s) and the additive effect of one's GFP level. The last error term, u_{1j}, indicates that the slope for DE is allowed to vary across persons and its variance is the variance of the random effect of DE. Moderation of the DE-outcome relationship is indicated by the cross-level interaction term, γ_{11}, significantly different from zero.

As noted and indicated by r_{ij}, u_{0j} and u_{1j}, parameters were allowed to vary across persons: this decision was based on both theoretical and statistical grounds. Theoretically, the days in this study can be viewed as being a random sample from a population of days and coefficients should therefore be modeled as random. We checked feasibility of random terms by the 95% confidence intervals of the variance components as well as likelihood ratio tests between models including and excluding a random term (e.g. one with and without u_{0j}) which all indicated that random effects were appropriate. In repeated measures over time, due to temporal stability of traits, observations temporally close to each other may be more highly correlated than those temporally far apart (Bolger et al., 2003). This problem of autocorrelation might lead to autocorrelated residuals influencing estimated standard errors. We therefore incorporated an AR(1) error structure into our models to control for these detrimental influences.

Individual differences in the daily measured social experiences can influence the parameter estimates (Nezlek, 2001). The daily predictors were person-mean centered to prevent this. A participant's coefficient for a daily experience (β_{1j}) therefore reflects increases or decreases from his or her average level of that experience. Just as in ordinary regression, each parameter represents the effect of that variable while the other predictors are zero. Since GFP scores in the present study are standard normal scores ($M = 0$, $SD = 1$), γ_{10} represents the effect of DE for persons with average GFP scores at average levels of the daily measured experience.

The random slope variance of the models including the cross-level interaction (M2) were compared with models excluding the interaction (M1); analogous to R^2, the decrease in variance from M1 to M2 can be viewed an indication of the relevance of the interaction effect (Hox, 2002).

Means and standard deviations for the daily measures in Table 1 were taken from the intercept-only models in the HLM analyses, i.e. models in which only the between and within person variances are estimated.

On a final note, following recommendations by Breevaart, Bakker, Demerouti and Hetland (2012), we examined the factor structures of the constructs at the daily level by means of multilevel confirmatory factor analysis (MCFA) in MPlus (Muthén & Muthén, 1998-2010). In MCFA, multiple covariance matrices are estimated corresponding to the number of levels (in our case two: between- and within-person) on which traditional CFA techniques are applied. Tested against the criteria of Hu and Bentler (1999), all analyses showed that creating a daily score by averaging over the different items for each of the daily measured constructs was justified. Furthermore, a MCFA model with all (positive and negative) mood adjectives loading on one single factor did not converge. A full model with correlated PA and NA factors showed adequate fit ($CFI = .91$, $RMSEA = .04$, $SRMR_{within} = .053$). The correlation between both factors was -.40. This indicated that daily PA and NA can be regarded as distinct, although related, concepts (Diener & Emmons, 1985). Results of the full MCFA analyses are available on request.

Appendix A3: Power Analysis

Given the very large number of observations (> 25,000 daily reports), the current study suffers from a problem that is rare in the psychological literature, namely too much statistical power. This means that any small – but trivial – effect will be found to be significant. Although the percentages of explained random slope variance by the interaction term (Table 4 and Table 5) already give an indication of effect size, to further test whether this abundance of power influenced our results a bootstrapping procedure was conducted. Four roughly equal groups were resampled from the current sample, after which a MANOVA was conducted on the variables of interest (i.e., the two GFP scores, and self-esteem, PA, NA, interpersonal conflict, relationship quality and impressions made on others, averaged over the diary days). If the test was non-significant, the four groups were retained, while if the test was significant, four new groups were sampled. In this way, comparable groups were formed. This procedure was repeated until 250 x 4 groups were sampled. For each of these groups, the models in Table 4 and Table 5 were estimated, and the coefficients for the interaction term, γ_{11}, in each model were saved. The distribution of these 1000 coefficients can be used to construct confidence intervals and subsequently test whether these coefficients are significantly different from zero at different nominal α-values (i.e., by checking whether the interval includes zero). Results for the 95% (α = .05) and 99% (α = .01) confidence intervals (CIs) for both the BFI and FIRNI based GFPs are presented in Table A3.1.

Table A3.1 shows that the results were very similar for both GFP measures. Overall, when self-esteem was the dependent variables, the 95% CIs of the interaction coefficients of the dichotomously measured conflict variable just included zero. The interaction coefficients of both the relationship quality and impressions made on others variable virtually all fall below zero, as expected, at the .05 and .01-level. Finally, for continuously measured conflict, both the 95% and 99% CIs clearly include zero. However, when we look at the percentage of coefficients > 0 (5th and 10th column of Table A3.1), we see that most coefficients are found to be in the expected direction. The same conclusion holds for the rest of the coefficients.

With negative affect as the dependent variable, we find similar results. The interaction effects of relationship quality and impressions made on others appear to be most robust in the sense that the confidence intervals largely excluded zero. The confidence intervals of the interaction effects of both interpersonal conflict variables more clearly included zero, which is in line with the values reported in Table 4 and Table 5 (a non-significant interaction effect for the continuously measured conflict variable, and a borderline ($p < .10$) significant interaction effect for the dichotomous conflict variable). Yet, for the dichotomous conflict variable, most coefficients (about 85%) fell below zero, as expected.

Finally, the interaction effects found for positive affect in the bootstrapping procedure correspond with those reported in Table 4 and Table 5. Thus, overall these results indicate that the effects we found are close to the "true" population values rather than a by-product of the size of our specific sample.

Table A3.1. Results from bootstrapping analyses on interaction effect coefficients from multilevel regression analyses.

	BFI				*FIRNI*			
	Median	95% CI	99% CI	% > 0*	*Median*	95% CI	99% CI	% > 0*
Self-esteem								
Interpersonal conflict	.01	-.013 – .043	-.024 – .058	83.5	.02	-.008 – .054	-.019 – .066	92.2
Interpersonal conflict (no/yes)	.03	-.004 – .066	-.016 – .077	95.8	.04	-.002 – .075	-.013 – .087	97.2
Relationship quality	-.05	-.094 – -.020	-.105 – -.007	100	-.05	-.087 – -.011	-.093 – -.002	99.8
Impressions made on others	-.06	-.097 – -.019	-.106 – -.004	99.8	-.06	-.103 – -.028	-.110 – -.016	100
Positive affect								
Interpersonal conflict	-.01	-.030 – .015	-.037 – .022	24.4	-.00	-.030 – .019	-.036 – .028	35.1
Interpersonal conflict (no/yes)	-.01	-.039 – .025	-.048 – .036	33.1	-.00	-.037 – .029	-.046 – .040	42.4
Relationship quality	.00	-.026 – .037	-.039 – .041	41.0	-.00	-.035 – .028	-.046 – .040	54.5
Impressions made on others	.03	-.005 – .059	-.013 – .069	5.3	.03	-.008 – .060	-.018 – .068	7.6
Negative affect								
Interpersonal conflict	-.00	-.032 – .022	-.043 – .031	62.9	-.01	-.030 – .025	-.039 – .032	57.2
Interpersonal conflict (no/yes)	-.02	-.052 – .018	-.060 – .030	84.2	-.02	-.052 – .016	-.063 – .029	85.5
Relationship quality	.03	-.005 – .056	-.016 – .066	94.3	.02	-.008 – .056	-.016 – .063	92.9
Impressions made on others	.03	-.001 – .064	-.013 – .078	97.1	.04	.008 – .066	.001 – .076	99.6

Note. CI = confidence interval. Based on 1000 bootstrapping samples.

BFI = Big Five Inventory; *FIRNI* = Five Individual Reaction Norms Inventory.

* For relationship quality and impressions made on others, the values correspond to the percentage of coefficients < 0.

Appendix A4: Investigating the role of Extraversion (E) and Neuroticism (N)

We tested whether the shared variance among the Big Five dimensions as reflected in the GFP added to the prediction in the outcomes beyond the unique variances of E and N, which have predominantly been linked to social outcomes in previous studies (e.g., Asendorpf & Wilpers, 1998; Watson et al., 1992). The GFP is assumed to be present in each of the Big Five factors; thus, each Big Five factor partly reflects the GFP and partly reflects the unique variance of the specific Big Five factor. Therefore, it would not make much sense to investigate the incremental validity of the GFP over Extraversion and Neuroticism, because by first controlling for E and N, one would then already control for the true variance of the GFP present in these two individual dimensions. It would be more viable to test whether the GFP shows incremental validity over the *unique* variances of E and N.

Therefore, we first created residualized Extraversion (E_{unique}) and Neuroticism (N_{unique}) factors by partialling out the variance attributable to the GFP from each factor (see Salgado, Moscoso, and Berges, 2013). These residualized factors are by definition uncorrelated with GFP scores. As such, E_{unique} and N_{unique} (Step 1) and the GFP (Step 2) can be entered in a hierarchical regression analysis and their relative importance can be inferred by simply comparing their proportion of variance explained by each, and their beta weights.

The outcomes of those analyses are described in Table A4.1 below. For both measures and for the prediction of all outcomes, we found that the GFP added a significant amount of explained variance beyond the unique variances of E and N. Focusing on the beta weights, we found that the GFP had the strongest effects on the outcomes compared to the unique variances of E and N. Yet, in several instances, the unique variance of E or N also had a significant effect on the outcome, when controlling for the GFP (although sometimes with a reversed sign compared to what one would expect from the literature; see Van der Linden et al. (2010a) for similar findings). Taken together, the results suggest that the associations between the GFP and the outcomes are not purely due to E + N, but rather due to the overlap between all Big Five factors.

Table A4.1. Summary of hierarchical regression analyses.

	BFI				*FIRNI*			
Variable	β	R^2	ΔR^2	ΔF	β	R^2	ΔR^2	ΔF
Self-esteem								
E_{unique}	.02				.03			
N_{unique}	-.16***	.02***			-.08**	.00*		
GFP	.52***	.29***	.27	470.67***	.56***	.31***	.31	548.36***
Positive affect								
E_{unique}	.05				.05			
N_{unique}	-.02	.00			.11***	.02***		
GFP	.43***	.18***	.18	275.08***	.41***	.18***	.16	249.39***
Negative affect								
E_{unique}	.08**				.04			
N_{unique}	.16***	.04***			.12***	.02***		
GFP	-.36***	.17***	.13	194.24***	-.39***	.17***	.15	221.70***
Interpersonal conflict								
E_{unique}	.08**				.00			
N_{unique}	.02	.01**			.06	.00		
GFP	-.08**	.01**	.01	6.93**	-.08**	.01**	.01	8.50**

(continued)

Table A4.1 (continued).

Variable	*BFI*				*FIRNI*			
	β	R^2	ΔR^2	ΔF	β	R^2	ΔR^2	ΔF
Interpersonal conflict (yes/no)								
E_{unique}	.10**				.04			
N_{unique}	.03	.01***			.02	.00		
GFP	-.08**	.02***	.01	7.40**	-.07*	.01*	.01	5.58*
Relationship quality								
E_{unique}	-.04				.08**			
N_{unique}	.06	.00			-.02	.00*		
GFP	.33***	.11***	.11	148.03***	.33***	.11***	.11	149.39***
Impressions made on others								
E_{unique}	-.08*				.10**			
N_{unique}	.09**	.01*			.01	.01*		
GFP	.45***	.20***	.20	240.20***	.38***	.15***	.14	163.55***

Note. E_{unique} = unique variance of Extraversion; N_{unique} = unique variance of Neuroticism; GFP = General Factor of Personality; BFI = Big Five Inventory; FIRNI = Five Individual Reaction Norms Inventory.

How Emotional Intelligence (EI) Might Get You the Job: The Relationship between Trait EI and Faking on Personality Tests

4

This chapter has been published as:
Pelt, D. H. M., Van der Linden, D., & Born, M. Ph. (2018). How emotional intelligence might get you the job: The relationship between trait emotional intelligence and faking on personality tests. *Human Performance, 31,* 33-54.

Abstract

This study examined trait emotional intelligence in relation to the ability to fake on personality tests. Undergraduate students (N = 129) were first instructed to fill out a personality inventory honestly, and subsequently in such a way as to maximize their chances of obtaining two very distinctive job positions (lawyer and file clerk). Participants were able to change their scores in line with the hypothesized job profiles. Regression analyses showed that EI statistically predicted faking ability to an equal degree in both job scenarios. Finally, EI showed incremental validity over general mental ability and the Big Five personality traits in predicting the ability to fake. Possible implications of the results for the predictive validity of personality tests are discussed.

Introduction

The widespread acceptance of the Big Five personality trait taxonomy has led to an increased use of personality assessments in personnel selection. This interest was fueled partly by meta-analyses showing the predictive power of personality on job performance and related outcomes (e.g., Barrick & Mount, 1991; Barrick et al., 2001).

The upswing of personality assessments was accompanied by criticism on their use. The main concern is the possibility to distort responses on personality tests, for example to answer in a socially desirable way (e.g., Ones et al., 1996). This may particularly be an issue in selection situations when the stakes are high. The tendency to respond in a socially desirable manner in order to maximize the chances of getting hired is often referred to as 'faking' on personality measures (Mesmer-Magnus & Viswesvaran, 2006), yet others use the more neutral term 'impression management' (Hogan et al., 2007). Research has shown that individuals *can* and *do* provide a more favorable view of themselves on personality measures in selection procedures (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Viswesvaran & Ones, 1999). Yet, whether this type of response distortion affects the predictive validity of personality tests remains rather unclear and is strongly debated in the literature (e.g., Cook, 2016; Dilchert et al., 2006; Hogan, Hogan & Roberts, 1996; Morgeson et al., 2007; Murphy & Dzieweczynski, 2005; Rosse, Stecher, Miller, & Levin, 1998; Tett & Simonet, 2011).

Earlier research using social desirability scales as an indication of faking has led some scholars to conclude that this type of response distortion has little effect on the predictive validity of personality constructs (e.g., Barrick & Mount, 1996; Hough et al., 1990; Li & Bagger, 2006). Controlling for scores on social desirability scales only has a small to negligible effect on the relation between personality constructs and work outcomes (Barrick & Mount, 1996; Li & Bagger, 2006; Ones et al., 1996; Schmitt & Oswald, 2006). On the other hand, social desirability scales have been criticized for their lack of construct validity: they appear to overlap with personality traits and thus seem to capture substantive and not error variance (Connelly & Chang, 2016; De Vries, Zettler, & Hilbig, 2014; Konstabel et al., 2006; McCrae & Costa, 1983; Ones et al., 1996; Uziel, 2010a). In this regard, they do not seem to do what they were originally intended for, namely detect faking (e.g., Holden, 2007, 2008).

More recent research indicated how faking can negatively affect the validity of personality test scores (Bing, Kluemper, Davison, Taylor, & Novicevic, 2011; Donovan, Dwight, & Schneider, 2014; Douglas, McDaniel, & Snell, 1996; Holden, 2007, 2008; O'Connell, Kung, & Tristan, 2011; Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011). In these studies, criterion validities were generally lower among participants who could be assumed to be faking (e.g., applicants) compared to those who were likely to respond truthfully (e.g., incumbents). Based on the suggestion that people differ in their amount of faking (McFarland & Ryan, 2000; Melchers et al., 2009; Viswesvaran & Ones, 1999; Zickar & Robie, 1999), scholars have warned that such differences can change the rank order of

candidates in selection procedures and consequently influence the selection outcome, i.e. who gets hired (Rosse et al., 1998).

Yet, whether this poses a genuine problem for the test's predictive validity – and the hiring organization – of course ultimately depends on whether those applicants who faked (and by result were hired) perform poorly on the job (Donovan et al., 2014). Some studies have indicated that this need not necessarily be the case, by showing how faking can have a *positive* effect on personality validities by showing *higher* validities under conditions in which faking is likely to occur (e.g. Blickle et al., 2009; Ingold et al., 2015; Klehe, Kleinmann, Nieß, & Grazi, 2014; Schmit, Ryan, Stierwalt, & Powell, 1995).

These contrasting findings might be due to the use of different definitions of what constitutes 'faking', different designs (within or between-subject) and settings (e.g., instructed or naturally occurring faking), different strategies for classifying 'fakers' (e.g. social desirability scales, magnitude of score changes or using 'bogus' items), different focal constructs (e.g., multiple Big Five factors or one single trait) and different criteria (e.g., supervisor rated performance or objective criteria). But more importantly, faking behavior appears to be a very complex phenomenon in which a multitude of factors simultaneously are at work (Marcus, 2009; Komar, Brown, Komar, & Robie, 2008). It is therefore imperative to identify what factors are responsible for the individual differences in faking behavior.

Several theoretical models have been developed to explain the variance in faking between individuals (Ellingson & McFarland, 2011; Goffin & Boyd, 2009; Marcus, 2009; McFarland & Ryan, 2000, 2006; Mueller-Hanson, Heggestad, Thornthon, 2006; Roulin et al., 2016; Snell, Sydell, & Lueke, 1999; Tett & Simonet, 2011). Although the models differ somewhat in the antecedents or how these are labeled, they largely overlap, and three core elements can be identified: (1) (perceived) motivation to fake (2) (perceived) ability to fake and (3) (perceived) opportunity/risk of faking. The current study focuses on the second antecedent, namely the ability to fake: because faking ability and job performance have previously been linked (e.g., Viswesvaran & Ones, 1999), knowing what factors are related to the ability to fake might help us understand whether faking poses a threat to the predictive validity of personality tests.

That is, if factors influencing one's ability to fake at the same time positively influence performance on the job, threats to validity by faking should be minimal. Scholars have indeed argued that how one behaves in selection procedures actually reflects one's true social effectiveness, i.e. genuine social skills and abilities (Hogan et al., 2007; Van der Linden et al., 2014b; Viswesvaran & Ones, 1999). In this line of argument, responding to a personality test is seen as a social interaction as any other where people try to convey a certain (i.e. positive) image of themselves and in which some will be more successful than others (Hogan, 1982, 2005; Marcus, 2009). This idea is supported by the fact that the same skills or traits that determine whether one is good at faking (or impression management) are likely to be the ones that are valuable and effective on the job (e.g., Blickle et al., 2009; Hogan & Shelton, 1998; Johnson & Hogan, 2006; Komar, Brown, Komar, & Robie, 2008).

A wide range of conceptually similar but more or less distinctive constructs have been gathered under the collective term social effectiveness (see Ferris et al. (2002), for an overview). These constructs have in common that they all relate to social competence and social abilities. Given that successful impression management or faking ability may depend on actual social knowledge and ability, one likely candidate among social effectiveness constructs is emotional intelligence (EI). Although scholars differ widely in their theoretical and operational definitions of EI, they all assume that high-EI individuals have the ability or tendency to act in socially effective ways (e.g., Ferris et al., 2002; O'Boyle, Humphrey, Pollack, Hawver, & Story, 2011).

The most important distinction in the current EI literature is between EI as an ability and EI as a trait. The former considers EI as a set of emotion-related cognitive abilities that can only be assessed through maximum performance tests, akin to the way cognitive intelligence would be measured. The trait conception considers EI as "a constellation of behavioral dispositions and self-perceptions concerning one's ability to recognize, process, and utilize emotion-laden information" (Petrides, Frederickson, & Furnham, 2004; p. 278).

The present study adopts the trait EI model. The sampling domain of trait EI includes facets such as emotion perception, emotion expression, emotion regulation, adaptability and social awareness/competence (Petrides, 2011). Using the trait EI model has the advantage that there are reliable and validated measures to assess the construct, and studies and meta-analyses suggesting that trait EI is related to a range of other-rated and objective outcome measures such as health (Martins, Ramalho, & Morin, 2010), social behavior and leadership (Mavroveli et al., 2007), and most relevant to the present study, supervisor-rated and objective job performance (Joseph, Jin, Newman, & O'Boyle, 2015; Joseph & Newman, 2010; O'Boyle et al., 2011). As far as trait EI relates to understanding one's own emotions and motivations and those of others, and entails the tendency to use this knowledge to be socially effective (e.g., Mavroveli et al., 2007; Sevdalis et al., 2007), it is plausible that trait EI may also play a role in optimizing one's chances of getting the desired job. Behaving effectively in selection settings has much to do with knowing what to say and do in order to create a favorable image of oneself in the eye of the hiring company (e.g., Levashina et al., 2014; Melchers et al., 2009): High levels of trait EI may allow one to adequately achieve this.

Specifically, the effect of trait EI on faking ability can be summarized as follows. When applying for a job, the applicant finds himself/herself in an evaluative situation (Ellingson, Sackett, & Connelly, 2007; Kleinmann et al., 2011; Schmit & Ryan, 1993); unsure about what the hiring company is looking for exactly, the applicant has to interpret the situational cues at hand (e.g. personality test items or interview questions) in order to leave a good impression (Klehe et al., 2012; Roulin et al., 2016). In other words, when applying for a job it can be expected that one imagines what behavior is required or expected on the job and to respond accordingly. It has been argued that people with higher levels of interpersonal skills are better able at interpreting situational cues and using this information to choose the adequate response or type of behavior (Roulin et al., 2016). It therefore seems reasonable to expect EI-related social competences – i.e., knowledge of what is

expected in social situations and the ability to act accordingly – to affect one's chances of providing the desirable responses during a selection situation.

The relation between EI and faking ability can conveniently be integrated within the existing theoretical models of faking behavior. Snell et al. (1999) and Ellingson and McFarland (2011) explicitly mentioned the influence of EI on faking but did not test it. Others have suggested a link between "analytical and behavioral skills" (Marcus, 2009) and "interpersonal skills" (Roulin et al., 2016), and faking ability. These traits or skills refer to social competences which can be linked to EI. McFarland and Ryan (2000, 2006) proposed that knowledge of the measured constructs should be positively related to faking ability. This knowledge is related to EI in the sense that personality test items are assumed to serve as cues, which some individuals are better able to decipher: we hypothesize that higher levels of emotional and social knowledge will make it easier to grasp what an item is trying to convey. Finally, Goffin and Boyd (2009) posited a link between social astuteness, social intelligence and social skills, and faking ability. They constructed a decision tree that summarizes the faking process at the item level in six consecutive steps: given their presumed higher level of social skills and knowledge, we can expect high-EI (versus low-EI) individuals to be better able to identify whether a certain response is job-relevant (step 2) and whether the trait measured by the item would be (dis)advantageous on the job (step 4; Goffin & Boyd, 2009, Fig. 2, p 157).

Research on other conceptually similar constructs provides initial evidence for this claim. For example, individuals' ability to correctly perceive and interpret assessed performance criteria in selection procedures positively influences their performance herein (Holden & Jackson, 1981; Kleinmann et al., 2011; Raymark & Tafero, 2009). However, this ability is by definition context-specific (Kleinmann et al., 2011; Melchers et al. 2009) and may be a more specific derivative of one's more general level of EI. By a better understanding of the consequences of given actions in general (Sevdalis et al., 2007), high-EI individuals can be expected to know what behaviors are required across various contexts, for example at home, in school or at work, with selection procedures being one of such contexts.

Although previous studies have examined the *fakability* of EI measures (Day & Carroll, 2008; Tett, Freund, Christiansen, & Coaster, 2012; Whitman, Van Rooy, Viswesvaran, & Alonso, 2008) and the consequences of faking on their validity (e.g. Choi, Kluemper, & Sauley, 2011), these studies did not yet take notice of EI as one of the potential factors that can actually *influence* faking behavior. After all, "high emotional intelligence may give individuals greater insight into the nuances or subtle ramifications of a given behavior, and this may allow them to choose more appropriate responses for a given situation." (Snell et al., 1999; p. 223). In light of the above, it can be expected that individuals high on trait EI are better able to fake on personality tests and to provide the desired responses that increase their chances of obtaining the desired job.

Hypothesis 1. Trait EI is positively related to the ability to fake on personality tests in selection procedures.

Our second hypothesis concerns the incremental validity of emotional intelligence over and above cognitive abilities and personality in predicting the ability to fake. General intelligence or general mental ability (GMA) has previously been found to be related to faking ability and impression management (Huffcutt, Roth, & McDaniel, 1996; Mersman & Shultz, 1998; Pauls & Crost, 2005; Vasilopoulos & Cucina, 2006). Adequate faking may be more cognitively demanding than responding honestly (Van Hooft & Born, 2012) and therefore require a prerequisite level of intelligence (Bing, Whanger, Davison, & VanHook, 2004); this notion is supported by studies finding personality scales to become more *g*-loaded under faking instructions (Bing et al., 2004; Mersman & Shultz, 1998) or when a more fake-resistant forced-choice item format is used (e.g., Christiansen et al., 2005; Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006). Furthermore, GMA may be instrumental in one's test-taking strategy (Snell et al., 1999) and in correctly understanding what is meant and measured by the items in a personality test (Pauls & Crost, 2005; Tett et al., 2012). Yet, for reasons outlined above, we expect trait EI to facilitate in assessing whether agreeing or disagreeing with an item will increase the chances of getting the desired job. Thus, although GMA as well as trait EI should be important in the explanation of the ability to fake, their effects should not fully overlap.

Whether and how GMA and EI influence one's ability to fake is related to the debate in the literature on how distinctive social effectiveness constructs and GMA are (Ferris et al., 2002; Ferris et al., 2001; Melchers et al. 2009). Conceptualized as a trait, the relations between EI and GMA appear to be weak (Petrides, Furnham, & Mavroveli, 2007a). Thus, it can be expected that trait EI will account for variance in the ability to fake when controlling for the effects of GMA.

In addition, trait EI has often been criticized for showing too much overlap with 'traditional' personality taxonomies such as the Big Five (e.g. MacCann et al., 2003; Schulte et al., 2004) and a lack of incremental validity beyond such taxonomies. However, these accounts fail to realize that the overlap with personality factors such as the Big Five is in line with the conceptualization of trait EI as partly interwoven with personality traits (Petrides, 2011; Van der Linden et al., 2017). As such, strong correlations between trait EI and the personality traits that define it can and *should* in fact be expected (Petrides, Pérez-González, & Furnham, 2007b). Despite this, several studies have shown that trait EI actually shows incremental validity over and above the Big Five in the prediction of, for example, life satisfaction, coping styles and stress (Andrei, Siegling, Aloe, Baldaro, & Petrides, 2016; Petrides et al., 2007b) and most importantly for the current study, job performance (O'Boyle et al., 2011). Based on this, trait EI can be assumed to entail affect-related variance – e.g. taking the perspective of others, social awareness – that is perhaps not directly encapsulated in existing personality taxonomies such as the Five Factor Model; variance that is positively related to the ability to fake.

Based on the previous discussion on the incremental validity of trait emotional intelligence our second hypothesis states:

Hypothesis 2. Trait EI will explain variance in the ability to fake on personality tests in selection procedures over and above the variance explained by general mental ability and the Big Five personality traits.

The Present Study

The present study builds on the previous work of Raymark and Tafero (2009) who assessed the effects of individual differences on the ability to fake. In their between-subject design, participants were either instructed to 'fake good' – i.e., try to leave the best possible impression – or to respond so that it would assure them of getting them the position of an accountant. They found that the produced personality profiles differed considerably between the two conditions. Although they were not the first to discover that differential instructional sets result in different personality profiles (Furnham, 1990; Mahar, Colognon, & Duck, 1995; Martin, Bowen, & Hunt, 2002; Pauls & Crost, 2005), they concluded: "Further research is needed, examining a number of different jobs, to determine whether individual differences in ability to fake toward a job are stable across jobs." (p. 101). The present study directly addresses this question by using a within-subject design to investigate faking towards two very distinctive jobs and the role of trait EI in doing so. In addition, the incremental validity of trait EI over personality and GMA is assessed. Finding effects of EI on the ability to fake over and above the effects of personality and GMA would further emphasize the relevance of trait EI as a factor in the explanation of faking ability.

Method

Sample and Design

Participants were undergraduate psychology students who participated for course credits. Of the total of 139 respondents, 10 participants responded incorrectly to the manipulation check (see *Procedure* section) and were excluded from the analyses. This resulted in a final sample of 129 participants (105 female = 81.4%), with a mean age of 20.39 years (SD = 3.64) ranging from 17 to 42 years old.

A key part of the study were two scenarios in which the participants were instructed to respond to a personality test in such a way that they would have the highest probability of getting the job, even though their answers to the items may not necessarily reflect how they really are. Henceforward, we refer to these two scenarios as the 'faking scenarios'. For these scenarios, we largely followed the procedures as described by Raymark and Tafero (2009).

Faking scenarios

Analogous to Raymark and Tafero (2009), we extracted the job profiles for our scenarios from a study by Raymark, Shilobod, and Steffensmeier (2004). In their study, job profiles – i.e. required score

patterns – were created after consulting test manuals and prior studies, and asking 30 experienced I/O-psychologists and 148 undergraduate students to judge which personality facets of the NEO Personality Inventory–Revised (NEO–PI–R; Costa & McCrae, 1992) would be positively or negatively related to performance in a variety of jobs.

The profiles were constructed at the facet level, because some facets *within* higher-order personality dimensions can be differentially related to different types of jobs (Raymark & Tafero, 2009). Therefore, focusing on facets captures subtle differences between jobs that would not be captured at the factor level. Moreover, previous studies have shown individuals to fake discriminatively, i.e. increasing scores on some facets (or even items) and decreasing them on others (Donovan et al., 2014; Goffin & Boyd, 2009; Scherbaum, Sabet, Kern, & Agnello, 2013).

Table 1 shows the personality profiles of the jobs used in the scenarios. From the Raymark et al. (2004) study, we selected the lawyer and file clerk positions because these two profiles included the most facets for which the relations with performance were hypothesized to be in opposing directions. For example, the NEO-PI-R facet gregariousness is hypothesized to be positively related to performance in the lawyer position, while it is hypothesized to be negatively related to performance in the file clerk position. For compliance, a facet of agreeableness, on the other hand, the reverse pattern was expected.

We also added facets that were positively related to performance in one job, and unrelated to performance in the other (e.g. trust, see Table 1). This allowed us to test whether trait EI was particularly related to picking out the job-relevant traits and faking accordingly. Finally, we included facets for which the relation with performance was the same in both jobs (e.g. orderliness) for comparison purposes and in order to further complicate identification of the job-relevant traits. In total, the profiles consisted of thirteen facets: three conscientiousness facets, four extraversion facets, two openness facets, three agreeableness facets and one emotional stability facet (see Table 1).

The construction of such mixed profiles complicates producing the correct profiles (Raymark & Tafero, 2009). Rather than just scoring high on all facets (i.e., the simple heuristic "higher is better"), participants will need to carefully consider whether responding positively or negatively to a certain item will produce the correct profile. This will result in a more robust test for the effect of EI on faking behavior (Raymark & Tafero, 2009).

Measures

Personality test

Personality assessment was done with an adapted version of Johnson's 120-item NEO-PI-R consisting of items from the *International Personality Item Pool* (IPIP; Goldberg et al., 2006). The IPIP is a public domain collection of items for the use of personality questionnaires. Johnson (2011) created a short 120-item questionnaire to reflect the facets and factors of the NEO-PI-R, with sufficient reliability and validity properties.

For most of Johnson's 120 IPIP items we found the original Dutch version of the IPIP-items as presented in Hendriks (1997). However, no Dutch versions were available for 49 of the items (40 %). If that was the case, a different item belonging to the same facet was taken from the full 300-item version for which a Dutch version *was* available. When this method did not provide an alternative item, a Dutch item was chosen from the original study by Hendriks (1997) that most closely matched the content of the English item. For 10 items (8 %), both methods did not lead to an alternative item and those ten were translated by the first author and checked by an undergraduate student: any discrepancies were discussed and translations altered when needed. One item was added to each facet to improve facet reliabilities. The same methods for the addition of items as for obtaining Dutch versions of items were used, as discussed above.

Table 1. Job profiles for faking scenarios.

File clerk	
Positive facets	Cooperation (A), Orderliness (C), Dutifulness (C), Self-discipline (C)
Negative facets	Gregariousness (E), Assertiveness (E), Activity (E), Excitement seeking (E), Adventurousness (O)
Neutral facets	Trust (A), Morality (A), Lack of Self-consciousness (ES), Emotionality (O)
Lawyer	
Positive facets	Orderliness (C), Self-discipline (C), Gregariousness (E), Assertiveness (E), Activity (E), Lack of Self-consciousness (ES)
Negative facets	Trust (A), Morality (A), Cooperation (A)
Neutral facets	Dutifulness (C), Excitement seeking (E), Emotionality (O), Adventurousness (O)

Note. Self-consciousness is keyed in the positive direction.
O = Openness to experience; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability.
Positive and negative facets based on Raymark, Shilobod, and Steffensmeier (2004).

The final personality test thus included 150 (120 + (6 x 5 x 1)) items, in a five-point Likert scale format. An example item of the gregariousness facet is "*Starts conversations*", an example item of the dutifulness facet is "*Keeps his/her promises*". Reliability coefficients for the facets ranged from .39 (immoderation) to .80 (self-discipline), with an average of .68. Although four values were low (< .60), these specific facets (immoderation, intellect, liberalism and sympathy) were all facets that

were not relevant for the faking scenarios (see Table 2). Reliability coefficients for the Big Five dimensions were adequate to good in the current sample (.86, .88, .76, .84, and .87 for emotional stability, extraversion, openness, agreeableness and conscientiousness, respectively).

Trait emotional intelligence

The short version of the *Trait Emotional Intelligence Questionnaire* (TEIQue; Petrides & Furnham, 2006) was used which currently is the most well-know and widely used measure of trait EI. It has the advantage over other self-report EI measures that its development is explicitly based on a theoretical framework (Petrides, 2011).

The TEIQue-SF consists of 30 items and is designed to measure global trait EI (comprised of four underlying factors, namely wellbeing, self-control, emotionality and sociability). Questions are in a 7-point Likert-scale format. The Dutch version has been validated in two samples (Petrides et al., 2010). Example items are "*Expressing my emotions with words is not a problem for me"* and "*I'm usually able to influence the way other people feel"*. The internal consistency of the global trait EI scale in the current sample was good (α = .88, see Table 3).

General mental ability

GMA estimates of the participants were collected independently from the current study, but in about the same time period. As part of their curriculum, students completed three tests of the Dutch intelligence test series by Drenth, Van Wieringen and Hoolwerf (2001): they consecutively completed the Verbal Analogies Test, Numerical Capacity (number series) Test, and Nonverbal Analogies (abstract reasoning) Test. All tests were completed in small groups (12 students) in silence, under the supervision of a test assistant. Students were given 40, 30 and 20 minutes for the verbal, number series and abstract reasoning test respectively, as indicated in the test manual. All participants completed a consent form stating that their results could be used for research purposes. Ability test scores could not be collected for all participants; leading to a total of 102 participants for whom complete information was available.

A principal component analyses was conducted on the scores on the three ability tests. A clear single factor indicating GMA, emerged, explaining 56.56% of the variance with factor loadings of .78, .69, and .78 for the verbal, numerical, and nonverbal reasoning test, respectively. Factor scores were saved using the regression method as an indication of participants' GMA.

Ability to fake

The *ability to fake* was operationalized as the difference between the participants' score on a facet from the faking scenario and their honest score on that facet (Raymark & Tafero, 2009). This difference score was keyed positive when the change in scores from the honest to the faking scenario was in the expected direction, and keyed negative when it was in the opposite direction.

There were nine *ability to fake* scores (i.e., keyed difference scores) for each scenario, one for each facet for which a specific relation with the job position was expected (see Table 1). These difference scores within each scenario were summed, resulting in two *overall ability to fake* scores: 1) The ability to fake towards the file clerk profile and 2) the ability to fake towards the lawyer profile. Higher scores reflected a higher ability to fake.

Opportunity to fake

Participants with a lower honest score on a facet that needed to be faked upwardly to get the job (for example, cooperation which was hypothesized to be positively related with the file clerk position) have more opportunity to fake than people with a higher honest score on that facet (Goffin & Boyd, 2009). Conversely, people with higher honest scores on a facet with a negative relation with a job (for example, trust in the lawyer scenario, Table 1) have more room to fake than people with lower honest scores on this facet. Therefore, analogous to the procedure followed by Raymark and Tafero (2009), we obtained a person's *opportunity to fake* score by taking the difference between the participants' honest score on a facet and the optimal score on that facet in the faking scenarios. For the positive facets, the optimal score was the highest possible score. For the negative facets, the optimal score was the lowest possible score. Again, in total there were nine *opportunity to fake* scores for the lawyer scenario and nine for the file clerk scenario (one for each job relevant facet, see Table 1). As for the ability to fake score, within each scenario these nine scores were summed to form two *overall opportunity to fake* scores, with higher scores implying more opportunity to fake.

Procedure

All materials were administered online. Participants could complete the questionnaires (except the cognitive ability measures) at their own convenience from any computer. At the beginning of the survey and before the start of the first personality and EI measure, we provided the instructions to answer as honestly as possible to reflect the participant's true score. Thereafter, the test battery consisted of the two faking scenarios. The honest condition should therefore be seen as the 'baseline' measurement, while the faking scenarios served as a task or 'test' in the true sense – in which one could either be good or bad. Operationalized in this way, i.e. as a maximum performance test, score changes (faking) truly reflect an *ability* in the current study.

The instructions for the scenarios included a formal job description from a Dutch website similar to O*NET ("werk.nl"), including regular tasks and duties. This was done to ensure that all participants had the same background information on the jobs. Possible biasing effects of differences in job familiarity or knowledge (Goffin & Boyd, 2009; Raymark & Tafero, 2009; Snell et al., 1999) were therefore eliminated with this procedure. To create an incentive, participants were informed that for every 50 participants, the person who did best in producing the correct job profiles was selected and rewarded with 10 Euros (approximately 11 US Dollars).

After completing all questionnaires, participants received a manipulation check, consisting of an item asking them to mark the two ways in which they had responded to the personality questionnaires in the faking scenarios. The response options were: *As honest as possible, Applying for the position of file clerk, As socially desirable as possible, Applying for the position of lawyer* and *Applying for the position of accountant*. Those participants (10 in total) who marked one or more false options were excluded from the analyses. As a final check, data screening analyses as described by Meade and Craig (2012) were conducted to assess the quality of our data, which showed that there were no influential outliers or no indications of irregular data points.

Statistical Analyses

Participants' overall ability to fake, overall opportunity to fake and trait EI scores were used as input for regression analyses in order to test H1. Analyses were done separately for the file clerk and lawyer position with the dependent variable being the overall ability to fake score. The overall opportunity to fake was included in the first step, and trait EI in the second step.

To test the incremental validity of trait EI over GMA and the Big Five (H2), hierarchical regression analyses were conducted. Opportunity to fake, GMA and each of the Big Five dimensions were simultaneously entered in Step 1, and EI in Step 2. Changes in R^2 between Step 1 and Step 2 were scrutinized for evidence in support of H2. Both hypotheses were tested against a one-sided p-value of .05, given the directional nature of our hypotheses.

A few notes should be made here on issues of multicollinearity in our tests of H1 and H2. A difference score is by definition correlated with the components from which it is derived (Burns & Christiansen, 2011; Pedhazur & Schmelkin, 1991). Thus, if trait EI is highly correlated with a personality facet in the honest condition, then this shared variance between EI and the facet itself could be responsible for any relation found between EI and the difference score based on this facet. We therefore first scrutinized the strength of the relations between EI and honest facet scores before proceeding to our test of H1.

Similar multicollinearity issues pertain to our test of H2, because the Big Five scores are by definition incorporated in the opportunity to fake score (i.e., the optimal facet score minus honest facet score). However, multicollinearity is only an issue for the interpretation of the beta weights of the predictors, not for the proportions of explained variance of (a set of) predictors (Kutner, Nachtsheim, & Neter, 2004). Thus, for our test of H2, we merely looked at increments in explained variance rather than at the specific effects of the individual predictors on the outcome.

Note that the opportunity to fake and ability to fake scores are difference scores. Some scholars have cautioned against the use of difference scores due to issues with unreliability and statistical artefacts such as the aforementioned autocorrelations with the measures from which they are derived (Burns & Christiansen, 2011; Edwards, 1994; Edwards & Parry, 1993). In the present study, however, these issues are addressed in the following ways. First, when a Participant X Treatment interaction is expected, difference scores are appropriate to use (McFarland & Ryan,

2000; Tisak & Smith, 1994). This was true in our case, since we hypothesized differences in score changes (i.e., faking ability) between people in the faking conditions (i.e., the "treatment") based on individual differences (i.e., emotional intelligence). Second, the issue of potential lower reliabilities of difference scores is partly mitigated because we use variables based on aggregated difference scores. Third, the limitation of spurious correlations is solved by controlling for the opportunity to fake (essentially keyed honest scores). This transforms our dependent variable from a "raw" difference score into a "regression-adjusted" difference score, (Burns & Christiansen, 2011). As such, the dependent variable, faking ability, can be interpreted as the part of the scores in the faking condition that cannot be explained by the honest score, hereby identifying who changed his/her score more (or less) than expected based on their initial trait standing (Burns & Christiansen, 2011).

In addition to reporting the results based on difference scores, we included polynomial regression analyses (Edwards, 1994; Edwards & Parry, 1993). Polynomial regression averts the problems associated with difference scores, and has the ability to give insight into the (in)congruence between variables (e.g. honest-faking scores) and a third variable (e.g., EI). Guided by the work of Edwards and colleagues (Edwards, 1994, 2002; Edwards & Parry, 1993), we compared the explained variance of a linear model and a quadratic model in hierarchical regression analysis, both for the file clerk and lawyer scenario:

$$EI = b_0 + b_1FS + b_2HS + e$$
$$EI = b_0 + b_1FS + b_2HS + b_3FS^2 + b_4FSHS + b_5HS^2 + e$$

where FS = score in faking scenario and HS = score when asked to respond honestly. If the quadratic model shows a significant increase in explained variance, the coefficients of this model are used to draw and interpret a response surface representing the relation between the honest scores and faking scores simultaneously, and EI. The stationary point and the principal axes are essential elements of this response surface (Edwards & Parry, 1993), because they respectively indicate the point where the slope is 0 in all directions and the lines along which the upward/downward curvature is greatest/smallest. Prior to the polynomial regression analyses, all facet scores were rescaled to its original 5-point scale, keyed in the same positive direction and summed within each scenario. These sum scores were then mean centered to facilitate interpretation of the regression model and response surface.[5]

[5] Technically, independent variables should be scale centered prior to the analyses (Edwards, 1994; Shanock, Baran, Gentry, Pattison, & Heggestad, 2010), however, in our case this resulted in considerable problems of multicollinearity. Mean centering of the variables solved this issue (Aiken & West, 1991), without altering the substantive interpretations of the response surface.

Results

Descriptive Statistics

Table 2 shows the mean facet scores in the honest and the faking conditions. Whether the participants actually faked – and if so, in the right direction – can be inferred from the differences in mean facet scores in the honest and faking conditions. Replicating the findings by Raymark and Tafero (2009), participants were overall able to produce the required job profiles quite well. Scores on the cooperation facet, for example, were significantly elevated in the file clerk condition ($t(128) = 12.56$, $p < .001$), while significantly lowered in the lawyer condition ($t(128) = -12.71$, $p < .001$). All facets were faked in the hypothesized direction in the lawyer condition. In the file clerk condition, only activity was faked in the non-predicted direction ($t(128) = 12.03$, $p < .001$) while the change in assertiveness between the honest and faking condition was not significant ($t(128) = -.11$, $p = .92$). Across both scenarios, changes in scores were largest for positive facets, followed by the negative facets and smallest for the neutral facets (i.e. for which no clear relation with the job positions were expected), as indicated by their mean effect sizes *d* (Table 2; 1.89, 1.08, 1.06 respectively). Overall, these effect sizes differed significantly from each other ($F(2, 23) = 5.30$, $p = .01$), although the only significant differences were between the positive facets on the one hand and the neutral ($p = .03$) and negative facets ($p = .03$) on the other.

Overall, participants were more inclined to increase their scores than to lower them, a finding also reported elsewhere (e.g., Raymark et al., 2004). Perhaps it is easier to identify desirable as opposed to undesirable traits for a job or perhaps participants deemed it useful to, in case of doubt, go for the general socially desirable direction. In line with this tendency, the neutral facets were mostly faked upward, with the exception of emotionality. However, this specific finding can probably be explained by the fact that the items of the emotionality facet (e.g., 'has crying fits') were likely to reside at the lower end of the social desirability spectrum. Interestingly, scores on dutifulness, a facet of conscientiousness, were not significantly altered in the lawyer condition ($t(128) = .53$, $p = .60$). Dutifulness was assumed to be a neutral facet in the lawyer condition: although it measures aspects such as keeping promises and being truthful which can generally be expected to be favorable assets for any type of job, based on these findings it seems as if participants were not sure whether being dutiful would ensure them of getting the lawyer position.

Table 2 shows that in the faking conditions, the reliabilities of the facets generally became lower, and in some cases were rather low (e.g. .19 for activity in the lawyer condition). Facet reliabilities have generally been shown to decrease under faking conditions (McCann, 2013; Ziegler, Danay, Schölmerich, & Bühner, 2010). We elaborate on this in the *Limitations and Future Studies* section. It also has to be noted here that we created the overall ability to fake measures by summing over multiple difference scores, hereby increasing the reliability of these composite measures (Raymark & Tafero, 2009).

Table 3 shows the means, standard deviations and correlations between all variables. Coefficient alphas, where applicable, are shown on the diagonal.

The correlation between EI and GMA was positive but not significant ($r = .14$), in line with their respective conceptualizations as a trait and an ability. As noted previously, of special interest from a multicollinearity perspective are the correlations between EI and the facets in the honest condition, since these facets were used to create our dependent variable. The maximum correlation between EI and the facets in the honest condition was .51 (with the lack of self-consciousness facet), while most correlations were much lower ($|\bar{r}| = .26$). Thus, although trait EI and the facets in the honest condition showed some overlap, their shared variance did not appear to be problematic for our test of H1.

Initially, there was a non-significant relation between the faking abilities on the two jobs. However, this appeared to be due to differences in the opportunity to fake (i.e. the possibility to go up or down the scale), which often was in the opposite direction in the two jobs (see Table 1). After controlling for opportunity to fake their respective opportunity to fake scores, the correlation between the faking ability scores in the two jobs became $r = .40$ ($p < .001$). This shows two things. First, it confirms that the participants showed some level of stability in the extent to which they faked. Second, it underlines the effect of the opportunity to fake and the need to control for this when testing hypotheses (Raymark & Tafero, 2009).

Hypothesis Testing

EI and faking ability (H1)

Controlling for the opportunity to fake, trait EI was significantly related to the *overall ability to fake* (i.e., the sum of the job-related difference scores) for both job positions (file clerk: $\beta = .13$, $p = .02$; lawyer: $\beta = .15$, $p = .01$; see Table 4, Model 1b). The beta values found for the clerk and the lawyer did not significantly differ from each other ($\Delta(b_{\text{laywer}} - b_{\text{file clerk}}) = -.28$, $p = .78$), indicating a similar influence of EI on the level of faking in both scenarios. This supports the predicted positive relation between trait EI and faking ability (Hypothesis 1). GMA related to *overall ability to fake* in the expected direction but the effect did not reach significance (Table 4, Model 2: $\beta_{\text{file clerk}} = .08$, $p = .13$; $\beta_{\text{lawyer}} = .09$, $p = .09$).[6]

[6] Analyses including GMA were performed on a subsample (79%) of the total sample (see *Measures* section). The relation between EI and the overall *ability to fake* scores controlling for their respective *opportunity to fake* scores did not change when the subsample was used instead of the total sample. Furthermore, the excluded participants did not differ from the subsample in terms of their opportunity adjusted overall faking ability scores (file clerk: $t(127) = -.48$, $p = .63$; lawyer: $t(127) = 1.28$, $p = .20$) and level of TEI ($t(127) = -.47$, $p = .64$).

Table 2. Mean facet scores in honest condition and faking scenarios.

	Honest			File clerk								Lawyer							
Facet	α	*M*	*SD*	*M*	*SD*	Δ*M*	*t*	*d*	H	α	r_{dd}	*M*	*SD*	Δ*M*	*t*	*d*	H	α	r_{dd}
Trust (A)	.71	16.96	3.00	18.43	3.04	1.47	-4.22	-0.48	NH	.76	-	13.04	3.46	-3.92	10.54	1.21	-	.72	.66
Morality (A)	.72	19.32	2.72	22.38	2.19	3.06	-10.63	-1.24	NH	.68	-	15.37	3.60	-3.95	10.79	1.23	-	.66	.63
Cooperation (A)	.63	17.53	2.83	20.81	2.37	3.28	-12.56	-1.25	+	.51	.36	12.93	3.38	-4.60	12.71	1.47	-	.71	.63
Orderliness (C)	.77	16.41	3.54	23.98	1.72	7.57	-23.12	-2.68	+	.79	.75	20.01	3.10	3.60	-8.65	-1.08	+	.73	.75
Dutifulness (C)	.64	19.88	2.37	23.36	2.05	3.48	-13.45	-1.57	+	.78	.66	20.07	3.51	**0.19**	-0.53	-0.06	NH	.76	-
Self-discipline (C)	.80	16.25	3.34	23.05	2.09	6.80	-19.49	-2.44	+	.63	.75	23.36	2.07	7.12	-20.06	-2.57	+	.77	.80
Gregariousness (E)	.68	17.25	3.21	13.33	3.20	-3.92	9.73	1.22	-	.70	.69	19.28	2.11	2.03	-6.13	-0.75	+	.29	.54
Assertiveness (E)	.67	15.49	3.00	15.45	3.10	**-0.04**	0.11	0.01	-	.70	.60	21.28	2.33	5.79	-16.71	-2.16	+	.47	.63
Activity (E)	.66	15.58	2.72	19.39	2.56	3.81	-12.03	-1.44	-	.50	.55	19.82	2.20	4.24	-13.87	-1.72	+	.19	.47
Excitement seeking (E)	.78	15.93	3.41	11.81	3.08	-4.12	9.83	1.27	-	.70	.76	19.89	2.70	3.96	-11.53	-1.28	NH	.63	-
Self-consciousness (ES)	.66	16.12	3.06	18.30	3.83	2.18	-5.18	-0.63	NH	.79	-	23.09	2.11	6.97	-21.35	-2.65	+	.70	.67
Emotionality (O)	.78	14.11	3.64	9.26	2.57	-4.85	12.57	1.54	NH	.67	-	8.22	2.06	-5.89	15.68	2.00	NH	.46	-
Adventurousness (O)	.61	14.76	2.70	12.52	3.05	-2.24	5.81	0.78	-	.56	.64	17.93	2.28	3.17	-10.53	-1.27	NH	.42	-

Note. All mean differences are significant ($p < .01$), except those in bold. Possible range between 5-25.

Reliabilities of the difference scores were calculated using the formula: $r_{dd} = (\sigma^2_d - \sigma^2_{ed})/ \sigma^2_d$ where $\sigma^2_{ed} = \sigma^2_h(1-r_{hh}) + \sigma^2_f(1-r_{ff})$ and *h* refers to the honest condition and *f* to the faking condition (McFarland & Ryan, 2006).

O = Openness to experience; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability.

H = hypothesized relationship; + = positive; - = negative; NH = no hypothesized relationship.

Polynomial regression analyses (H1)

In the file clerk condition, the addition of the quadratic and interaction terms (Step 2) did not result in a significant increase of explained variance of EI (Table 5). Thus, in the file clerk condition, faking was best described by a linear model, indicating a significant main effect of EI on faking ($\beta = .17$, $p = .04$) independent of honest scores. This finding further supports Hypothesis 1.

In the lawyer condition, the quadratic and interaction term in Step 2 accounted for a significant portion of additional variance in trait EI beyond the linear terms (Step 1). In this model, the coefficients for the linear effect of the faking score and the linear and quadratic effects of honest score were significant (Table 5). The surface plot corresponding to the full model showed a complex saddle-shaped surface. The stationary point was found at X = -.25 and Y = .51. The 1st principal axis ran along the line Y = .47 - .14X, while the 2nd principal axis was found well outside the range of the honest and faking scores at Y = 2.24 + 7.02X.

The response surface indicated that, in line with H1, at a given honest score, higher faked scores were generally associated with higher EI levels. However, this was less so the case when honest scores were high: EI was more strongly associated with higher faked scores at lower honest scores, while at high honest scores, the surface was essentially flat.

Incremental validity of EI over GMA and Big Five personality traits (H2)

EI showed incremental validity over the effects of GMA and personality in the prediction of faking ability (Table 4). Entering EI (Model 3b), after including both GMA and the Big Five (Model 3a), resulted in a significant increase in explained variance ($\Delta R^2_{\text{file clerk}} = .03$, $p = .02$ and $\Delta R^2_{\text{lawyer}} = .04$, $p = .00$).[7] In contrast, adding the Big Five simultaneously in Step 2 after entering GMA and EI in Step 1 did not result in significant increases in explained variance in the ability to fake ($\Delta R^2_{\text{file clerk}} = .02$, $p = .56$ and $\Delta R^2_{\text{lawyer}} = .05$, $p = .08$). Thus, trait EI adds unique variance beyond personality in the explanation of the ability to fake, but the reverse could not be supported.

[7] To further address the multicollinearity issues, we also ran models including the Big Five traits but excluding the opportunity to fake scores from the analyses: results were identical.

Table 3. Means, standard deviations, reliability coefficients and intercorrelations between variables.

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
File clerk																
1. Overall ability to fake	27.64	17.46	-													
2. Overall opportunity to fake	83.93	13.42	.72**	-												
Lawyer																
3. Overall ability to fake	42.22	18.07	.01	-.25**	-											
4. Overall opportunity to fake	91.71	12.38	-.31**	-.41**	.71**	-										
5. Trait Emotional Intelligence	5.01	0.63	.35**	.32**	-.19*	-.44**	.88									
6. GMA	0.00	1.00	.29**	.30**	.07	-.03	.14	.62								
7. Openness	96.99	9.91	.33**	.30**	.02	-.12	.21*	.19	.76							
8. Conscientiousness	104.46	11.63	-.31**	-.49**	-.28**	-.42**	.28**	-.27**	-.02	.87						
9. Extraversion	102.00	12.60	.52**	.66**	-.46**	-.72**	.62**	.10	.25**	.19*	.88					
10. Agreeableness	110.52	10.51	-.31**	-.49**	.24**	.31**	.06	-.19	.06	.36**	-.03	.84				
11. Emotional Stability	103.16	11.95	.12	.19*	-.20*	-.24**	.68**	.13	-.05	.23**	.48**	.18*	.86			
12. Trust (A)	16.96	3.00	.05	-.04	.15	.24**	.14	.01	.11	.04	.12	.49**	.24**	.71		
13. Morality (A)	19.32	2.72	-.47**	-.60**	.11	.19*	.00	-.33**	-.08	.54**	-.12	.76**	.10	.13	.72	
14. Cooperation (A)	17.53	2.83	-.42**	-.60**	.32**	.46**	-.08	-.15	-.19*	.26**	-.25**	.74**	.21*	.24**	.60**	.63
15. Orderliness (C)	16.41	3.54	-.38**	-.62**	-.11	-.24**	.04	-.25*	-.11	.74**	-.14	.26**	-.02	.01	.38**	.23*
16. Dutifulness (C)	19.88	2.37	-.24**	-.49**	-.08	-.15	.25**	-.21*	-.02	.76**	.08	.45**	.21*	.15	.60**	.30**
17. Self-discipline (C)	16.25	3.34	-.23*	-.31**	-.42**	-.49**	.21*	-.26**	.01	.80**	.32**	.37**	.16	.09	.48**	.22*
18. Gregariousness (C)	17.25	3.21	.33**	.39**	-.37**	-.54**	.40**	.02	.08	.12	.70**	.09	.33**	.14	.02	-.07
19. Assertiveness (E)	15.49	3.00	.36**	.59**	-.55**	-.73**	.45**	.15	.13	.16	.72**	-.28**	.36**	-.09	-.20*	-.41**
20. Activity (E)	15.58	2.72	.19*	.24**	-.45**	-.61**	.23**	-.09	.29**	.41**	.54**	.10	.03	.09	.12	-.14
21. Excitement seeking (E)	15.93	3.41	.60**	.69**	-.05	-.24**	.36**	.22*	.24**	-.22*	.57**	-.26**	.29**	-.05	-.36**	-.30**
22. Self-consciousness (ES)	16.12	3.06	.33**	.39**	-.41**	-.58**	.51**	.22*	.02	.10	.61**	-.04	.62**	.10	-.08	-.06
23. Emotionality (O)	14.11	3.64	-.03	-.12	.04	.02	-.27**	-.07	.47**	.03	-.11	.07	-.57**	-.09	.14	-.11
24. Adventurousness (O)	14.76	2.70	.43**	.48**	-.08	-.25**	.38**	.08	.32**	-.01	.41**	-.04	.34**	.19*	-.20*	-.07

(continued)

Table 3 (continued).

		15	16	17	18	19	20	21	22	23	24
	File clerk										
1.	Overall ability to fake										
2.	Overall opportunity to fake										
	Lawyer										
3.	Overall ability to fake										
4.	Overall opportunity to fake										
5.	Trait Emotional Intelligence										
6.	GMA										
7.	Openness										
8.	Conscientiousness										
9.	Extraversion										
10.	Agreeableness										
11.	Emotional Stability										
12.	Trust (A)										
13.	Morality (A)										
14.	Cooperation (A)										
15.	Orderliness (C)	.77									
16.	Dutifulness (C)	.52**	.64								
17.	Self-discipline (C)	.46**	.50**	.80							
18.	Gregariousness (C)	-.03	.14	.26**	.68						
19.	Assertiveness (E)	-.16	.00	.25**	.29**	.67					
20.	Activity (E)	.14	.28**	.52**	.24**	.43**	.66				
21.	Excitement seeking (E)	-.32**	-.26**	-.14	.16	.42**	.05	.78			
22.	Self-consciousness (ES)	-.08	.05	.12	.49**	.57**	.24**	.22*	.66		
23.	Emotionality (O)	.06	.03	.04	-.08	-.05	.18*	-.23**	-.28**	.78	
24.	Adventurousness (O)	-.04	-.05	.05	.30**	.15	.29**	.35**	.23**	-.28**	.61

Note. * $p < .05$. ** $p < .01$. Reliabilities on the diagonal, where applicable. O = Openness to experience; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability.

Additional Analyses: Trait EI and Faking on the Facet Level

In addition to the overall ability to fake, the effect of EI on the ability to fake on each of the specific facets separately was investigated. Due to the relatively moderate reliabilities of the difference scores (Mdn_{lawyer} = .63 and $Mdn_{\text{file clerk}}$ = .66), however, these results should be interpreted with due caution.

Significant relations between EI and faking on the specific facets are reported in Table 6. For most individual facets, no significant effects of trait EI on score changes were found. For the relations that were significant, EI appeared to influence the ability to fake on lack of self-consciousness, morality and cooperation in the lawyer condition, and gregariousness, excitement seeking and cooperation in the file clerk condition. The strengths of the relations between EI and faking ability were roughly equal across facets ($.12 < \beta < .18$).

There are three important things to note about these results. First, EI was both related to faking on facets that were (hypothesized to be) faked upwards and downwards. In other words, EI was not merely related to 'faking good' (that is, faking towards the socially desirable end of the scale) but also related to one's ability to 'fake bad'. Second, EI had an effect on faking on the cooperation facet, which was faked in opposing directions in the two scenarios as expected (see Table 2). And third, EI was related to faking on job-relevant traits *only*, as no relations were found between EI and changes in scores on facets for which no relationship were proposed (Table 1). For example, morality was expected to be faked downwardly in the lawyer position, while no relation was hypothesized in the file clerk position. While scores were altered in both scenarios (Table 2), only a significant effect of EI was found in the lawyer condition. Again, however, no firm conclusions should be derived from these results.

Table 4. Summary of hierarchical regression analyses for variables predicting the ability to fake.

		File clerk							Lawyer						
Model	Independent variables	*b*	*S.E.*	β	$r_{xy.z}$	R^2	ΔR^2	ΔF	*b*	*S.E.*	β	$r_{xy.z}$	R^2	ΔR^2	ΔF
1a	Overall opportunity to fake	.94	.08	.72**	.72	.52**		137.03**	1.03	.09	.71**	.71	.50**		125.24**
1b	Overall opportunity to fake	.88	.08	.68**	.69	.53**	.01*	4.12*	1.12	.10	.77**	.71	.52**	.02*	4.86*
	TEI	3.61	1.78	.13*	.18				4.37	1.98	.15*	.19			
2	Overall opportunity to fake	.92	.10	.69**	.69	.52**		52.54**	1.03	.10	.72**	.72	.52**		52.71**
	GMA	1.39	1.25	.08	.11				1.66	1.24	.09	.13			
3a	Overall opportunity to fake	.88	.31	.66**	.28	.52**		14.75**	1.26	.24	.88**	.48	.54**		15.53**
	GMA	1.56	1.36	.09	.12				2.35	1.39	.13*	.17			
	Openness	.10	.15	.06	.07				.07	.14	.04	.05			
	Conscientiousness	.08	.23	.06	.04				.21	.17	.14	.13			
	Extraversion	.05	.25	.04	.02				.22	.20	.16	.11			
	Agreeableness	-.02	.18	-.01	-.01				.02	.17	.01	.01			
	Emotional Stability	-.05	.13	-.03	-.04				-.19	.13	-.13	-.14			
3b	Overall opportunity to fake	.87	.30	.66**	.29	.55**	.03**	5.88**	1.29	.23	.89**	.50	.58**	.04**	9.74**
	GMA	1.42	1.32	.08	.11				2.19	1.33	.12†	.17			
	Openness	.03	.15	.02	.02				-.02	.14	-.01	-.02			
	Conscientiousness	.02	.22	.02	.01				.15	.16	.10	.10			
	Extraversion	-.04	.25	-.03	-.02				.11	.20	.08	.06			
	Agreeableness	.02	.17	.01	.01				.07	.16	.04	.05			
	Emotional Stability	-.25	.15	-.18†	-.17				-.46	.15	-.31**	-.29			
	TEI	7.47	3.08	.27**	.24				9.65	3.09	.34**	.31			

Note. † $p = .05$. * $p < .05$. ** $p < .01$. All one-sided.

TEI = Trait Emotional Intelligence; $r_{xy.z}$ = partial correlation controlling for other variables in the model.

The beta coefficients for TEI represent the effect of EI on faking ability independent of faking opportunity; however, faking opportunity (i.e. honest scores) can still, through its overlap with EI, affect the relation between EI and faking ability. The partial correlation between EI and faking ability, controlling for faking opportunity, thus captures the 'purest' effect of EI on faking ability.

Table 5. Results from polynomial regression analyses predicting trait emotional intelligence.

	File clerk						Lawyer					
Model	*b*	*S.E.*	*β*	R^2	ΔR^2	ΔF	*b*	*S.E.*	*β*	R^2	ΔR^2	ΔF
Step 1:				.13**		9.60**				.23**		18.31**
Faking score	.39	.19	.17*				.38	.17	.17*			
Honest score	-.71	.18	-.33**				1.02	.18	.45**			
Step 2:				.14**	.01	.39				.28**	.05*	2.83*
Faking score	.41	.20	.18*				.49	.18	.22**			
Honest score	-.73	.18	-.34**				1.06	.18	.47**			
Faking score squared	.23	.45	.04				.50	.38	.11			
Honest score squared	-.39	.44	-.08				-1.16	.48	-.19*			
Faking x Honest score	.14	.68	.02				-.48	.64	-.06			

Note. * $p < .05$. ** $p < .01$. All two-sided.

Table 6. Summary of hierarchical regression analyses for variables predicting the ability to fake on the facet level.

		File clerk							Lawyer						
Dependent variable	Independent variable	*b*	*S.E.*	β	$r_{xy.z}$	R^2	ΔR^2	*F*	*b*	*S.E.*	β	$r_{xy.z}$	R^2	ΔR^2	*F*
Ability to fake:															
Self-consciousness	OTF – Self-consciousness								1.08	.07	.89**	.81	.69	.01*	139.28**
	TEI								.78	.34	.13*	.20			
Morality	OTF – Morality								.79	.11	.52**	.53	.30	.03*	26.96**
	TEI								1.16	.49	.18**	.21			
Cooperation	OTF – Cooperation	.69	.07	.66**	.67	.46	.02*	53.73**	.86	.11	.59**	.59	.35	.01*	34.35**
	TEI	.60	.31	.13*	.17				.78	.47	.12†	.15			
Gregariousness	OTF – Gregariousness	.95	.10	.66**	.66	.52	.01*	69.20**							
	TEI	.92	.49	.13*	.17										
Excitement Seeking	OTF – Excitement Seeking	1.01	.09	.72**	.73	.59	.01*	92.34**							
	TEI	.84	.46	.11*	.16										

Note. † $p = .05$. * $p < .05$. ** $p < .01$. All one-sided.

TEI = Trait Emotional Intelligence; OTF = Opportunity to fake; $r_{xy.z}$ = partial correlation controlling for other variable in the model.

Discussion

The present paper showed that trait EI was related to the ability to fake on personality tests, and that the effects were similar when asked to fake towards two jobs with very divergent profiles. That is, when the job indicated that higher scores on certain traits were better, high-EI individuals shifted their scores more strongly in that direction than low-EI individuals. Similarly, when a job indicated that lower scores on certain traits might be better, higher EI was associated with stronger effects in that direction. The nature of the effects and the effect sizes were roughly in line with those found for other traits known to be related to the ability to fake, such as openness (Raymark & Tafero, 2009), integrity (McFarland & Ryan, 2000) and efficacy of self-presentation (Pauls & Crost, 2005). In addition, trait EI showed incremental validity over GMA and the Big Five personality dimensions in its relation with the ability to fake. Finally, although not the main focus of this study, in the present sample no significant effect of GMA on the ability to fake was found. Because the participants were all university students (and from the same study, Psychology) there may have been a restriction of range in intelligence, reducing relationships with criteria (Hunter & Schmidt, 2004). Using the standard deviations of the population from the test manual, we tested whether this was true. Corrected for range restriction, the correlation between GMA and faking ability (independent of faking opportunity) increased from .11 to .14 ($p = .08$, one-sided) in the file clerk condition and from .13 to .17 ($p = .04$, one-sided) in the lawyer condition. Thus, homogeneity of the sample may have partly been responsible for the non-significant relation between GMA and faking ability.

Importantly, our main findings were robust across methods of analyses: the regression-based results were confirmed in polynomial regression analyses, a sophisticated technique gaining popularity in faking research (e.g., Peterson et al., 2011). Results from the lawyer condition implied that the effect of EI on faking was more pertinent when honest scores were low compared to when honest scores were high. A possible theoretical explanation could be that with high honest scores, faking ability becomes less relevant, because there may be less *need* to fake in order to leave a good impression (Goffin & Boyd, 2009). Put differently, when honest scores are low and self-presentation is required to get the job, we can expect individual differences in EI-related competences to become more pronounced. However, there could also be a methodological explanation: the lawyer scenario included three extraversion facets – with a relatively strong *positive* relation to EI – which needed to be faked upward. Thus, those with higher honest scores on those facets are also more likely to have higher EI scores, reducing the likelihood of finding a positive effect of EI on faking for this particular group of individuals. Given that the findings did not generalize across scenarios, this explanation appears to be likely. Nonetheless, the results do emphasize that honest scores are an important factor in faking behavior and therefore should not be overlooked (Goffin & Boyd, 2009; Marcus, 2009; McFarland & Ryan, 2000).

Overall, the present study has implications for both the EI literature as well as the faking literature, and connects these two fields theoretically. First, EI appeared to be a relevant factor in the ability to fake across jobs. We found that individuals who can fake towards the profile of one job are also able to fake towards the other to a considerable degree ($r = .40$ with respective opportunities to fake controlled). This result supports previous findings implying that some

people are consistently better at faking or impression management than others (Ingold et al., 2015; Klehe et al., 2014; McFarland & Ryan, 2000). The present study provided initial evidence that emotional intelligence can be partly responsible for such individual differences in faking ability. In the sense that trait EI reflects dispositions related to genuine social competence or social skills that can be expected to have positive effects on the job, practitioners might be less worried about the effects of EI on faking personality tests in terms of their the predictive validity. Put differently, practitioners trying to safeguard personality tests from faking may be throwing the baby out with the bathwater; by doing so, valid trait variance related to self-presentational skills they might actually seek in employees will be removed. This is not improbable, given that trait EI has been shown to be related to job performance (Côté & Miners, 2006; Joseph et al., 2015; Joseph & Newman, 2010). For example, in a large meta-analysis, O'Boyle et al. (2011) showed that trait EI has criterion validity for job performance of around .26. Moreover, trait EI measures showed an incremental validity of 5.2% ($p < .05$) above and beyond intelligence and personality. Obviously, this study provides only one piece of the puzzle: One would need information on both performance in selection procedures *and* on the job in order to arrive at more firm conclusions about this dual effect of EI. Promising evidence comes from research that has shown that the ability to identify selection criteria, a concept similar to EI, is positively related to both performance in selection procedures and on the job (Kleinmann et al., 2011).

The findings of the present study are in line with impression management (Hogan et al., 2007) or self-presentational (Hogan, 1982; Marcus, 2009) theories of faking, which state that people are *always* concerned with the impression they make on others in social interactions and in this regard, responding to a personality test in selection procedures is no exception. Imperative here is the argument that individuals with higher levels of EI – indicative of one's general social effectiveness – would engage and be more effective in impression management not only during a selection procedure but also in their everyday (social) life. This could be the reason why high-EI (versus low) individuals are rated as being more prosocial by others (Petrides, 2011).

Finally, our study has implications for the faking literature at large. Scores on one and the same facet were raised in one scenario, while lowered in the other, in line with the job profiles (Raymark et al., 2004). This implies that providing a higher score on a facet does not need to increase one's chances of getting the job; whether this will be the case will depend on the job one is applying for. In fact, elevating a score on one and the same facet could increase one's chances of getting hired in one job, but decrease them in the other. In addition, increasing scores on non-relevant traits would not increase someone's chances of getting the job at all. Fake good (e.g., Dunnett, Koun & Barber, 1981; Mersman & Shultz, 1998) and fake-as-the-best-applicant-for-a-'desired job' (e.g., McFarland & Ryan, 2000) cannot account for these effects: such studies consider 'higher scores' to be 'better', but the present study showed that this is not necessarily the case. Therefore, we agree with Raymark and colleagues (Raymark et al., 2004; Raymark & Tafero, 2009) that in faking studies *a priori* hypotheses on which facet (or factor) scores will be altered and in which particular directions are essential in correctly estimating the effects of faking in selection procedures.

Limitations and Future Studies

When interpreting the results, the lowered reliabilities of the facets under the faking conditions warrant some consideration. They stand in contrast to reliabilities reported in fake good studies, in which consistently choosing for the most socially desirable response option generally results in increases in internal consistencies (e.g., Douglas et al., 1996). In the current study, apparently, participants scrutinized each individual item as to whether it would increase or decrease their chances of getting hired, and answered accordingly. Investigation of the items of facets with low reliabilities in the faking scenarios confirms this idea. For example, for the gregariousness facet in the lawyer condition, the item "Likes large parties" was solely responsible for the low reliability value. Interestingly, the mean score on this item was lower in the faking condition compared to the honest condition, while mean scores on the remaining four items were higher. Clearly, participants considered a preference for large parties (perhaps indicative of irresponsibility or lack of self-control) as undesirable for getting the job, while endorsing the other items (e.g. "Starts conversations") as desirable. This indeed suggests that participants very carefully selected their response to each individual item (see Donovan et al., 2014 for similar results).

Further, given our student sample, the potentially limited generalizability of the results should be taken into account. Students may lack the relevant knowledge needed when applying for a job, and are perhaps less motivated to do well in the faking scenario compared to when a real job is at stake (Mersman & Shultz, 1998). On the other hand, we presented all participants with the same job information and provided a monetary incentive, which can be assumed to at least partly deal with the aforementioned limitations. Also in relation to generalizability of findings, some have argued that instructed faking studies are limited in terms of validity because they reflect maximum rather than actual faking behavior (Smith & Ellingson, 2002). Although this might be true, the aim of the present study first and foremost was to confirm that EI is related to the ability to fake in the first place. Nevertheless, additional research with samples from the working population in real life settings is needed to confirm the present findings.

The present study has exclusively focused on the influence of EI on the *ability to fake*. Models of faking behavior include other aspects such as motivational (e.g. the need for the job) and situational factors (e.g. possible verification of test results) as well (e.g., Goffin & Boyd, 2009; McFarland & Ryan, 2006; Roulin et al., 2016). The fact that individuals with higher levels of trait EI are better *able* to fake does not necessarily mean that, in real life, they would also have an increased tendency to do so. In this sense, it would be interesting to investigate how EI and motivational and situational factors interplay in the prediction of faking behavior.

On the predictor side, the present study focused on trait EI; alternatively, future studies could investigate how faking ability relates to other social effectiveness constructs, for example the ability to identify the criteria in selection procedures (Kleinmann et al., 2011). Furthermore, although we demonstrated incremental validity of EI over GMA and the Big Five with respect to faking ability, there are other traits against which incremental validity could be tested (for example, core-self evaluations; Joseph et al., 2015; Kluemper, 2008). Finally, given the theoretical

debate on how EI should be conceptualized, the relation between faking ability and trait EI vis-a-vis ability EI should be tested empirically.

Conclusion

The present study provides new insights into the factors associated with faking. Specifically, it showed that EI positively relates to the ability to fake on personality tests across two very distinctive jobs. The findings from the present study may provide some reassurance for those who tend to be pessimistic on the use of personality tests: yes, personality tests can be faked, but it might be that successfully doing so requires the same skills and knowledge that also lead to a genuinely better performance on the job. This does not necessarily mean that faking will *never* pose a problem. The notion that people *can* fake all the more calls for an adequate assessment of people's traits and skills as well as the specific conditions influencing whether people actually *do* fake.

The Motivation and Opportunity for Socially Desirable Responding Does Not Alter the General Factor of Personality

5

This chapter has been submitted for publication as:
Pelt, D. H. M., Van der Linden, D., Dunkel, C. S., & Born, M. Ph. (2019). The motivation and opportunity for socially desirable responding does not alter the general factor of personality.

Abstract

Socially desirable responding (SDR) may affect the factor structure of personality questionnaires and may be one of the reasons for the common variance among personality traits. In this study, we test this hypothesis by investigating the influence of the motivational test-taking context (development vs. selection) and the opportunity to distort responses (forced-choice vs. Likert response format) on personality questionnaire scores. Data from real selection and assessment candidates (total $N = 3{,}980$) matched on gender, age, and educational level were used. Mean score differences were found between the selection and development groups, with smaller differences for the FC version. Yet, exploratory structural equation models (ESEM) showed that the overall factor structures as well as the general factor were highly similar across the four groups. Thus, although SDR may affect mean scores on personality traits, it does not appear to affect factor structures. This study further suggests that the common variance in personality questionnaires is consistent and appears to be little influenced by motivational pressures for response distortion.

Introduction

Organizations increasingly use personality questionnaires as part of their selection procedures (König, Klehe, Berchtold, & Kleinmann, 2010), because such measures can help them improve the selection of employees. Research has shown that personality test scores moderately but consistently predict job performance and organizational citizenship behavior, even on top of cognitive abilities (Schmidt & Hunter, 1998; Cook, 2016). Yet, the increased use of personality questionnaires by practitioners has also raised concerns about potential response distortion by applicants completing a personality questionnaire. Response distortion in order to maximize the chances of obtaining a desired job, rather than to answer honestly, is referred to as socially desirable responding (SDR) or faking (Mesmer-Magnus & Viswesvaran, 2006; Ones et al., 1996). A vast body of literature has been devoted to the investigation of the prevalence and magnitude of faking (Birkeland et al., 2006; Viswesvaran & Ones, 1999), its antecedents (e.g., Ellingson & McFarland, 2011; McFarland & Ryan, 2000, 2006; Roulin et al., 2016), and the predictive validities of personality questionnaires (e.g., Cook, 2016; Morgeson et al., 2007; Rosse, Stecher, Miller, & Levin, 1998).

In the present study, we focus on the effect of SDR on the construct validity of personality questionnaires, an area that has also received considerable attention in the faking literature (Ellingson et al., 2001; Ellingson, Sackett, & Hough, 1999; Joubert et al., 2015; Marshall et al., 2005; Schmit & Ryan, 1993; Smith & Ellingson, 2002). It is important for personality questionnaires to preserve their construct validity when used in motivated settings (such as selection procedures), because construct and criterion validity are interrelated (Cronbach & Meehl, 1955). One way of assessing the construct validity of a questionnaire is by investigating its factor structure. In the current study we add to the literature on social desirability by inspecting the influence of test-taking context and the item format on the factor structure of a personality inventory.

Scholars have argued and shown that SDR in selection contexts can negatively affect the construct validity of the adopted personality questionnaire. In a classic study, Schmit and Ryan (1993) showed that the standard factor structure of the NEO-FFI (Costa & McCrae, 1989) completely disappeared when administered in an applicant sample. More specifically, inflated factor intercorrelations were found in an applicant sample but not in a non-motivated sample. These authors argued that individual differences in the tendency to respond with an ideal employee in mind (i.e., SDR) uniformly added variance to allegedly independent traits, hereby increasing their overlap and hence their observed intercorrelations. Consequently, their best fitting model included a general factor, which they labeled the "ideal-employee-factor", to account for this additional common variance. This finding, i.e., inflated correlations between personality factors or scales due to a larger portion of shared variance (i.e., the emergence of a general factor) under motivated settings, has been reported a number of times (e.g., Ellingson et al., 1999; Klehe et al., 2012).

In contrast, however, other studies have shown that although the mean scores on personality traits may indeed shift, the factor structures of personality questionnaires are invariant across motivated and non-motivated groups (Ellingson et al., 2001; Marshall et al., 2005; Smith, Hanges, & Dickson, 2001; Smith & Ellingson, 2002). These latter findings suggest that, although people may lower or increase their scores somewhat depending on the context, the

rank order between individuals seems to remain rather similar. Consequently, it has been argued that social desirability introduces little if any systematic error to the measured personality characteristics (Ellingson et al., 2001; Smith & Ellingson, 2002). In sum, to date, the effect of the selection context on the construct validity of personality questionnaires remains unclear.

Interestingly, the two contrasting views outlined above are reflected in two relatively recent streams of research. The first relates to whether social desirability scales – traditionally developed to detect "liars" or "fakers" – measure social desirability as a response set (i.e., error or method artefact) or whether these scales might actually capture substantive trait variance (Lönnqvist et al., 2007). Uziel (2010a) reviewed the literature on social desirability scales and concluded that they should mainly be reinterpreted as measures of interpersonally oriented self-control. Recent meta-analytic results have confirmed this idea, showing that social desirability scales appear to largely measure trait-like tendencies, rather than only response styles (Connelly & Chang, 2016). These findings naturally lead to the question whether social desirability in itself should be reconsidered as a trait-like construct. Indeed, Klehe et al. (2012) showed how Schmit and Ryans' general ideal-employee-factor was related to job-related performance, a relation which could be explained by the ability to identify the criteria in the selection procedure, abbreviated as ATIC (Kleinmann et al., 2011), an ability akin to social skills and social effectiveness. If social desirability is really related to genuine social skills and abilities, then this might also explain why controlling for social desirability appears to have little (or even a negative) influence on the criterion validity of personality measures (e.g., Ingold et al., 2015; Schmitt & Oswald, 2006).

The second stream of research is focusing on the so-called General Factor of Personality (GFP; Figueredo et al., 2004; Musek, 2007). The GFP reflects the shared variance of lower-order traits and its nature is nearly identical to the ideal-employee factor as described above. Specifically, in terms of the Big Five model, high GFP individuals would, on average, be relatively open-minded, diligent, sociable, friendly, and emotionally stable. Studies have consistently shown that a general factor is present in personality measures, even under non-motivated circumstances, i.e., settings in which we can assume respondents to answer in a truthful way (Van der Linden et al., 2010a). Several scholars consider the GFP to reflect a substantive and meaningful construct. In this tradition, the leading current interpretation is that the GFP reflects a general social effectiveness factor (Van der Linden et al., 2016), which is supported by its large overlap with emotional intelligence (Van der Linden et al., 2017), part of which appears to be genetic (Van der Linden et al., 2018). In further support of this interpretation, the GFP has been found to be related to a wide range of outcomes from different life domains, such as popularity and likeability (Van der Linden et al., 2010b), leadership (Pelt et al., 2017), reduced delinquent behavior (Van der Linden et al., 2015) and job performance (Pelt et al., 2017).

Mirroring the debate on social desirability as described in previous paragraphs, however, there is a similar dispute around the GFP with several scholars suggesting that the GFP is not substantive, but rather an artefact arising from the way personality traits are measured (e.g., Ashton et al., 2009; Chang et al., 2012). One important explanation for the GFP is in fact the tendency to respond in socially desirable ways (Irwing, 2013): that is, the evaluative content (i.e. social desirability level) of personality items may trigger the motivation to self-enhance in some people more than in others (Bäckström et al., 2009). In this way, because people respond both

to the personality content and evaluativeness of items (Biderman et al., 2018), scales tend to become more correlated, leading to the emergence of a general factor. Related to this, Dunkel et al. (2016) tested the various sources of variance in the GFP, and concluded that although the general factor partly reflects measurement error such as social desirability bias, the lion share of the variance could be attributed to genuine social effectiveness.

Van der Linden et al. (2011) also directly tested this social desirability explanation for the GFP by comparing the general factors extracted from a selection and an assessment sample. Naturally, the level of self-enhancement motivation was expected to be higher in the selection sample compared to the assessment sample, because there was a job at stake. If the GFP was entirely due to SDR induced by the level of self-enhancement motivation, then the authors expected the GFP to be markedly different – or even disappear – in the assessment sample compared to the selection sample. However, the study showed that although participants in the selection setting indeed scored higher on all personality dimensions, suggesting a general tendency to present oneself in a more favorable way, the factor structure remained the same in both settings. Thus, the GFPs were highly similar across the two groups, which led them to the conclusion that it is unlikely that the GFP merely is the product of SDR.

Some recent studies lead to somewhat different conclusions. In line with Van der Linden et al. (2011), Anglim et al. (2017) showed that most statistical properties of the HEXACO questionnaire – including factor loadings and the size or importance of the general factor – were equal in applicant and non-applicant samples. They did show, however, that the general factor was likely to represent a faking factor as indicated by 1) significantly higher mean levels on the general factor in the applicant sample than in the non-applicant sample, with this mean difference being larger than for any of the HEXACO dimensions and 2) strong relations between item loadings on the general factor on the one hand and indicators of item social desirability (e.g., standardized item mean differences between honest-applicant conditions from a separate instructed faking study) on the other. Yet, the authors acknowledged that this study "does not resolve the ongoing debate about whether it reflects substance or bias" (p. 679); the fact that scores on a social desirability factor can be increased in selection contexts does not automatically mean that such a factor cannot have predictive value or be substantiated by other-reports (Chen et al., 2016). In addition, recently a number of studies (MacCann et al., 2017; Schermer et al., 2019a; Schermer et al., 2019b) using an experimental laboratory design in which participants are instructed to fake have shown that the general factor did become more prominent under such instructions. The contrasting findings between these experimental studies and Van der Linden et al. (2011) might be ascribable to the instructed nature of the former and the use of student samples; such studies have often been criticized for a lack of ecological validity (Smith & Ellingson, 2002). The influence of social desirability in the selection context on construct validity and the general factor in personality tests – also for recommendations to practitioners using the tests in the field – is thus best studied in real-life, naturally occurring situations.

The Forced-Choice Method

Although the effect of social desirability in the selection context on the construct validity of personality questionnaires is not clear empirically, it has mostly been viewed as a nuisance factor. As such, several attempts have been made to reduce its influence, one of them being presenting personality items in a different format. In the previous discussion, all personality questionnaires referred to were traditional Likert-type format questionnaires in which respondents were asked to indicate the extent to which they disagree/agree with a statement on a certain (mostly 5-point or 7-point) scale. With this format, especially in combination with the transparency of personality items (Kuncel & Tellegen, 2009), it may be relatively easy to deduce what the desirable response is in a selection context. In other words, Likert-type items provide a relatively large opportunity to "fake".

To reduce this opportunity, the so-called multidimensional forced-choice (FC) format was developed (Waters, 1965), where respondents are presented with two or more statements loading on different traits paired in their levels of social desirability, and forced to choose to indicate which is "most like them" (and sometimes also which is "least like them"). When responding to FC questionnaires, it is impossible to exclusively endorse socially desirable items, preventing one from leaving a favorable image across all traits. In addition, the FC format can reduce response biases related to personality measurement such as acquiescence (Cheung & Chan, 2002) and halo effects (Bartram, 2007).

Although these properties appear to be favorable compared to the Likert format, they come at a cost. That is, when FC questionnaires are scored with traditional methods based on classical test theory (Brown & Maydeu-Olivares, 2011), they yield *ipsative* scores (Hicks, 1970; Meade, 2004) and consequently measurement difficulties arise. For example, by being forced to choose dependencies in the data arise, resulting in negative correlations between traits even though, in reality, they may be unrelated or even positively related. In addition, FC questionnaires yield relative rather than absolute trait standings (Baron, 1996) influencing the obtained score profiles of the respondent; as noted, it precludes scoring high (or low) on all measured traits simultaneously. From an applied perspective, this makes comparing trait standings *across* individuals, for example in selection situations, problematic.

While these concerns are theoretically valid, empirically it has been shown that traditional sum scores based on FC questionnaires are relatively good approximations of trait scores (Lee, Lee, & Stark, 2018), especially when specific precautions are taken (e.g., Salgado, Anderson, & Tauriz, 2015). Using traditional sum scores, previous studies have shown how the FC format reduces the possibility of response distortion. For example, mean score differences between honest participants and participants instructed to fake tend to be smaller, yet not vanish, with the FC compared to the Likert format, and relations with criteria in motivated conditions appear to be less attenuated with the FC (vs. Likert) format (Christiansen et al., 2005; Jackson, Wroblewski, & Ashton, 2000; Vasilopoulos et al., 2006).

In terms of construct validity of FC questionnaires, however, much less is known (Salgado et al., 2015); only a few studies have directly compared factor structures across Likert and FC versions of the same instrument. This is presumably due to the aforementioned difficulties associated with scoring FC measures, leading to biases in factor loadings in factor analysis

(Brown & Maydeu-Olivares, 2011). However, the recently developed Thurstonian IRT model (Brown & Maydeu-Olivares, 2011, 2012) allows the extraction of normative (vs. ipsative) scores form FC inventories. Several studies have used this model to score FC questionnaires and showed that indeed normative scores (i.e., normally distributed, and allowing for score profiles with exclusively high or low scores) and good results in terms of validity can be obtained (Anguiano-Carrasco, MacCann, Geiger, Seybert, & Roberts, 2015; Guenole, Brown, & Cooper, 2016; Joubert et al., 2015; Lee, Joo, & Lee, 2019; Morillo et al., 2016; but see Dueber, Love, Toland, and Turner (2019) for remarks on limitations of the model).

Two studies adopted this model to directly test the influence of the item format on construct validity of personality inventories. Joubert et al. (2015) showed that the covariance matrices of the Likert and FC version of Occupational Personality Questionnaire (SHL, 2013) were largely equivalent. Although they compared groups assumed to have differed in their motivation to leave a good impression on the personality questionnaires (i.e., a training sample and selection samples), they excluded a group that completed the Likert version in a selection context. Also, they did not explicitly investigate the role of the general factor in the different groups. Irwing (2013) did explicitly investigate whether the GFP in the OPQ Likert and FC version where comparable, and showed this to be largely the case. In that study, however, samples were used from low-stakes settings. As such, it is unclear whether the general factor found in Likert and FC versions of the same instrument still converge when motivational pressures are present.

The Present Study

Until now, studies have predominantly focused on either the influence of context (non-motivated vs. motivated test setting) on the factor structures of personality tests, or the influence of item format (Likert vs. FC), but have rarely combined both approaches simultaneously. While the context has shown to influence the motivation to distort responses, the item format defines the opportunity to do so. Thus, by leaving out an applicant group completing a Likert-type questionnaire, Joubert et al. (2015) omitted a test against the "worst case scenario" in which people are motivated to fake and had ample chance to do so. Consequently, it is possible that the factor structure of a FC questionnaire can be replicated with a Likert questionnaire in low-stakes settings, but the question remains whether this is true when the Likert questionnaire is administered in high-stakes settings.

In addition, if there are fundamental differences in the factor structure across Likert and FC-based instruments then differences between instruments with different item formats might be found even in the absence of motivational pressures. Thus, we first need to establish the equivalence of the construct validity between the instrument types under low-stakes settings – acting as a baseline – before we can make statements about the equivalence under motivated settings.

In summary, to fully understand whether factor structures overall – and the general factor more specifically – are affected by SDR, the effects of item format and test-taking context need to be disentangled. This is achieved by a full context X item format design (Table 1). Such a

design is used in the current study containing four groups with potentially differing levels of SDR on the basis of their motivation and opportunity to distort their responses.

The main goal of the current study was to thoroughly test the influence of social desirability on the construct validity of personality tests in general – and the general factor more specifically – by comparing the factor structures across these four groups. If social desirability represents a response set related to situational pressures, then we would expect it to introduce error variance in the selection context, negatively influencing the instrument's construct validity, and thus to lead to differences in factor structures between the selection and development groups. Because it is assumed that the FC format will reduce the amount of error variance that is introduced, when social desirability is seen as a response set, the expectation would be that the factor structures of the development and selection group will be more similar for the FC version than for the Likert version. Following the same line of reasoning, if the general factor is markedly different across the four groups, then this would imply that it mostly captures situation-specific variance emanating from response distortion. Alternatively, if the general factor is highly similar across the groups, then this would point in the direction of this factor reflecting a relatively stable personality characteristic, being less susceptible to situational pressures for response distortion.

As many studies on social desirability are limited in their use of students instructed to distort their responses (e.g., Pavlov, Maydeu-Olivares, & Fairchild, 2018; Smith & Ellingson, 2002), one particular asset of the present study is that it addresses the effect of social desirability in actual, real-life settings with participants either applying for real jobs or going through an assessment for their own career development.

Table 1. Research design (2 x 2).

		Context	
		Development	Selection
Item format	Likert	Low motivation High opportunity	High motivation High opportunity
	Forced-choice	Low motivation Low opportunity	High motivation Low opportunity

Method

Sample and procedure

Data were collected from the database of a large test development and publishing firm which develops instruments specifically for the HR market. Their clients mostly consist of selection and assessment agencies, recruitment agencies, career counseling firms, individual counselors/psychologists, and HR-departments of medium to large companies. The database was searched for candidates who completed a personality questionnaire (in Likert or FC format) as part of a career development process or a selection procedure and for which background information (gender, age and education) was available. The applicants in the sample applied to

a wide range of job positions at a large number of different organizations. Data were collected between July 2011 and March 2015.

These requirements resulted in an initial sample of 9,212 candidates (FC development; *N* = 1,673, FC selection; *N* = 1,473, Likert development; *N* = 3,325; Likert selection; *N* = 2,741). To reduce the influence of differences in demographics between the groups, a matching procedure as described in Anglim et al. (2017) was conducted in order to create samples with equal distributions in terms of age, gender and education. The procedure is based on strata sampling, therefore we created 5 x 2 x 3 = 30 different strata based on respectively age (<25, 26 – 35, 36 – 45, 46 – 55, >56), gender and educational level (low, middle, high). In each group, a similar number of participants was randomly sampled from each stratum (see Anglim et al., 2017, for a detailed description of the procedure). This resulted in a final total sample of 3,980 participants, with an equal number of 995 participants in each of the four groups, and with no significant differences in age (FC development: *M* = 39.0, *SD* = 10.0, range 17-63; FC selection: *M* = 38.6, *SD* = 10.0, range 18-66; Likert development: *M* = 39.0, *SD* = 9.9, range 17-62; Likert selection: *M* = 38.6, *SD* = 10.1, range 17-62) and gender (63.4% male in all four groups). The same was the case for educational level (5.7% lower, 23.1% middle and 71.2% higher educated in all four groups).

Measures

Likert personality questionnaire

To measure personality, the Work-related Personality Inventory (WPI; Ixly, 2012) was used, which is based on the Five Factor Model (FFM). The WPI consists of 25 lower level facets, which combine into five higher order factors (see Table 2), namely 1) Stability, 2) Structure, 3) Exuberance, 4) Influence, and 5) Sociability, that overlap with the Big Five dimensions (Ixly, 2012, 2014). The Stability factor of the WPI overlaps with Emotional Stability (reversed Neuroticism) of the Big Five (*r* = .82 and *r* = .69 with the Dutch version of the *Big Five Inventory* (BFI; Denissen, Geenen, Van Aken, Gosling, & Potter, 2008) and the *Five-Factor Personality Inventory* (FFPI; Hendriks et al., 1999), respectively, as reported in the WPI manual; Ixly, 2012, 2014), Structure with Conscientiousness (*r* = .63 and *r* = .76), Sociability with Extraversion (*r* = .78 and *r* = .77). Exuberance mostly overlaps with Openness (*r* = .69 and *r* = .56), but also partly overlaps with Conscientiousness (*r* = .50 with the BFI). The Influence factor mainly overlaps with Extraversion (*r* = .57 with the BFI and FFPI), but is also moderately related to lower Agreeableness (*r* = -.46) and higher Openness of the FFPI (*r* = .40). In 2012, the WPI received a favorable rating from the COTAN, the official national test-auditing authority from the Dutch Association of Psychologists (NIP). This favorable rating implies that the committee has evaluated the empirical evidence supporting the quality of this instrument and concluded that it is reliable and a construct valid assessment of personality according to current main theories, in this case the FFM.

The Likert version of the WPI consists of 276 items, with a 5-point scale ranging from *Totally disagree* to *Totally agree*. All items are formulated in the third person; an example item of the Trust facet states "*Is quick to trust strangers*". For this version, facet scores were calculated by summing the items belonging to a facet and subsequently standardizing them to put the

scores of the two instrument types on the same metric (see below). Reliabilities of the facets were adequate to good and comparable between the development (M = .90, SD = .02, ranging between .84 and .94) and selection context (M = .88, SD = .03, ranging between .78 and .93).

Forced-choice personality questionnaire

The forced-choice version was previously constructed (not as part of the current study) based on the Likert version: items were paired in terms of their attractiveness based on their mean scores in the normative sample, a common method in the literature (e.g., Heggestad, Morrison, Reeve, & McCloy, 2006; Vasilopoulos et al., 2006). Each item consists of two statements with a four-point rating scale to indicate the preference of one statement over the other. The forced-choice version is multidimensional in the sense that the items in each item pair load differently on different facets. In line with Brown and Maydeu-Olivares (2012), item pairs consist of items with both positive statements, both negative statements, and combinations of positively and negatively worded items. The total forced-choice questionnaire consists of 225 item pairs. An example item is presented in Figure 1.

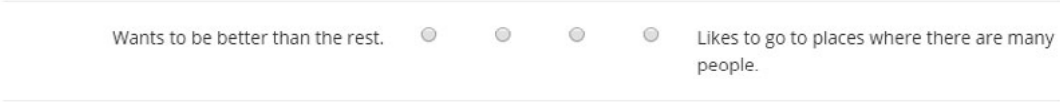

Figure 1. Example item of the FC version of the WPI. On the left is a statement measuring the Competition facet, the statement on the right measures Gregariousness.

Normative scores for the FC version of the WPI were derived by applying the Thurstonian IRT model to the WPI-FC data in Mplus (Muthén & Muthén, 1998-2010), after which trait scores were obtained with the Bayes *maximum a posteriori* (MAP) estimation method. Mplus syntax was created by the Excel macro provided by Brown and Maydeu-Olivares (2012). The present FC questionnaire includes a rating scale with four options (Figure 1) to indicate one's preference of one statement over the other, thus the model employed here is the ordinal extension of the Thurstonian IRT model. Recently it has been shown that this model can successfully be used to recover personality trait scores (Brown & Maydeu-Olivares, 2018) from FC surveys. Empirical reliabilities (Brown & Maydeu-Olivares, 2011) were adequate and roughly equal in the development (M = .84, SD = .05, ranging between .76 and .92) and selection context (M = .82, SD = .05, ranging between .72 and .91).

Statistical analyses

Standardized facet score differences between contexts within instruments were first investigated to assess the extent to which the FC method reduces the possibility of distorting responses. Based on the finding that FC tests can also be faked (e.g., Heggestad et al., 2006; Pavlov et al., 2018), we expect score differences between development and selection context for the Likert as well as the FC instrument. Yet, given that the FC format reduces the opportunity

for response distortion, we can expect the development-selection difference to be smaller for this instrument type compared to the Likert instrument.

Exploratory Structural Equation Modeling (ESEM)

To investigate whether the factor structures of personality questionnaires were equivalent across the four groups, we used the measurement invariance approach within the exploratory structural equation modeling (ESEM) framework (Asparouhov & Muthén, 2009). ESEM, in combination with target rotation, allows prespecifying a theoretical loading matrix with primary (targeted) loadings and (untargeted) cross-loadings. In the current study, this means that, for example, loadings of the facets intended to load on the Influence factor are estimated freely, while the loadings of all other facets on that factor are targeted (not fixed, they are still allowed to vary) at zero. ESEM reports factor solutions akin to EFA procedures while also providing model fit values typically found in confirmatory factor analysis (CFA) procedures, allowing for model comparisons.

ESEM has several advantages over CFA when studying the effects of response distortion on the overall factor structure and more specifically the general factor found in personality inventories (Lee, Mahoney, & Lee, 2017). By estimating cross-loadings, in other words by allowing for relations between items or facets and untargeted, presumably construct-irrelevant factors, ESEM (vs. CFA) controls for both inflated factor correlations and inflated general factor loadings (Arias et al., 2018; Lee et al., 2017; Marsh, Morin, Parker, & Kaur, 2014; Morin, Arens, & Marsh, 2016). It hereby provides a more accurate picture of the extent to which presumably independent constructs overlap more when pressures for response distortion are present.

The general factor was specified in a bifactor model, in which the 25 facets directly load on the general factor, in addition to their specified domain factor (Figure 2). Bifactor models have gained popularity in the personality literature in recent years (Arias et al., 2018; Biderman et al., 2018) as they allow appropriate decomposition of variance attributable to the general factor and the specific domain factors (Rodriguez, Reise, & Haviland, 2016). Furthermore, the bifactor approach is line with the oft-used definition of social desirability as a factor influencing all personality traits to the degree depending on the desirability or evaluativeness of the trait (Biderman et al., 2018; Smith & Ellingson, 2002). In line with how bifactor models are traditionally conceptualized (e.g., Reise, 2012), we used orthogonal target rotation, which means that the general factor and all domain factors are uncorrelated.[8] All models were fitted using the robust maximum likelihood estimator (MLR) in Mplus version 6.12 (Muthén & Muthén, 1998-2010).

We followed the procedure (and Mplus syntax) for testing measurement invariance outlined by Morin et al. (2016), in which consecutively more restricted models are tested and compared in terms of model fit. In line with Morin et al. (2016), the following models were tested (see Table 3):

1. A model in which five correlated factors are estimated, separately for the development and selection groups (labeled as **Oblique** model).

[8] In some studies, the specific domain factors are allowed to correlate with each other (Anglim et al., 2017; Biderman et al., 2018). We also estimated our models allowing for domain factor correlations but all substantive conclusions remained unchanged.

2. A model with one general factor directly loading on all facets, in addition to five specific domain factors estimated separately in both groups, without any imposed constraints (**Configural** model). This bifactor model[9] and the Oblique model are compared in terms of model fit to assess whether the addition of the general factor is a significant improvement over the correlated factors model.
3. A model in which all loadings (including cross-loadings) are constrained to be equal across both groups (**Weak** invariance model).
4. A model in which all loadings and facet intercepts are constrained to be invariant (**Strong** invariance model). If strong invariance can be demonstrated, then latent factor means from this model can be compared to investigate whether differences between the groups on these factors exist.
5. A model in which all loadings, intercepts, and facet uniquenesses (residuals) are constrained to be equal between the groups (**Strict** invariance).
6. A model in which, additionally, factor variances and covariances are constrained to equality across groups (labeled **Variance-Covariance** model).
7. A final model in which, in addition to all previously specified constraints, the **latent means** are constrained to be equal across groups.

The two instrument types differ in terms of their format, and consequently, in terms of the scores they yield: the Likert version results in regular sum scores, while the FC version (being scored by IRT) yields theta (i.e., standard normal) scores. Thus, because the facet scores from the two different instruments are on different metrics, testing the invariance of intercepts or latent means across instrument types is unjustified. Consequently, only Model 1 through Model 3 were estimated for the four groups simultaneously. The full sequence of model tests were estimated *within* instrument types and results of these parallel analyses are compared to derive conclusions on similarity of factor structures across item formats.

Goodness of fit of each of the models was evaluated using the comparative fit index (*CFI*; Bentler, 1990), the Tucker-Lewis Index (*TLI*; Tucker & Lewis, 1973), the root mean square error of approximation (*RMSEA*; Steiger, 1990) and the standardized root mean residual (*SRMR*). *CFI* and *TLI* values > 0.90 and *RMSEA* and *SRMR* values of < 0.08 are generally considered adequate (Hu & Bentler, 1999).

[9] An alternative to the bifactor model would be to fit a hierarchical model in which the domain factors load on a subordinate general factor (e.g., Van der Linden et al., 2010). However, this model fit worse (*CFI* = .826, *TLI* = .726, *RMSEA* = .113, and *SRMR* = .042 for the configural model) than the bifactor model (*CFI* = .873, *TLI* = .769, *RMSEA* = .104, and *SRMR* = .028; Table 3), and was therefore, and for the reasons outlined in the text, discarded.

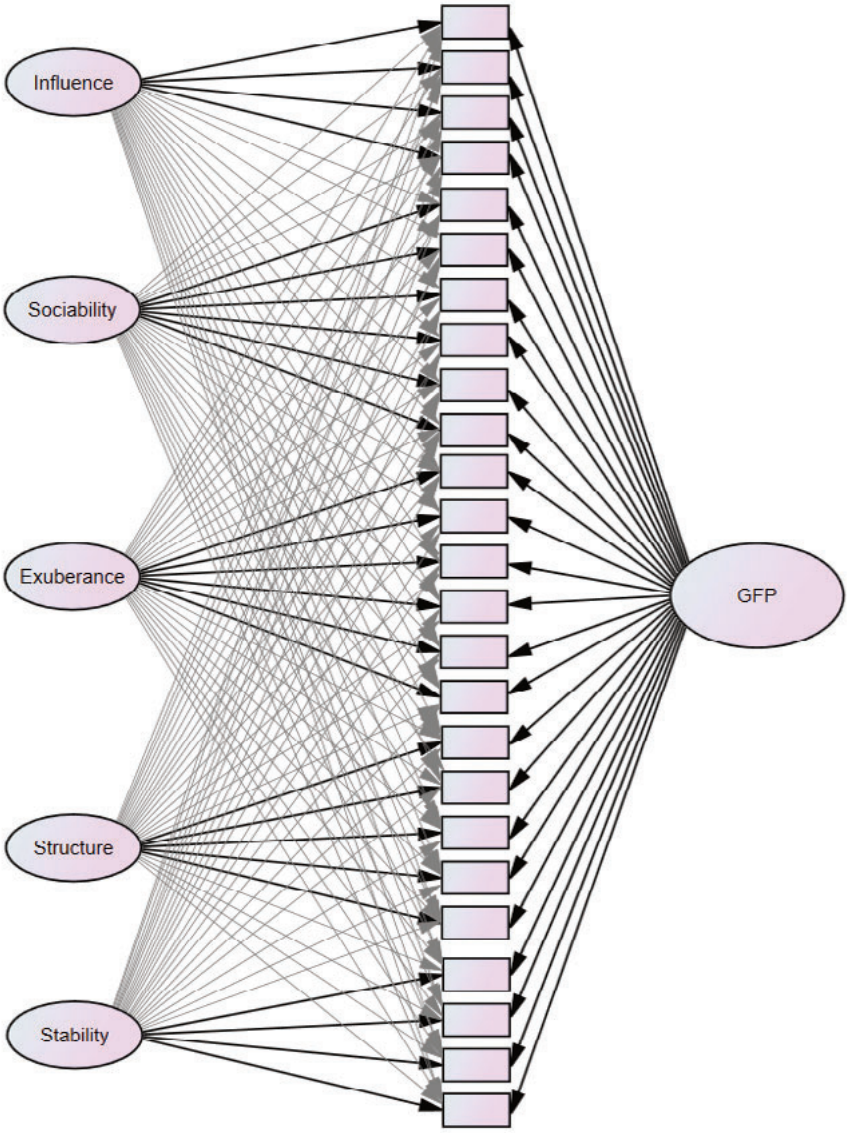

Figure 2. Schematic representation of bifactor ESEM fitted in the current study. GFP = General Factor of Personality. Black arrows represent targeted loadings, gray lines indicate untargeted cross-loadings. Rectangles are observed variables, ellipses are latent factors.

Invariance of parameters across groups constrained in each step was assessed by comparing the fit of the more restrictive model with the previous model. We followed the guidelines provided by Chen (2007) and Cheung & Rensvold (2002) stating invariance is implied when decrements in *CFI* and *TLI* values are < .01 and increases in *RMSEA* values are < .015 when moving to a more restrictive model. It has been shown, however, that the models' information criteria can sometimes lead to different conclusions compared to fit values when testing measurement invariance (Morin et al., 2016). Therefore, in addition, we inspected the Akaike Information Criterion (*AIC*; Akaike, 1987), the Bayesian Information Criterion (*BIC*; Schwartz, 1978) and the sample-size-adjusted BIC (*ABIC*; Sclove, 1987) values between models, for which lower values indicate better fitting models.

Results

Descriptive statistics

Table 2 shows the standardized differences in mean facet scores between the Likert and the FC versions of the personality survey. As expected, the difference between the development and selection context was significantly larger ($t(48) = 2.41$, $p = .02$) for the Likert version (average $\| d \| = .32$) than for the FC version (average $\| d \| = .24$). In general, facet scores were higher in the selection context than in the development context. The exceptions were the facets 'Regularity' and 'Independence' which, for the Likert version, were *lower* in the selection context while for the FC version no significant differences were found. Apparently, these two facets are deemed socially undesirable when applying for a job. The pattern of score differences were, to a large extent, similar between the two instrument types: the correlation between the *d*-values across the 25 facets was .81. For the factors, the average *d*-value was -.39 for the Likert version and -.30 for the FC version.

Overall, these findings confirm that with the FC format, it may have been more difficult to score in the desired direction. This was also reflected in smaller differences in facet variances between the development and selection contexts for the FC version (average *SD* 0.95 vs. 1.02 for selection and development, respectively), compared to the Likert format (0.89 vs. 1.03).

Table 2. Means, standard deviations, and effect sizes of facets and factors for development and selection by item format.

	Likert					Forced-Choice				
	Development		Selection			Development		Selection		
	M	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*
Status (I)	-.02	1.02	.15	.91	-.17***	-.04	.99	.14	.87	-.19***
Dominance (I)	-.09	1.06	.29	.79	-.41***	-.09	.97	.20	.92	-.31***
Competition (I)	.04	.98	.14	.96	-.11*	-.05	.97	.16	.87	-.22***
Self-presentation (I)	-.04	1.03	.19	.88	-.24***	-.02	1.01	.14	.87	-.17***
Need for contact (SO)	-.15	1.06	.17	.84	-.33***	-.04	.99	.10	.84	-.15**
Gregariousness (SO)	-.20	1.05	.24	.84	-.46***	-.06	.96	.14	.86	-.22***
Self-disclosure (SO)	-.09	1.04	.10	.89	-.19***	.02	.99	.07	.86	-.05
Trust (SO)	-.05	1.01	.08	.93	-.13**	.01	.92	.07	.84	-.08
Friendliness (SO)	-.21	1.04	.17	.89	-.40***	-.07	.95	.13	.84	-.22***
Attentiveness (SO)	-.21	1.03	.04	.94	-.25***	-.05	.92	.06	.82	-.13**
Energy (EXU)	-.22	1.05	.24	.85	-.48***	-.16	.94	.21	.87	-.40***
Personal growth (EXU)	-.12	1.02	.18	.85	-.33***	-.15	.93	.19	.87	-.38***
Perseverance (EXU)	-.24	1.05	.18	.89	-.43***	-.23	.89	.18	.86	-.46***
Adaptability (EXU)	-.19	1.04	.21	.84	-.42***	-.13	.91	.18	.89	-.34***
Originality (EXU)	-.10	1.05	.24	.84	-.36***	-.01	.95	.13	.91	-.15**
Independence (EXU)	.15	.97	-.08	.94	.24***	.05	.94	.00	.90	.06
Orderliness (STRC)	-.17	1.04	.09	.96	-.26***	-.08	.94	.03	.86	-.12**
Precision (STRC)	-.16	1.02	.07	.99	-.23***	-.15	.88	.07	.94	-.25***
Regularity (STRC)	.06	1.02	-.18	.93	.25***	-.02	.91	-.06	.89	.04
Conformity (STRC)	-.14	1.01	.07	.92	-.22***	-.16	.85	.13	.91	-.33***
Deliberation (STRC)	-.10	1.05	.13	.90	-.23***	-.12	.94	.07	.88	-.21***
Self-confidence (STAB)	-.15	1.06	.25	.83	-.42***	-.13	.98	.20	.89	-.34***
Optimism (STAB)	-.23	1.03	.29	.78	-.57***	-.12	.90	.20	.84	-.37***
Frustration-tolerance (STAB)	-.23	1.06	.20	.91	-.44***	-.15	.93	.15	.86	-.33***
Resilience (STAB)	-.23	1.03	.26	.88	-.51***	-.19	.95	.20	.87	-.43***
Mean					-.28					-.23
Mean (absolute)					.32					.24
Influence					.63***					.20***
Sociability					-.04					.18**
Exuberance					.17*					.20**
Structure					.11					.28***
Stability					.18**					.31***
GFP					.38***					.47***

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. Latent mean differences are derived from the Strong invariance models. I = Influence; SO = Sociability; EXU = Exuberance; STRC = Structure; STAB = Stability.

Results from ESEM analyses

Measurement Invariance

The fit values of the different models included in the measurement invariance tests are reported in Table 3.

Invariance across both contexts and item formats

First, we discuss the results from the tests of Model 1 through Model 3 based on all four groups simultaneously. The Oblique model showed poor fit in comparison to common guidelines (with the exception of the *SRMR* value). Reasons for this can be found in the intercorrelations among the factors (Table 4). It appears that differences between the instruments were present, even in the absence of pressures for self-enhancement. The main culprit was the Structure factor, which showed the most diverging correlations across the two instrument types (see also Table 7). Within instruments, however, the correlations did not differ too dramatically across contexts. Although differences between the specific factors appeared to exist, adding the general factor in the bifactor (Configural) model did, however, significantly improve model fit indicated by the fit values and information criteria compared with the Oblique model. Thus, we continued with our invariance test for this superior model.

Evidence for the invariance test for equality of factor loadings (Weak versus Configural) appeared to be mixed: the decrease in *CFI* value exceeded the cutoff of .01, and both the *AIC* and *ABIC* were higher in the Weak model than in the Configural model, all indicating non-invariance. However, the *BIC* value was lower for the Weak model. In addition, the *TLI* value actually *increased* by a fairly large amount ($\Delta TLI = .058$), while the *RMSEA* actually *decreased* ($\Delta RMSEA = -.014$), indicating better model fit. Given that the *RMSEA* and *TLI* values impose a stronger penalty for model complexity than the *CFI* (Kenny, 2015), it appears more likely that invariance of factor loadings across the four groups held.

However, because of the mixed evidence, it is informative to investigate the factor loadings from the Configural model to identify possible differences between the groups. Table 5 and Table 6 show the factor loadings from the configural model for respectively the FC and Likert versions of the personality questionnaire. Focusing on the five factor model structure (ignoring the general factor first), we find that the ESEM bifactor model has recovered the targeted factor structure quite well in each of the four groups. That is, the targeted loadings were relatively high, while the cross-loadings were relatively low: the mean absolute targeted loadings ranged between .34 and .64, while the mean absolute cross-loadings ranged between .08 and .15. In general, the facets showed their highest loading on their targeted factor, with some exceptions (e.g., Self-presentation on the Sociability factor in the FC version in both contexts, and Perseverance on the Exuberance factor in all four groups).

Table 3. Goodness-of-fit statistics and information criteria for the models estimated.

	χ^2	*df*	*CFI*	*TLI*	*RMSEA*	*RMSEA 90% CI*	*SRMR*	*AIC*	*BIC*	*ABIC*
Four groups										
Oblique Model	10382.998	740	0.828	0.721	0.114	[0.112 - 0.116]	0.036	219439	223590	221493
Bifactor Configural	7780.107	660	0.873	0.769	0.104	[0.102 - 0.106]	0.028	216728	221382	219030
Bifactor Weak	9300.665	1027	0.852	0.827	0.090	[0.088 - 0.092]	0.057	218371	220717	219531
Forcea-Choice										
Oblique Model	7857.398	370	0.789	0.659	0.143	[0.140 - 0.145]	0.038	104107	105953	104905
Bifactor Configural	5804.954	330	0.846	0.720	0.129	[0.126 - 0.132]	0.029	102114	104185	103009
Bifactor Weak	5609.254	444	0.855	0.804	0.108	[0.106 - 0.111]	0.033	102072	103505	102691
Bifactor Strong	5694.453	463	0.853	0.809	0.107	[0.104 - 0.109]	0.033	102098	103424	102671
Bifactor Strict	5792.855	488	0.851	0.817	0.105	[0.102 - 0.107]	0.035	102168	103355	102681
Bifactor Variance-Covariance	5851.231	509	0.850	0.823	0.103	[0.100 - 0.105]	0.044	102231	103300	102693
Bifactor Latent Means	6015.954	515	0.845	0.820	0.104	[0.101 - 0.106]	0.058	102401	103436	102848
Likert										
Oblique Model	2583.819	370	0.890	0.822	0.078	[0.075 - 0.080]	0.034	115332	117179	116131
Bifactor Configural	2048.528	330	0.921	0.857	0.070	[0.067 - 0.073]	0.026	114614	116684	115509
Bifactor Weak	1917.152	444	0.920	0.893	0.060	[0.058 - 0.063]	0.033	114641	116073	115260
Bifactor Strong	2132.976	463	0.917	0.893	0.060	[0.058 - 0.063]	0.034	114684	116011	115258
Bifactor Strict	2498.906	488	0.900	0.877	0.064	[0.062 - 0.067]	0.043	115029	116216	115542
Bifactor Variance-Covariance	2680.125	509	0.892	0.873	0.065	[0.063 - 0.068]	0.065	115223	116292	115685
Bifactor Latent Means	2866.48	515	0.883	0.864	0.068	[0.065 - 0.070]	0.084	115419	116454	115866

Note. CFI = comparative fit index; *TLI* = Tucker–Lewis Index; *RMSEA* = root mean square error of approximation; *CI* = confidence interval; *AIC* = Akaike information criterion; *BIC* = Bayesian information criterion; *ABIC* = sample-size-adjusted *BIC.*

Table 4. Latent factor intercorrelations for development and selection by item format.

	Forced-Choice					Likert					
	I	SO	EXU	STRC	STAB		I	SO	EXU	STRC	STAB
I	-	.30**	.44**	-.19**	.36**	I	-	.13**	.31**	-.15**	.21**
SO	.36**	-	.19**	-.40**	.34**	SO	.16**	-	.39**	.09**	.42**
EXU	.46**	.31**	-	.03	.53**	EXU	.35**	.27**	-	.24**	.43**
STRC	.09*	-.28**	-.04	-	-.11**	STRC	-.17**	.02	.18**	-	.15**
STAB	.42**	.29**	.38**	.05	-	STAB	.25**	.36**	.42**	.02	-

Note. * $p < .05$. ** $p < .01$. Correlations below the diagonal are based on the development group, correlations above the diagonal on the selection group. I = Influence; SO = Sociability; EXU = Exuberance; STRC = Structure; STAB = Stability.

.

It should be noted here that the factor loadings in Table 5 and Table 6 again show that some differences in the factor structures exist between the FC and Likert version, irrespective of context. For example, in the FC version, Trust had a relatively strong negative untargeted loading on the Structure factor in both contexts, while this was not found in the Likert version. Thus, differences in factor structures were more notable *across* instrument types than *within* instrument types across contexts. Still, congruence analyses showed that the five specific domain factors were highly similar across the four groups (most congruence coefficients being >.90; Table 7).

Focusing on the general factor, the results showed that this factor was well-defined in all four groups with an average loading of .43 in both the FC development group (range -.49 – .90) and the FC selection group (range -.52 – .89), .42 in the Likert development group (range -.35 – .81) and .44 in the Likert selection group (range -.22 – .75). In all four groups, the GFP was defined by high Dominance, high Gregariousness, high Energy, high Self-confidence, and low Regularity. However, the general factors from the two different instrument types had somewhat different 'flavors': Independence was only a strong indicator of the general factor in the FC version (λ of about .75 in both contexts). Similarly, competitiveness loaded on the general factor in the FC version (λ of about .55 in both contexts) but not in the Likert version (loadings <.30 in both contexts). Thus, the response format partly shaped the content of the general factor (see *Discussion*). However, a more formal test in the form of congruence analysis nevertheless showed the general factors to be essentially equivalent across groups (congruence coefficients approaching unity; Table 7).

Invariance across contexts within item formats

Although the previous analyses indicated that the factor structures were similar across the four groups, other relevant aspects (such as latent mean differences) could not be investigated. Therefore, we turn to our invariance tests within instruments across contexts.

Forced-Choice. Generally speaking, for the FC version, in absolute terms, the fit values of the models were below common guidelines (Hu & Bentler, 1999), with the exception of the *SRMR* value.[10] However, in relative terms, adding the invariance constraints did not result in decreases in model fit exceeding the recommended cutoffs (ΔCFI and $\Delta TLI < 0.01$ and $\Delta RMSEA < 0.015$). The information criteria sometimes provided mixed evidence. For example, moving from the Weak to Strong model, the *AIC* value increased, while both *BIC* and *ABIC* decreased, indicating a better fitting model when factor loadings are constrained to be equal. However, in combination with the small decreases in fit indices found, it appears that measurement invariance between the development and selection group was largely supported for the FC version. An exception was the final Latent Means model; when testing invariance of latent means, information criteria are considered better indicators than changes in fit values (Fan & Sivo, 2009; Morin et al., 2016). The information criteria of the Latent Means model were all higher

[10] Based on modification indices, correlated residuals were freed until fit values of the configural model were more acceptable (ten in total). The full sequence of invariance tests was then carried out starting from this baseline model, however, the substantive conclusions reported in the text remained unchanged.

than the information criteria of the Variance-Covariance model, indicating that the latent means presumably were different across the four groups (see below).

Likert. The conclusions pertaining to the FC version largely applied to the Likert version, with some exceptions. Yet, overall, the fit values of the models for the Likert version were higher than for the FC version, and indicating adequate fit from the Configural model onwards. A second difference compared with the FC version was that Strict invariance did not seem to hold; although the increase in the *RMSEA* value was below .015 ($\Delta RMSEA = .004$), the decreases in *CFI* and *TLI* values were too large ($\Delta CFI = -.017$ and $\Delta TLI = -.016$), and all information criteria were higher in the more restrictive model.[11] Investigation of uniquenesses indicated that the error variances in the selection group were lower than the error variances in the development group (average u = .37 for selection, .46 for development). This was not the case for the FC version (average u = .28 for selection, .32 for development). These findings mimic those found for the facet variances (Table 2) and previous studies (e.g., Anglim et al., 2017), and could indicate that in the selection context, facet scores become more reliable because applicants respond in a more consistent fashion in line with an ideal response (Anglim et al., 2017; Griffith et al., 2007; cf. MacCann, 2013). In addition, evidence for the invariance of the variance-covariance matrix across contexts was mixed, given that the decreases in fit values were below the recommended cutoffs, while the information criteria increased compared to the Strict model. Finally, as for the FC version, invariance of latent means did not seem to hold based on the information criteria.

Latent mean differences

The invariance tests showed that latent mean differences appeared to exist between the development and selection group for both instruments. Based on the parameters from the Strong invariance model, for the Likert version, the latent mean of the general factor in the Selection group was $d = .63$ higher than in the development group, while in the FC version this difference was .20 (see Table 2). In terms of the specific factors, for the Likert version, the latent mean differences of the Influence ($p = .56$) and Exuberance ($p = .08$) factors were not significant. For the FC version, all latent mean differences of the specific factors were significant and somewhat larger (average $d = .29$) than for the Likert version (average $d = .16$). Apparently, differences between the development and selection context were absorbed by the general factor in the Likert version, while in the FC version, the differences were more evenly spread out across the general and specific factors. These findings appear to be in line with the idea that the FC format complicates elevating scores on all traits simultaneously, instead causing people to elevate their scores more on specific traits (e.g., Brown, 2008).

[11] We did find, however, that by freeing only three uniquenesses (for Energy, Originality and Positivity) based on the modification indices, it was possible to keep the reduction in model fit according to the *CFI*, *TLI*, and *RMSEA* below the recommended cutoffs.

Table 5. Standardized factor loadings for configural bifactor exploratory structural equation model solution for development and selection, forced-choice.

	Development						Selection					
	GFP	I	SO	EXU	STRC	STAB	GFP	I	SO	EXU	STRC	STAB
Status	.60***	.63***	-.04	.19***	.02	-.03	.53***	.60***	-.08**	.23***	.07**	-.05*
Dominance	.90***	.14***	.06**	.06*	.11***	.08**	.89***	.11**	.06*	.06*	.09***	.07**
Competition	.56***	.68***	-.17***	.12***	.01	-.10***	.54***	.68***	-.19***	.17***	.04	-.06**
Self-presentation	.82***	.18***	.36***	-.18***	.06*	-.05	.83***	.17***	.31***	-.22***	.04	-.07**
Need for contact	.55***	.19***	.61***	.02	-.06*	-.03	.53***	.17***	.62***	.05	-.11***	.04
Gregariousness	.81***	.03	.41***	-.06**	.04	.01	.83***	.03	.38***	-.07***	.02	.09***
Self-disclosure	.38***	-.02	.61***	-.10***	-.14***	-.21***	.35***	-.07**	.61***	-.18***	-.11***	-.23***
Trust	.31***	-.12***	.42***	-.08*	-.39***	.03	.35***	-.15***	.44***	-.04	-.35***	.06*
Friendliness	.62***	-.02	.58***	.01	-.16***	.05*	.61***	-.06*	.54***	.02	-.18***	-.01
Attentiveness	.19***	-.24***	.51***	.15***	-.08*	-.09*	.21***	-.19***	.49***	.14***	-.10**	-.04
Energy	.71***	.14***	.11***	.37***	.15***	.13***	.71***	.14***	.09***	.43***	.07**	.04
Personal growth	.58***	.24***	-.05	.61***	.14***	.03	.58***	.22***	-.12***	.61***	.13***	.06*
Perseverance	.27***	.14***	-.14***	.40***	.65***	.13***	.28***	.16***	-.10***	.49***	.64***	.10***
Adaptability	.77***	-.07**	.12***	.49***	-.03	.20***	.77***	-.06**	.10***	.49***	-.03	.20***
Originality	.71***	-.06	.01	.14**	-.20***	-.05	.71***	-.02	.03	.16***	-.15***	-.01
Independence	.74***	-.26***	-.36***	.12	-.12	-.18***	.76***	-.20***	-.32***	.08	-.17***	-.20***
Orderliness	.03	-.08**	-.20***	.11**	.68***	-.07*	.05	-.05	-.21***	.05	.69***	-.03
Precision	-.22***	.05	-.22***	.25***	.66***	-.05	-.26***	.09***	-.21***	.27***	.70***	.08***
Regularity	-.49***	-.08**	-.01	-.14***	.64***	-.18***	-.52***	-.07*	-.06*	-.15***	.61***	-.14***
Conformity	-.27***	.20***	.20***	.19*	.44***	.38***	-.30***	.15***	.17***	.23***	.51***	.34***
Deliberation	-.08	.05	-.40***	.22***	.51***	.10*	-.05	.05	-.35***	.23***	.54***	.14***
Self-confidence	.82***	.07*	-.15***	-.07**	.14***	.38***	.82***	.07**	-.11***	-.09***	.15***	.34***
Optimism	.75***	.05*	.11***	.01	-.16***	.41***	.76***	.03	.13***	.03	-.15***	.35***
Frustration-tolerance	.23***	-.10**	-.11**	.19***	-.03	.71***	.28***	-.10***	-.05	.16***	.09**	.76***
Resilience	.47***	-.09***	-.07**	.10***	.08***	.71***	.50***	-.06**	-.06*	.12***	.19***	.68***

(continued)

Table 5 (continued).

	Development						Selection					
	GFP	I	SO	EXU	STRC	STAB	GFP	I	SO	EXU	STRC	STAB
ECV	.47						.48					
ω_h	.74						.74					
ω	.95						.95					
Relative omega	.78						.78					

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. Targeted loadings in brackets. I = Influence; SO = Sociability; EXU = Exuberance; STRC = Structure; STAB = Stability; ECV = explained common variance; ωh = coefficient omega hierarchical; ω = coefficient omega total.

Table 6. Standardized factor loadings for configural bifactor exploratory structural equation model solution for development and selection, Likert.

	Development						Selection					
	GFP	I	SO	EXU	STRC	STAB	GFP	I	SO	EXU	STRC	STAB
Status	.38***	.68***	-.09**	.25***	.02	.03	.26***	.75***	-.04	.24***	.00	.06*
Dominance	.75***	.29***	-.01	.02	-.20***	-.10**	.74***	.22***	-.08**	.00	-.26***	-.13
Competition	.28***	.65***	-.18***	.20***	.02	-.09**	.25***	.68***	-.17***	.19***	-.03	-.08**
Self-presentation	.61***	.47***	.22***	-.24***	-.23***	-.20***	.52***	.51***	.15***	-.26***	-.26***	-.19***
Need for contact	.50***	.14***	.55***	.07*	-.03	.04	.45***	.20***	.48***	.06	-.01	.09*
Gregariousness	.73***	.06	.29***	-.14***	-.06*	.09	.70***	.09**	.26***	-.18***	-.07*	.06
Self-disclosure	.33***	-.04	.55***	-.15***	-.10**	-.11**	.28***	.00	.42***	-.14***	-.08*	-.14**
Trust	.28***	-.10**	.41***	-.10**	-.09**	.15***	.21***	-.09*	.35***	.07*	-.19***	.07
Friendliness	.55***	-.02	.63***	.07**	.07**	.05	.66***	-.06*	.48***	.05*	.11***	.06
Attentiveness	.27***	-.17***	.61***	.19***	.20***	.00	.37***	-.28***	.51***	.20***	.23***	-.07
Energy	.72***	.06	.10***	.36***	.19***	.02	.75***	.08**	.10***	.30***	.12***	.05
Personal growth	.50***	.29***	.03	.55***	.06*	.01	.45***	.23***	.06*	.55***	.15***	-.04
Perseverance	.52***	-.10*	.00	.35***	.50***	-.05	.60***	.03	.03	.30***	.43***	.00
Adaptability	.68***	.03	.08**	.39***	-.18***	.16***	.69***	-.02	.12***	.36***	-.16***	.11***
Originality	.57***	.08	-.03	.25***	-.19***	-.05	.66***	-.02	-.03	.25***	-.17***	-.02
Independence	.27***	.04	-.28***	.15*	-.10**	-.26***	.23***	.15**	-.18***	.17**	-.21***	-.24***
Orderliness	.31***	-.13***	-.07*	.05	.64***	-.01	.46***	-.10**	-.10**	-.03	.56***	.00
Precision	.13**	-.09**	-.08**	.15***	.67***	-.03	.28***	-.02	-.02	.13***	.69***	.00
Regularity	-.35***	.04	.04	-.19***	.63***	-.20***	-.22***	-.04	.07*	-.15***	.57***	-.10**
Conformity	-.04	.01	.38***	.15***	.54***	.22***	.11*	-.15***	.32***	.15**	.55***	.26***
Deliberation	.14**	-.05	-.20***	.11*	.45***	.28***	.26***	-.07*	-.19***	.09*	.48***	.14**
Self-confidence	.81***	-.02	-.17***	-.15***	.07**	.23***	.69***	.11***	-.19***	-.15***	.03	.31***
Optimism	.64***	-.09**	.16***	.07*	.07*	.32***	.61***	-.11***	.15***	.11***	.05	.31***
Frustration-tolerance	.36***	-.13***	.13***	.06	.10***	.65***	.38***	-.14***	.08**	.07*	.15***	.59***
Resilience	.57***	-.09**	-.02	-.04	-.01	.55***	.50***	-.08	-.03	-.07*	.08**	.55***

(continued)

Table 6 (continued).

	Development						Selection					
	GFP	I	SO	EXU	STRC	STAB	GFP	I	SO	EXU	STRC	STAB
ECV	.43						.45					
ω_h	.73						.76					
ω	.93						.93					
Relative omega	.79						.82					

Note. ** $p < .05$. ** $p < .01$. *** $p < .001$. Targeted loadings in brackets. I = Influence; SO = Sociability; EXU = Exuberance; STRC = Structure; STAB = Stability; ECV = explained common variance; ωh = coefficient omega hierarchical; ω = coefficient omega total.

General Factor Saturation and Explained Total Variance

The relevance of the general factors in each of the four groups was estimated by OmegaH or ω_h, which is a measure of explained variance by a higher-order factor (McDonald, 1999; Revelle & Zinbarg, 2009; Rodriguez et al., 2016) and relative omega (by dividing ω_h by ω, a measure of explained variance by all sources of common variance), which indicates how much of the *reliable* variance is captured by the general factor. Finally, the explained common variance (ECV) by the general factor was calculated, which indicates the relative strength of the general factor relative to the specific domain factors.

Table 5 and Table 6 (lower panels) show the ECV, ω_h and relative omega values calculated based on the loadings from the Configural model for the FC and Likert version respectively.[12] The results indicate that all three indicators of general factor saturation were highly similar across the four groups. Thus, the context and item format did not appear to have a strong influence on the size or importance of the general factor present in the personality inventories.

However, the size of the general factor in itself does not give direct insight into whether it mostly captures substantial or error variance. Indications of these different types of variances can, however, be inferred from comparing the sources of explained total variance (ETV) in the bifactor model with those in the Oblique model (five correlated factors). In the Oblique model, each facet's total explained variance can be divided in variance attributable to the specific domain factor (S), cross-loadings (CR), and error (uniqueness or u). In the bifactor model, the total explained variance is attributable to either the GFP, S, CR, or u. Consequently, when the GFP is introduced in the bifactor model, it must by definition absorb either variance attributable to, S, CR, or u. If the GFP in the bifactor model takes up mostly uniqueness variance from the Oblique model, then this would suggest that the GFP reflects error rather than variance attributable to the substantive specific traits. If the explained variance due to S decreases considerably and moves to the GFP, then this would be more in line with the GFP capturing substantive trait variance. The final source of variance, CR, will most likely be a mix of substance and error: the cross-loadings indicate (partial) unintended overlap between facets, which can be either conceptually meaningful, or due to, for example, similarity in wording of items within the facets.

[12] The most restrictive model that held for both instrument types was the Strong invariance model. Therefore, these analyses were also carried out on the parameters from the Strong invariance models. However, the results were virtually identical and therefore not reported here.

Table 7. Congruence coefficients for specific domain factors and the general factor across combinations of context and item format based on configural bifactor exploratory structural equation model solution.

	FC Development						FC Selection						Likert Development					
	GFP	I	SO	EXU	STRC	STAB	GFP	GFP	I	SO	EXU	STRC	GFP	I	SO	EXU	STRC	STAB
FC Selection	1.00	.99	.99	.99	.99	.98												
Likert Development	.98	.90	.94	.95	.90	.90	.98	.92	.94	.97	.89	.94						
Likert Selection	.97	.90	.92	.91	.89	.91	.97	.91	.92	.92	.89	.94	.99	.96	.99	.98	.98	.97

Note. FC = Forced-choice; GFP = General Factor of Personality; I = Influence; SO = Sociability; EXU = Exuberance; STRC = Structure, STAB = Stability.

Table 8 shows the amounts of explained variance attributable to each of the aforementioned sources in each of the four groups. First, in all four groups, it becomes clear that when the GFP is introduced in the bifactor model, the largest reduction in ETV can be found for specific domain variance, followed by variance attributable to the cross-loadings. Thus, the GFP appears to absorb mostly variance from the specific domain factors, rather than variance attributable to measurement error. The fact that this result is found in all four groups suggests that this finding is unaffected by circumstances related to the motivation and opportunity to self-enhance. Note that this information only informs us that the makeup of the GFP appears to be consistent across the four groups, not what the makeup exactly is. That is, it provides some evidence that the GFP is not wholly artefactual, yet, this does not imply evidence for the claim that it is substantial.

Table 8. Decomposition of explained total variance in oblique and bifactor exploratory structural equation model.

	GFP	S	CR	u	-ΔS	-ΔCR	-Δu	%S	%CR	%u
Forced-Choice Development										
Oblique model		.50	.16	.33						
Bifactor model	.33	.27	.10	.30	.23	.06	.03	72	19	9
Forced-Choice Selection										
Oblique model		.52	.14	.34						
Bifactor model	.33	.27	.10	.30	.25	.04	.04	75	15	11
Likert Development										
Oblique model		.42	.12	.46						
Bifactor model	.25	.25	.08	.43	.17	.04	.03	71	17	12
Likert Selection										
Oblique model		.40	.11	.49						
Bifactor model	.24	.22	.08	.46	.18	.03	.03	74	14	12

Note. GFP = General Factor of Personality; S = Specific domain factors; CR = cross-loadings; u = uniqueness/error; %S = Percentage of GFP variance attributable to specific domain variance; %CR = Percentage of GFP variance attributable to cross-loadings; %u = Percentage of GFP variance attributable to uniqueness/error variance.

Discussion

With regard to the main goal of the present study, we showed that the overall factor structures as well as the general factor were robust to differences in the item format of the questionnaire and test-taking context. Overall, the factor structure of the five factors and the general factors were nearly identical irrespective of whether one filled in the surveys in a selection or development context, and irrespective of whether the Likert or the FC-version was used. In addition, over the instrument and context types, the level of explained variance by the general factor also was highly similar.

Our findings are in line with previous studies that did not find any differences in factor structures between high-stakes and low-stakes settings (e.g., Anglim et al., 2017; Ellingson et al., 2001; Marshall et al., 2005). Other previous studies, however, did find such differences. Explanations for the divergent findings might be found in methodological differences between our study and other studies (ESEM vs. CFA (e.g., Schmit & Ryan, 1993), where CFA may lead to inflation of general factor loadings) or differences in the study setting (real-life vs. laboratory setting with faking instructions; e.g., Ellingson et al., 1999; Schermer et al., 2019a, 2019b). Another explanation may lie in the specific samples used for the low-stakes settings (career development vs. students; e.g., Schmit & Ryan, 1993). The use of a career development group in the current study may also have been responsible for the somewhat lower mean differences we found between the development and selection context compared to previous studies; a career development process is notably different from, for example, a setting in which students complete a personality survey for research purposes. Because of this setting, the development group might have responded slightly more socially desirable than other groups that are often used as a comparison for applicants. Interestingly, our effect sizes are highly comparable to those found in another study using a development group (Ellingson et al., 2007), and were largest for (facets of) Emotional Stability as is often found (Birkeland et al., 2006) in the literature. Thus, our results appear to be largely in line with previous studies on this topic.

The results from the current study contribute to insight in the construct validity of the social desirability factor and provides additional information on the extent to which it might reflect a trait or an artefact. First, the fact that mean differences in facet scores and general factor scores were present between the development and selection context implies that the social desirability factor, in line with (modern) trait theories (e.g., Fleeson & Jayawickreme, 2015; Mischel & Shoda, 1995), is partly a function of the motivational context in which it is assessed. This conclusion was supported by the finding that, compared to the Likert version, the influence of the motivational context was smallest for the FC version.

Yet, it also became clear that the differences in means over the motivational contexts and the different survey types coexisted with similarity in factor structures. This similarity implies that the content and covariation among traits remains largely the same across contexts and item formats (i.e., people scoring high on Self-confidence also generally scored high on Resilience,

regardless of whether a FC or Likert questionnaire was completed and regardless of the test-taking context). In the same manner, similarity of the general factor across the groups makes it plausible that this factor is not simply only the result of response distortion evoked by the context; if one would assume that the larger part of the social desirability factor was due to unrealistic self-enhancement – which should be increased in selection procedures and reduced by the forced-choice format – then it could be expected that the factor structure would differ more between the groups, which was not the case.

In general, the main findings described above can be summarized in the following simple theoretical model;

$$OPT = TG + UPT + FG + C + e,$$

where OPT is the observed personality trait (item or facet) score in a selection context, TG represents "true" GFP variance, UPT is the unique variance of the personality trait, FG is "faked" GFP variance (caused by self-enhancement on all personality traits due to the test-taking context), C captures contextual factors (such as the job type one is applying for or the presence of warnings that faking will be detected; e.g., Dwight & Donovan, 2003; Furnham, 1990) and *e* represents unsystematic measurement error. Results from the present study imply that, when personality questionnaires are used in the selection context, levels of social desirability *can* be altered, but that score shifts appear to mainly take place in the FG part of the equation. For FC questionnaires, it appears that it is harder to alter the FG score, leading to smaller shifts in observed scores. Yet overall, a significant amount of variance still remains in the TG and UPT part, enough to keep the rank-orders between traits the same, retaining the factor structures and level of general factor saturation. Alternatively, if the GFP had truly represented merely a faking factor and hence the score shifts had mainly taken place in the TG parts, then the factor structures and the general factor would have been more different between the four groups.

The model presented above can easily be extended by the incorporation of interactions, for example between TG and FG: it may be hypothesized that those with high TG (i.e., socially desirable) scores "fake" less because they do not need to, while those with lower TG scores add a certain constant to all the traits (captured by the FG score), yet not enough to fundamentally change the rank-orders of persons across traits (Anglim et al., 2017, Ellingson et al., 2001). Again, this combination of response processes would result in mean shifts in observed scores but similar factor structures. As an example of an UPT x FG interaction, higher Conscientiousness might be associated with reduced faking (e.g., McFarland & Ryan, 2000). One can also think of interactions between FG and context (C); the type of job one is applying for may influence the extent and nature of applicant faking (Furnham, 1990). However, what most distinguishes the above model from previous models (e.g., Tett & Simonet, 2011) is that it incorporates a true social desirability factor (TG); consequently, hypotheses should be formulated in terms of relations between unique personality trait variance

(UPT) – that is what remains after accounting for true individual differences in social desirability – and faking (FG).

It should be noted here that no definite conclusion can be drawn whether the general factor in personality is substantive or artefactual based on the results of the current study. We have shown that the general factor is unlikely to be mainly caused by response distortion due to the motivation to self-enhance; yet, this is not the same as showing that it is a substantive factor. The only way to properly investigate the substance of the general social desirability factor is by relating it to external criteria, such as other-reports (see Chen et al., 2016), or objective outcomes (see for example, Pelt et al., 2017). Nevertheless, the results of the present study are informative regarding the role of the general factor in selection and assessment contexts and whether or not it should be treated as a nuisance or a potentially relevant construct.

A practical implication of the current study underlines previous statements that practitioners need not worry too much about the effect of SDR on construct validity, and that the FC format can remedy some of their concerns in terms of score shifts. Practitioners may thus use both Likert and FC questionnaires in both career counseling and selection contexts, without significant loss of construct validity. Accordingly, the choice for the instrument type should rather be based on other factors, such as their criterion validity or applicant reactions towards them. In terms of applicant reactions, studies have shown that FC questionnaires elicit more negative attitudes, because of their cognitive demands and frustration related to being forced to choose between two equally (un)attractive options (Converse et al., 2008). Regarding criterion validity, a recent meta-analysis has shown higher predictive validities for FC questionnaires compared to Likert questionnaires (Salgado et al., 2015). However, these meta-analytic data were gathered under low-stake settings and the results should be replicated based on data obtained in high-stake settings. In addition, because FC questionnaires are more cognitively demanding to complete (Vasilopoulos et al., 2006), part of the higher criterion validity might actually be due to cognitive abilities rather than the intended measured personality constructs. Higher saturation with intelligence of FC questionnaires might also explain why the (general) factor structures of the two instrument in the current study differed somewhat even when both were completed in the development context; perhaps, individual differences in intelligence influenced the content of the facets in the FC version to a larger extent than in the Likert version, slightly altering the (general) factor(s) and reducing model fit of the ESEM models of the FC instrument.

Limitations and Future Research

Although our between-subject design served the purposes of the current study, it comes with its limitations (Ellingson et al., 2007). For example, the motivated and non-motivated groups were not formed at random. On the other hand, the matching procedure ensured that the groups were equivalent at least in terms of demographic variables. Still, our findings might have been affected by group differences on confounding variables related to social desirability as a response style (e.g.,

competitive worldviews; Roulin et al., 2016), as a trait (e.g., self-control; Uziel, 2010a), or both (e.g., integrity; De Vries et al., 2014; McFarland & Ryan, 2000, 2006). A repeated measures within-subject design using the same four groups as in the current study would be an interesting follow-up study as it allows for controlling for and investigation of aforementioned confounders (see Van Geert el al., 2016).

In the previous section, we have outlined a simple theoretical model that can be tested in the future. The current study can be regarded as a first step in disentangling the different sources of variance, but more studies are needed. Possibilities for further studies include variance decomposition studies (e.g., McCrae, 2018) comparing self and other reports based on Likert and FC versions of personality. The current study has shown that the FC method can reduce some rater bias in self-ratings, while Bartram (2007) showed a similar effect for other-ratings. Thus, when using the FC format for both self and other-ratings, more bias-free estimates of traits (facets, Big Five domains or the GFP) might be obtained.

The present study focused on the effect of item format and test-taking context on the construct validity of the GFP; a next logical step would be to focus on its criterion validity. That is, the substantive interpretation of the GFP would be even more plausible if similar associations between the GFP and criteria (e.g., job performance) are found in each of the four groups of the present study. In our view, a full, perhaps meta-analytic, investigation of the criterion validity of the GFP extracted from FC and Likert questionnaires (Pelt et al., 2017), both under low-stakes and high-stakes settings, would therefore be appropriate at this point (e.g., Salgado et al., 2015). In addition, with the advent of new psychometric methods to retrieve normative scores from FC questionnaires such as the Thurstonian IRT model, data from previous studies can be reanalyzed to get a clearer picture of the criterion validity of FC questionnaires – and the GFPs extracted from them.

A correlate that would be useful to take into account in future studies is cognitive ability. First, the relation between the GFP and cognitive abilities under low-stakes settings remains unclear (Dunkel, Van der Linden, Beaver, & Woodley, 2014; Loehlin et al., 2015). Second, under high-stakes settings, the association between the GFP and intelligence has been found to be positive and inflated – although this relation was only found in an instructed, laboratory setting (MacCann et al., 2017) and not with real job candidates (Schermer & Goffin, 2018). The positive finding is consistent with the idea that intelligence is related to identifying what behavior is required in selection situations (Bing et al., 2004; Geiger, Olderbak, Sauter, & Wilhelm, 2018). Given that FC questionnaires make this identification harder, the relation between the GFP and intelligence may be stronger for these types of inventories (Vasilopoulos et al., 2006). Consequently, it might be that the GFP-criterion relations for different item formats and contexts will be moderated by cognitive abilities (Salgado et al., 2015).

Concluding remarks

SDR or faking continues to be of interest to both researchers and practitioners. The current study shows that in terms of construct validity, previously expressed concerns about the effect of SDR may be unwarranted in the sense that factor structures were highly similar across groups differing in their motivational pressures and opportunities to distort responses. The same equivalence was found for the general factor in the personality questionnaire. These results seem to suggest that social desirability in general, and the common variance among personality traits more specifically, appears to be more consistent and robust than previously thought.

Summary and General Discussion

6

Self-report personality questionnaires are commonly used by organizations as part of their selection procedures (Kantrowitz et al., 2018). Although meta-analyses have shown self-report personality test scores to predict important work-related outcomes, some organizations are still reluctant to adopt personality questionnaires because of the 'problem of social desirability' (Cook, 2016). That is, concerns exist that applicants respond in order to present the best image of themselves to increase their chances of getting hired, rather than to respond in an honest fashion in line with how they would typically behave, hereby negatively affecting the validity of personality measures. In this light, social desirability is regarded as a factor that introduces measurement bias to the valid measurement of personality traits, implying a reduction of the validity and practical utility of personality questionnaires in selection procedures.

This interpretation of social desirability as a bias is widely shared in the personnel selection literature (Edwards, 1953; Morgeson et al., 2007; Ziegler et al., 2011a), yet others contend this view and, in contrast, attach substantive meaning to social desirability. In this alternative perspective, social desirability is seen as a substantive trait, that is a stable individual difference variable with important consequences for outcomes in and beyond selection procedures. Different substantive interpretations of social desirability exist (e.g., He & Van de Vijver, 2013; Uziel, 2010a), but at the core lies that it relates to how one interacts with other people and how one presents himself or herself to others.

Three different streams of research relevant to the field of personnel selection have suggested that social desirability might be a stable characteristic of a person related to knowing what to do and how to behave in social situations. First, this substantive view is consistent with recent findings that a sizeable general factor can be found in personality measures, which reflects the socially desirable poles of traits such as the Big Five. This factor is labeled the General Factor of Personality (GFP; Figueredo et al., 2004; Musek, 2007) and appears to be a robust factor related to emotional intelligence (e.g., Van der Linden et al., 2016). As such, it has been labeled a social effectiveness factor or simply a general social desirability factor (Bäckström, 2007; Van der Linden et al., 2017). The second stream of research is formed by studies showing that social desirability scales – traditionally developed to detect whether people are engaging in response distortion – appear to measure valid trait variance, as indicated by moderate overlap between self- and other-ratings of social desirability scale scores (around $r = .30$), and moderate correlations between self-report social desirability ratings and other-ratings of personality (around $r = .40$; De Vries et al., 2014; McCrae & Costa, 1983; Paulhus, 1991). The third and final stream of research has revealed that knowing what to do in selection situations, that is acting in socially desirable ways, reflects the social skills of an applicant; social skills which can of course also be applied in people's daily lives and on the job (e.g., Kleinmann et al., 2011).

In light of the accumulating evidence from the three streams described above, this dissertation aimed to contribute to the support for the substantive interpretation of the social desirability factor present in self-report personality questionnaires. By doing so, this dissertation

aimed to contribute to the broader debate – substance versus artefact – on social desirability in selection situations and beyond. The main research question of this dissertation thus stated:

> *Research Question: Can social desirability in self-report personality questionnaires be regarded as a substantive factor in personnel selection?*

Four empirical studies were presented on the construct validity, antecedents, and criterion validity of social desirability as measured by the socially desirable component in personality inventories or highly related traits (i.e., trait emotional intelligence). In the present final chapter, the main findings of the current dissertation and how they relate to the four specific research questions are successively summarized and discussed which, collectively, will provide answers to the main research question formulated above. Subsequently, the strengths and limitations of these studies, as well as practical implications and suggestions for future research are described. The chapter ends with a general conclusion.

Summary of Main Findings

Criterion Validity

> *Research Question 1: Does social desirability in self-report personality questionnaires predict work outcomes?*

In the study presented in **Chapter 2**, data from multiple meta-analyses were combined in order to test the associations between the GFP and several other-rated, or objectively measured work-related outcomes. If the GFP reflects one's tendency for socially desirable behavior, then it should positively influence the way one interacts with colleagues, customers, clients, and supervisors. Presumably, this will also result in better performance ratings or in higher objective performance, such as more sales. In addition, higher GFP-levels increase the chance of being selected or accepted as a leader by being better able to do what is required in social situations (Van der Linden et al., 2014a; Van der Linden et al., 2016). Therefore, we hypothesized positive relations between the GFP and leadership, organizational citizenship behavior, and job performance (Burns et al., 2017; Sitser et al., 2013; Van der Linden et al., 2010a), and a negative relationship with counterproductive work behavior.

A relatively strong relation was found between the GFP and an other-rated overall measure of job performance (r = .31), especially when compared with Conscientiousness (r = .27) and Emotional Stability (r = .13), which are commonly regarded as the most relevant personality traits in terms of job performance (Barrick & Mount, 1991) – although the values are not particularly large in terms of absolute effect sizes (Cohen, 1988). The same pattern was observed for the relations with other work-related criteria (leadership, organizational citizenship behavior, and

counterproductive work behavior); the absolute corrected correlations between the GFP and the criteria ranged from .13 (performance for professionals) to .49 (leader emergence), with a mean of .34. In comparison, the mean absolute corrected correlations between Conscientiousness and Emotional Stability and these criteria were .26 (range .12–.42) and .15 (.02–.26), respectively.

Relatively high associations were found between the GFP on the one hand, and team performance, training performance, and the criteria related to leadership. Team performance relates to cooperativeness, the quality of interpersonal relations, and the ability to work with others. Trainings often include social-evaluative settings (Uziel, 2010a), in which interpersonal competences can be expected to play an important role (Viswesvaran et al., 2001). Finally, as noted, the GFP has previously been linked to leadership (e.g., Van der Linden et al., 2014), based on the argument that those who show more socially desirable and socially effective behavior are more likely to be appointed and accepted as leaders by peers. These results are thus in line with the theoretical substantiations of the GFP as a social effectiveness factor evolved through pressures for pro-social behavior (Van der Linden et al., 2016).

Taken together, the aforementioned results positively answer Research Question 1. The finding of different correlations between the GFP and performance across specific job types was further informative for this first research question. The lowest relation (corrected $r = .13$) was found for professional jobs, a group that included engineers, architects, and accountants. It could be argued that such jobs are characterized by a set of abilities and knowledge which have a smaller social component in daily work tasks. Consequently, the relation between a factor representative of social desirability (i.e., the GFP) would have less influence on the performance in these jobs as rated by others. The strongest relation was found for managerial performance (corrected $r = .31$), which is in line with previous studies that have linked social effectiveness with performance in managerial positions (Semadar et al., 2006). Theoretically this makes sense, as we know from previous studies that to be successful and accepted in managerial positions, one needs to combine both *getting ahead* and *getting along* (Hogan & Shelton, 1998; Marinova, Moon, & Kamdar, 2013; Semadar et al., 2006). In reference to Research Question 1, it thus appears that social desirability indeed predicts other-rated or objective work outcomes, but in some job types more than in others. These results are in line with the finding that scores on impression management scales show criterion validity in some jobs, but not in others (Ispas et al., 2014).

Importantly, two characteristics of the study described in **Chapter 2** render the artefactual account of the GFP or social desirability less plausible. First, the criteria were based on other-ratings or objective outcomes, which precludes the influence of common-method bias on the results. Second, the relation between the GFP and job performance was investigated while controlling for social desirability as measured by the *Balanced Inventory for Socially Desirable Responding* (BIDR; Paulhus, 1984). Specifically, the relation between the GFP on the one hand, and self-deceptive enhancement (SDE) and impression management (IM) on the other were strong ($r = .66$ and $r = .55$, respectively). Yet, partialling out the effect of SDE and IM on the relation between the GFP and job performance had a negligible effect: the GFP-performance relation ($r = .31$) increased only slightly

when SDE was controlled for (r = .33), while controlling for IM resulted in a slight attenuation (r = .29).

This finding is informative for the discord on what social desirability scales exactly measure: although social desirability scales were traditionally developed to be measures of faking on personality questionnaires, recent results suggest that these scales partly capture valid (personality) trait variance. That is, if social desirability scales are interpreted as measures of bias, the results indicate that removing this bias has little influence on the associations found, supporting the substantive interpretation of the GFP. However, taking the substantive perspective, the combination of findings (i.e., strong GFP-social desirability scale overlap, but negligible attenuation of the GFP-job performance relation by social desirability scale scores) suggests that social desirability scales may capture *redundant* variance (Smith & Ellingson, 2002), without added value for the prediction of job performance beyond the GFP. Given the typically small relations found between social desirability scales and job performance (meta-analytic corrected r = .10 for SDE and r = .12 for IM; Li & Bagger, 2006), the reverse pattern is less plausible (i.e., the GFP is redundant with respect to social desirability scales).

The sizeable associations between the GFP on the one hand, and SDE and IM on the other, support the notion that social desirability – as measured by social desirability scales – resides at the higher levels of the personality trait hierarchy, just as the GFP (Connelly & Chang, 2016; Paulhus & John, 1998). Previous studies have identified the meta-factors Stability (the combination of Conscientiousness, Agreeableness, and Emotional Stability) and Plasticity (Openness and Extraversion), which are proposed to lie at an intermediate level between the Big Five domains and the GFP (Digman, 1997; DeYoung et al., 2002). Stability and Plasticity are reflected in the two-dimensional (*egoistic* vs. *moralistic*) model of social desirability proposed by Paulhus and John (1998). Stability conceptually aligns with egoistic bias and reflects the tendency of stressing one's exceptional qualities and social and intellectual status. Plasticity aligns with moralistic bias reflecting tendencies related to claiming to have an overly large ability to control malignant impulses. In socio-analytic theory, which has been applied to applicant faking (e.g., Blickle et al., 2009; Hogan & Blickle, 2018; Ingold et al., 2015), Stability and Plasticity are referred to as the factors that signify one's motives for *getting along* and *getting ahead*. As a social effectiveness factor, the GFP reflects the combination of *getting along* and *getting ahead* (Irwing, Booth, Nyborg, & Rushton, 2012). Consequently, the finding that the associations between the GFP and social desirability scales are high – and higher than those found for the Big Five (meta-analytic *r*s between .03 and .42 for IM, and *r*s between .19 and .54 for SDE; Li & Bagger, 2006) – provides support for the location of social desirability scales at higher levels in the personality trait hierarchy.

A relevant finding from the study presented in **Chapter 2** for the personality literature was that the Big Five dimensions added little unique variance in the prediction of the outcomes beyond the GFP. The results imply that an important part of the presumed criterion validity of the Big Five dimensions might actually be due to GFP, i.e. the shared variance among the Big Five dimensions. Note that testing the other way around, that is whether the GFP adds unique explained variance

on top of the Big Five dimensions is less meaningful. After all, the GFP is assumed to be partially present in each of the specific personality dimensions. Therefore, controlling for the latter means eliminating the true variance of the general factor in the outcomes.

It does not, however, imply that the Big Five dimensions lose their merit as important predictors for work-related outcomes. As noted, each Big Five dimension is made up of variance due to the general factor and variance unique to the specific dimension (e.g., Extraversion); some criteria will be better explained by the component related to social effectiveness (i.e., the GFP), while other criteria are better predicted by the unique variance(s) of specific Big Five dimensions. This phenomenon is known as the bandwidth-fidelity trade-off: predictions of criteria tend to be optimal when predictor and criterion are conceptually aligned at the same level (Ones & Viswesvaran, 1996). Sitser et al. (2013), for example, showed that specific performance dimensions (e.g., handling customer complaints) were predicted by (facets of) Agreeableness but not by the GFP.

If the GFP indeed reflects social effectiveness, then one could hypothesize that for criteria which are more 'loaded' with social effectiveness, the role of the GFP should be larger and hence the relations between the unique variance of the Big Five dimensions and these criteria should be relatively small whereas relations with the GFP should be comparatively large. To a large extent, this idea could be confirmed in **Chapter 2**. Results showed that the importance of the GFP compared to the Big Five was relatively large for organizational citizenship behavior, which entails showing socially desirable behavior towards colleagues or the organization. At the same time, for reasons outlined above, the relative importance of the GFP compared to the Big Five was relatively small for job performance in professional jobs. Comparable findings are found in the literature on emotional intelligence, a construct which is conceptually and empirically similar to the GFP (Van der Linden et al., 2017). The relation between EI and job performance is stronger in jobs requiring more 'emotional labor' (Joseph & Newman, 2010; O'Boyle et al., 2011), i.e., in which controlling one's emotions and dealing with other people is crucial (e.g., customer service, health care, police). Similar findings are reported in **Chapter 2** for the GFP, in that it showed sizeable associations with performance for police officers (r = .29) and with leadership (rs between .32 and .49), a position that is also defined by a high level of emotional labor (Humphrey, Pollack, & Hawver, 2008).

Research Question 2: Does social desirability in self-report personality questionnaires relate to (daily) criteria of social effectiveness?

Research Question 1 and **Chapter 2** specifically focused on whether relations between social desirability, as captured by the GFP, and work-related outcomes could be found. Research Question 2 extended the scope of this question by asking to what extent social desirability is reflected in people's everyday social lives. That is, if the socially desirable component in self-report personality questionnaires represents social effectiveness, then this should be reflected in people's daily (social) experiences. Ultimately, if this component (i.e., the GFP) is related to knowing what to do and say

and how to act in social interactions, then we can expect this to positively influence the social relationships with others, and presumably, lead to higher levels of well-being because of this.

This idea was tested in the study described in **Chapter 3**. The study was based on secondary data (the Berlin Diary study by Denissen and colleagues, 2005 – 2008) of a sample of 1,223 German participants who completed a diary for 25 days in which they daily reported their levels of positive affect, negative affect, and self-esteem, as well as a number of daily social experiences and events. This large sample was a strong asset of the study, although the sample was skewed in terms of gender (1,055 women, 86%). The average age was 29.47 (SD = 10.49, ranging from 13 to 72 years old). The sample was more mixed in terms of marital status (39% were single, 40% in a steady relationship without being married or engaged) and education (about 50% of the sample was relatively highly educated). Before completing the diaries, the respondents completed two Big Five measures from which we extracted the GFPs. Although the data were collected more than 10 years ago, there is no reason to assume that the relationship between fundamental constructs such as personality, well-being, and social relationships has changed since then.

Therefore, the data was found eligible to test our three hypotheses. First, it was hypothesized that the GFP would show relations with daily social experiences (relationship quality, impressions made on others, interpersonal conflict and relationship quantity), being indicators of social effectiveness. Second, it was hypothesized that the relation between the GFP on the one hand and self-esteem, positive affect, and negative affect on the other would be mediated by daily social experiences and events. Third and finally, it was expected that the GFP would moderate the associations between the daily reported social events and daily levels of well-being and mood in such a way that the effects would be stronger when GFP-levels were low.

The hypotheses were largely supported in the study presented in **Chapter 3**. In line with the first hypothesis, small negative (r = -.08 for both GFP measures) relations were found between the GFP and interpersonal conflict, and moderate to large positive associations with relationship quality (r = .33 for both measures) and the impressions made on others (r = .45 for the BFI-based GFP, r = .38 for the FIRNI-based GFP). With respect to the second hypothesis, the most important mediators of the relations between the GFP and all three outcomes were relationship quality (indirect / total effect ratio between 17% and 28% across the two GFP measures) and daily impressions made on others (indirect / total effect ratio between 16% and 49%). However, the effect of the GFP on self-esteem and mood did not appear to be substantively mediated by the number of interpersonal conflicts. The third hypothesis, on moderation, was supported for self-esteem and negative effect, but not for positive affect. One of the reasons for this latter unexpected finding that was mentioned in the chapter was that the average positive affect levels of the participants were more moderate (M = 2.87 on a 5-point scale) compared to their average self-esteem and negative affect levels. That is, the average self-esteem level of the sample was relatively high (M = 3.89), while the average negative affect level was relatively low (M = 1.82). At the more moderate positive affect levels, it may matter less whether one reacts to daily experiences in a socially desirable manner in terms of affective reactions.

Overall, the results from the study described in **Chapter 3** suggest that higher GFP scores allow for easier navigation through the social world with better daily social interactions and higher quality relationships with others. Given that social relationships are strong indicators of social effectiveness (Argyle, 2001; Denissen et al., 2008c), this chapter provides evidence for the hypothesis that the GFP indeed reflects social effectiveness. It also provides an explanation for why GFP scores are found to be associated with higher levels of adaptation in terms of well-being (e.g., Musek, 2007): social relationships are considered as highly important for people's well-being (Argyle, 2001), and as such, the higher quality of social interactions associated with the GFP appears to be an important mechanism for the GFP-well-being relationship. Finally, the finding that the GFP is related to better affective reactions to social interactions further supports the notion that the GFP is a social effectiveness factor. The finding of reduced reactivity due to the GFP is in line with the idea that the social skills underlying the GFP allow for adequate reactions to (social) setbacks (Hengartner et al., 2017) such as conflicts, reducing fluctuations in mood, and thus allowing one to be better adjusted on the whole.

With respect to the context of selection and assessment, a relevant finding in the study reported in **Chapter 3** was that, compared to the other variables tested, the GFP was relatively strongly related to leaving favorable impressions on others (around $r = .40$). In addition, leaving good impressions on others was one of the most important mediators of the relation between the GFP on the one hand, and self-esteem, negative affect and positive affect on the other. These findings underline the interpretation of GFP as a social desirability factor, but in a veridical sense: presenting oneself favorably appears to be an important part of personality, with important consequences in terms of well-being, but presumably also in terms of success in one's working life. In other words, the findings showing a positive relationship between the GFP and positive impressions on others fit with those from previous studies showing that higher GFP scores are associated with better evaluations in selection procedures, and with the findings from **Chapter 2** and Van der Linden et al. (2010b) that higher GFP scores lead to higher ratings of job performance by supervisors. As noted previously, results from studies on the relation between social desirability scales and personality also suggest such a substantive interpretation of social desirability or impression management (e.g., McCrae & Costa, 1983; Uziel, 2010a; Roth & Altmann, 2019).

Antecedents

Research Question 3: Does social desirability in self-report personality questionnaires in the selection context relate to social competences?

The idea of social desirability as a substantive construct rests on the assumption that it is predicated on social skills or competences: that is, in order for one to behave in socially desirable ways, one needs to know which behavior is adequate given a certain (social) situation, and the social skills to be able to show this type of behavior (Argyle, 1969; Hogan & Shelton, 1998; Marcus, 2009; Roulin et

al., 2016). In the study described in **Chapter 4**, we tested this idea in the domain of personnel selection.

Applied to the selection context, providing desirable responses involves reading social and emotional cues, interpreting the information provided by the cues, and acting on them in appropriate ways (Roulin et al., 2016). One way of testing whether social competences are indeed an antecedent of social desirability is thus by investigating whether social competences are indeed related to providing the desirable responses (labeled *the ability to fake*) in selection procedures. This was done in the study in **Chapter 4**, in which social competences were measured by trait emotional intelligence (EI) since this construct entails understanding one's own emotions and those of others and the tendency to use this knowledge act in socially effective ways (Mavroveli et al., 2007; Petrides, 2011). Trait EI has been found to show a large amount of overlap with the GFP (meta-analytic corrected $r = .86$; Van der Linden et al., 2017).

An experimental within-subject design was used to test whether trait EI was positively related to providing the desired responses in selection procedures. A sample of 129 undergraduate students completed a personality inventory multiple times. First, they were instructed to answer the items of the inventory in an honest fashion. In this 'honest' condition, they also completed a trait emotional intelligence questionnaire. Subsequently, two scenarios were presented in which they were asked to fill out the personality questionnaires as to maximize their chances of obtaining the jobs of file clerk and lawyer. It was hypothesized that personality facet scores would shift into the direction of the prespecified job profiles under the instruction of 'chance maximization' – indicating faking ability – and that EI was positively related to this ability. In addition, it was hypothesized that EI would show incremental validity over the Big Five personality traits and general mental ability in the prediction of faking ability.

Results supported both hypotheses: although relatively small in terms of effect size, trait emotional intelligence predicted the ability to fake in both scenarios to a similar degree. In addition, trait EI scores explained a unique part of variance in the ability to fake, over and above the Big Five and general mental ability ($\Delta R^2 = .03$ in the file clerk condition and $\Delta R^2 = .04$ in the lawyer condition).

An interesting other finding was that the ability to fake towards the file clerk profile and towards the lawyer profile showed a moderate correlation ($r = .40$). This result has important implications for the predictive validity of personality questionnaires in the selection context. It suggests that some people are consistently better at self-presentation than others (Ingold et al., 2015; Klehe et al., 2014; McFarland & Ryan, 2000). **Chapter 4** further showed that trait EI might be part of the reason why some people are better at presenting themselves than others, as indicated by the positive effects on the ability to fake in both scenarios. However, this effect of trait EI will most probably not be limited to selection procedures only. Following the thought that trait EI reflects genuine social competence or social skills, we can expect it to have positive effects beyond the selection procedure, for example on the job by knowing how to (inter)act with colleagues and supervisors. Studies have indeed shown that trait emotional intelligence is positively related to job

performance (Côté & Miners, 2006; Joseph et al., 2015; Joseph & Newman, 2010; Pekaar, Bakker, Born, & Van der Linden, 2018; Pekaar, Van der Linden, Bakker, & Born, 2017).

Thus, in more general terms, self-presentation may be regarded to reflect general social skills that people can also use during their jobs or everyday social interactions. Consequently, instead of considering self-presentation a bias that distorts the validity in predicting performance, it may actually enhance the predictive validity of personality measures. This idea is supported by empirical evidence from previous studies finding that controlling for social desirability can actually reduce the criterion validity of personality questionnaires (e.g., McCrae & Costa, 1983; Viswesvaran & Ones, 1999). It should also be noted here, however, that other studies do find that controlling for social desirability increases the criterion validity of personality test scores (e.g., Douglas et al., 1996; Ziegler, MacCann, & Roberts, 2011b); clearly, more research is needed on this topic. Based on the results from **Chapter 4**, however, we would suggest that social skills should be taken into account when the relations between personality and criteria are investigated, because social skills will influence these relations (see Witt & Ferris, 2003, and Blickle, Wendel, & Ferris, 2010, for empirical evidence for this claim).

Additional evidence for the argument that the socially desirable component in self-report personality questionnaires is based on social skills comes from studies on the ability to identify criteria (ATIC) in selection procedures, a concept similar to trait EI, although ATIC is regarded as a cognitive concept (Kleinmann et al., 2011). ATIC has been found to relate to both performance in selection procedures across different assessment types (e.g. personality, assessment center exercises and interviews), and to be related to job performance (Kleinmann et al., 2011). In addition, ATIC research has shown that disclosing what constructs are measured in the selection procedure increases the construct validity of the selection procedure, while at the same time decreasing its criterion validity (Kleinmann et al., 2011). This finding is what we would expect if social skills related to decoding what behavior is required form the explanation for both performing well in selection situations as on the job.

The most direct evidence for the idea of social skills (in the form of ATIC) underlying both performance in selection procedures and at work comes from a study by Klehe et al. (2012), who found the so-called "ideal employee factor" (Schmit & Ryan, 1993; i.e., in our view the GFP) to be positively related to job performance, a relation that could be explained by ATIC. In other words, their findings show that the general factor is predicated on social skills, which allows for higher performance on the job. In sum, the results from **Chapter 4**, and the findings by Klehe et al. (2012) seem to indicate that knowing what to say and do in selection procedures, but also more generally speaking acting in socially desirable ways, requires social (EI) and/or more cognitive (ATIC) skills which one can also use in one's everyday lives to attain one's (social) goals.

Construct Validity

Research Question 4: Does social desirability in self-report personality questionnaires vary with the test-taking context and item format?

Social desirability has predominantly been regarded as a factor resulting from a person x situation interaction (e.g., Ziegler & Buehner, 2009). In some situations, for example in selection procedures, some individuals might be more motivated to respond in a socially desirable way than others, compared to in other situations. In fact, this argument has been brought forward for why the GFP is in fact spurious: because individual differences in socially desirable responding adds uniform systematic variance to all (presumably) independent Big Five traits, this type of responding would drive all Big Five dimensions into the same direction leading to the emergence of a general factor (Bäckström, 2007; Bäckström et al., 2009; Petterson et al., 2012). Following this argument, it can be expected that a measure of social desirability *as a bias* might be more apparent, i.e., that the general factor is larger in situations where people are motivated and have the opportunity to present themselves in socially desirable ways.

In the study presented in **Chapter 5**, this idea was tested by investigating the influence of test-taking context (career advice vs. selection, the latter being known to induce a higher motivation to self-enhance; e.g., Birkeland et al., 2006) and item format (Likert vs. forced-choice (FC), the latter being known for reducing the opportunity to self-enhance; e.g., Christiansen et al., 2005; Jackson et al., 2000) on the general social desirability factor in self-report personality questionnaires. As such, it was investigated whether the properties of the social desirability factor are dependent on external factors; if the general factor changes considerably due to the test-taking context and item format, then it is less plausible that scores on this social desirability factor are representative of the extent to which someone truly possesses such traits. Alternatively, it was hypothesized that the general factor in personality self-report questionnaires is robust to external factors and thus would not change as a function of test-taking context and item format. Thus, this chapter focused on the construct validity of the socially desirable component in self-report personality questionnaires.

A fully crossed 2x2 design (development forced-choice, selection forced-choice, development Likert, and selection Likert) was used. Previous studies have mainly focused on either motivation (e.g., Anglim et al., 2017; Ellingson et al., 2001; Schmit and Ryan, 1993; Smith & Ellingson, 2002) or opportunity to respond in a socially desirable way (Christiansen et al., 2005; Jackson et al., 2000; Vasilopoulos et al., 2006). Alternatively, studies focused only on lower order traits such as the Big Five rather than on higher-order factors such as the GFP (e.g., Joubert et al., 2015). As such, the fully crossed design described in **Chapter 5** formed a contribution to the extant literature. Furthermore, because in previous studies differences in results were found between real-life selection procedures and experiments in which participants are instructed to 'fake' (e.g., MacCann et al., 2017; Schermer et al., 2019a; Schermer et al., 2019b; Van der Linden et al., 2011), a vital feature

of the current study was the use of data from real candidates undergoing career advice or selection procedures.

These data were collected between July 2011 and March 2015 by a large psychological test development and publishing firm specifically tailored to the HR market Responses of 3,980 candidates distributed over the four groups in our design (development FC, selection FC, development Likert, and selection Likert) were analyzed. Each condition consisted of 995 participants, matched on gender, age, and educational level. The personality questionnaire used was the Work-related Personality Questionnaire (WPQ; Ixly, 2012), an instrument that has its theoretical foundation is the Big Five model. Both the Likert and FC version of the WPQ yield 25 facet scores forming five domain factors.

The three main findings of the study can be summarized as follows. First, scores on facets, domain factors, and general factor were generally higher (thus more in the socially desirable direction) in the selection context than in the development context, with the difference being larger for the general factor than for the domain factors and facets. Second, the mean differences between the development and selection context were smaller for the FC questionnaire than for the Likert-scale questionnaire. At a first glance, these two results combined seem to suggest that the general factor in personality questionnaires at least partly reflects a 'faking factor'. Given that score differences occurred due to motivational context, and that the effect of motivations to distort responses could be reduced by limiting the opportunity to do so by means of the item format, it appears that social desirability is partly the result of the characteristics of given situations and circumstances.

However, the third main finding was that, despite mean differences, the lower order factors and general factor were highly similar across the four groups, both in terms of their content and their importance (size). This similarity implies that the content and between-person rank order across traits remained largely the same across contexts and item formats. Specifically, if one would assume that the larger part of the social desirability factor was due to unrealistic self-enhancement – which should be most prominent in the selection context and limited by the forced-choice format – then it could be expected that the general factors would differ between the groups, which was not the case.

The answer to Research Question 4 is thus mixed in the sense that social desirability, at least as measured by the general factor in self-report personality questionnaires, appears to be a mix of valid trait variance and variance due to measurement biases. Yet, decades of research on organizational research methods have taught us that any self-report measure captures both veridical trait variance and method variance (Cote & Buckley, 1987; McCrae, 2015, 2018). In this sense, the GFP does not differ in any way from, say, the Big Five personality traits. In a recent study, different possible sources of GFP variance (socially desirable response bias, positive self-evaluation and social effectiveness) were investigated (Dunkel et al., 2016). A relatively large share was attributable to social effectiveness, which was measured by calculating the similarity of a person's

personality profile with a highly socially effective profile as judged by three independent raters (Dunkel et al., 2016).

It should also be noted here that the fact that scores on a social desirability factor can be increased in selection contexts does not automatically mean that such a factor cannot have predictive value or be substantiated by other-reports (Chen et al., 2016). In fact, if the GFP is a measure of social desirability, then we *should* expect it to be activated and thus scores to be higher in situations in which people are deemed to look as desirable as possible (i.e., in selection procedures). This effect of the selection procedure on observed scores can be expected to be smaller for the Big Five domains, because these are composed of variance due to social desirability (i.e., the GFP) and specific variance related to the specific domain (e.g., Extraversion). In any case, evidence that GFP scores are inflated or that the general factor becomes larger in the selection context in itself does not lead to the conclusion that a 'true' GFP does not exist; the question is whether it still meaningfully relates to (external) criteria.

The arguments described in the previous paragraph are also relevant for a recent stream of studies showing that the properties of the GFP do change because of the effects of applicant faking (MacCann et al., 2017; Schermer et al., 2019a, 2019b). The results from these studies are in contrast which those found in the study described in **Chapter 5**. However, these studies are all conducted in the lab, in which participants are instructed to fake, i.e., effectively asked to deliberately lie.

There are multiple reasons why the results from such instructed laboratory studies might differ from studies conducted in the field. First, such situations do not reflect real-life selection situations in which people are perhaps more subtle in the way they present themselves favorably, and in which applicants also have a lot to lose when deliberate lying is detected (Marcus, 2009). Second, instructed faking studies reflect maximum faking behavior (Smith & Ellingson, 2002), i.e., how much people *can* fake. Note that in **Chapter 4**, an instructed faking design was also used; however, the goal of the study in **Chapter 4** was to find whether people were *able* to fake. As such, the faking instruction facilitated the creation of an ability or maximum performance measures. Yet, when looking at the properties of the GFP, or personality measures in general, it is more informative to investigate to what extent they *are* affected by circumstances than to what extent they *can* be affected, under the least favorable conditions (Smith & Ellingson, 2002; Smith & McDaniel, 2011). Furthermore, finding higher GFP scores under faking instructions *in itself* does not provide evidence against a substantive interpretation; in fact, again, when we explicitly ask people to present themselves as favorably as possible, we would expect it to be reflected in a factor that is proposed to capture social desirability.

Strengths and Limitations of the Four Empirical Studies

The empirical studies presented in the previous chapters contribute to both the GFP literature specifically and the personnel selection literature in a number of ways. The first strength of this dissertation is that it adds to the growing body of empirical evidence in favor of the hypothesis that

the GFP represents social effectiveness. As a general social desirability factor, it can be expected to exert a broad influence on the daily lives of people, both in and outside the selection and work context. In three of the four empirical studies (**Chapters 2, 3,** and **5**), the GFP proved to be a stable personality trait related to daily social interactions, well-being, and a wide range of work-related outcomes. A second strength of this dissertation was that the study designs reduced possible alternative explanations, such as common method bias or socially desirable responding, for example because objective or other-rated criteria (**Chapter 2**) or methods to reduce response distortions (**Chapter 3**) were used. Taken together, based on the results presented in the current dissertation, it appears likely that the GFP represents a substantive trait rather than a statistical or methodological artefact associated with the use of personality questionnaires.

Second, three studies presented in this dissertation used responses from large and relatively heterogeneous samples, collected in communities and real-life selection and assessment situations. In **Chapter 2,** data from previously published meta-analyses were used on which structural models were tested; the number of studies in these original meta-analyses ranged between 6 and 239, and the total number of respondents between 1,144 and 144,117. The study in **Chapter 3** used a German community-sample of 1,223 respondents with more than a total of 20,000 daily reports, while the study described in **Chapter 5** was based on a total sample of 3,980 real selection and assessment candidates. In both the faking and GFP literature, findings diverge between studies conducted in the lab with students, and studies based on real-life applicants (MacCann et al., 2017; Schermer et al., 2019a; Schermer et al., 2019b; Van der Linden et al., 2011). It is therefore important that field studies such as the ones presented here are conducted, also from the perspective of ecological validity and generalizability of the obtained results.

A final strength of the current dissertation lies in the methodological advances it brings to the social desirability and faking literature, as well as the GFP literature. Different and novel statistical methods were used, and multiple analytical strategies were employed to verify the obtained results. For example, in **Chapter 5**, a novel method based on item response theory was used to extract normative data from forced-choice questionnaires (i.e., the Thurstonian IRT model; Brown & Maydeu-Olivares, 2011, 2012), after which exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) was conducted, a relatively new framework which relaxes the unrealistic assumptions of traditional confirmatory factor analysis (CFA). Several methods for detecting careless responding proposed by Meade and Craig (2012) were carried out in **Chapter 4,** and polynomial regression analyses (Edwards, 1994) were conducted in order to avoid the problems associated with the use of difference scores. In addition, to the best of our knowledge, the current dissertation presents the first study in which the GFP is used in multilevel analyses to investigate its relations with daily social experiences and well-being (**Chapter 3**). Finally, in **Chapter 2**, different competing structural models were tested on meta-analytic correlation matrices. In fields of research were controversy is high, as is the case for both social desirability and the GFP, it is of importance that research is conducted in a rigorous fashion so that conclusions are not based on

methodological fallacies. In this dissertation, such pitfalls were tried to be avoided by the use of different analytical strategies described above.

Some limitations of the studies in this dissertation are worth mentioning. The first limitation is that three out of four empirical studies were based exclusively on self-reports, introducing possible effects of common method variance on the results (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, but see also Spector, 2006). This issue might be less problematic than it appears, as the main question of this thesis was whether the socially desirable component of *self-report* measures of personality can be interpreted as a factor representing social effectiveness. For this reason, it seems less problematic that self-report measures of personality were used in **Chapter 2** and **Chapter 5**. In **Chapter 4**, self-report trait EI was used as a predictor of faking ability. Although faking ability was manipulated in an experimental design, the question remains whether the results hold when other-reported trait EI is used as the predictor. In **Chapter 3**, all measures of personality, well-being and daily events were based on self-reports, so in this case, concerns for the influence of common method variance are justified. However, the use of diary data is assumed to alleviate at least some of the problems associated with self-report measures, in particular recall bias and more importantly, social desirability bias (Barta et al., 2013). Furthermore, by person-mean centering our daily variables, in other words by looking at fluctuations in daily variables, individual differences in response tendencies are eliminated (Beal, 2015), reducing the influence of common method biases (see Derks, Bakker, Peters, & Van Wingerden, 2016). In addition, it is implausible to find cross-level interaction such as those found in **Chapter 3** under circumstances with high levels of common method bias present (Lai et al., 2013).

In sum, regarding the limitation of using self-reports, it appears unlikely that common method bias associated with the self-report measures has affected the results presented in **Chapter 3**. Still, by exclusively using self-reports, our measures actually reflect *perceived* events and actions. The same event can be interpreted in different ways by different individuals (e.g., Von Dras & Siegler, 1997). This may have affected our results, especially when our variables of interest (e.g., the GFP) influence how events are perceived; for example, it might be that lower GFP scores are associated with being selected in or interpreting a social interaction as an interpersonal conflict (e.g., Bolger & Schilling, 1991; Larsen, 1992). A replication of the study using multiple sources of information would be desirable in the future to address this limitation.

A second limitation relates to the samples used in the presented studies. Although some characteristics of the samples were strong assets, such as their size and being obtained from the field (**Chapter 5**), other characteristics were less than ideal. For example, the samples were relatively highly educated in **Chapter 3** and **Chapter 5**, limiting the generalizability of our findings. In addition, in **Chapter 3**, the majority of the sample consisted of women (86%). The same limitation in terms of generalizability holds for **Chapter 4**, in which a student sample with faking instructions was used. An additional limitation of the student sample is the restriction of range in terms of cognitive abilities. This restriction of range was possibly responsible for not replicating the relation between cognitive ability and faking ability from previous studies (e.g., Tett et al., 2012; Vasilopoulos

& Cucina, 2006). Thus, this study needs to be replicated in a field study with real applicants. However, this will probably be hard to achieve. Multiple personality measurements at different time points are needed; one where people are assumed to respond honestly, and other measurements when these people apply for different jobs.

In terms of limitations, a third and final note about causality should be taken into account when interpreting the presented results. Certain aspects of the studies presented in this dissertation contribute to robustness of the relations found (e.g., using meta-analytic or diary data). Yet, in some cases, the relations found are cross-sectional in nature. For example, in the diary study presented in **Chapter 3**, the daily variables all were measured at the same time at the end of the day, rendering it impossible to establish the causal order of the associations found. In this case, for conclusions about causality of the relations, a design in which predictor and outcome variables are separated in time would be needed.

Practical Implications

This dissertation has shown support for the notion that the general factor in personality measures is stable across situations that elicit socially desirable responding to different degrees. The dissertation also provides evidence for relatively strong relations between the GFP and other-rated or objectively measured work-related outcomes such as job performance. Based on these findings together, a first practical implication of this dissertation is that practitioners and recruiters might want to include a GFP measure in their selection procedure. On a general level, GFP scores provide a good indication of the level of well-being of a person (e.g., Musek, 2007). As such, a person's GFP score reflects his or her level of adaptation to his or her social environment (Dunkel, 2013) and might thus indicate to what extent someone is affected by (social) challenges (Dunkel, Van der Linden, & Kawamoto, 2019; Van der Linden et al., 2016; Hengartner et al., 2017); having such information would generally be regarded as useful in selection situations. Note that a GFP score should not *replace* Big Five domain scores or facet scores in selection procedures; in many cases, applicants would probably like to see and discuss a detailed report and nuanced view of their personality in their selection procedure, rather than a single score. Seeing such an elaborate profile rather than a single score would probably also contribute to the face validity and perceived fairness of the application procedure (Hausknecht, Day, & Thomas, 2004). Yet, a GFP score could be a useful additional piece of information, and practically, obtaining a GFP score is a relatively easy task given that it can easily be extracted from the Big Five or any other personality questionnaire commonly used in selection procedures.

The current dissertation has shown that the assessment of someone's GFP-level might be more valuable for some jobs than for others. For jobs in which social competences and skills are required, obtaining a GFP score on top of the Big Five domain scores would have added value. At the same time, interpersonal contacts such as dealing with colleagues, supervisors, and clients are part of virtually any job, especially since working in teams has become increasingly popular in the

world of work (e.g., Morgeson, Reider, & Campion, 2005; Peeters, Van Tuijl, Rutte, & Reymen, 2006; Salas, Cooke, & Rosen, 2008). Therefore, practitioners might always want to opt for a measure of social effectiveness to inform their hiring decisions. In addition, social skills are commonly regarded as an important part of 21st century skills which are becoming increasingly popular in the future (Ananiadou & Claro, 2009); as an organization, it may thus be beneficial to get an idea of the interpersonal skills of an applicant (as indicated by one's GFP score).

Second, this dissertation furthermore implies that attempts to control for social desirability in selection procedures will most likely be futile or even have negative consequences during personnel selection. Apart from some exceptions (e.g., Douglas et al., 1996; Donovan et al., 2014), research shows that efforts to control for socially desirable responding by statistically removing variance captured by social desirability scales typically has a negligible effect on the criterion validities of personality measures (Li & Bagger, 2006; McCrae & Costa, 1983; Ones et al., 1996). This is in line with research on social desirability scales showing that these tend to measure consensually valid personality traits, rather than a response style (e.g., McCrae & Costa, 1983; Uziel, 2010a; Zettler et al., 2015). The current dissertation has provided evidence for a substantial interpretation of social desirability captured by the general factor in personality measures, also when administered in the selection context. Taken together, social desirability appears to be a substantive personality trait with important consequences for performance on the job; by controlling for this variance, part of what truly drives the relation between personality and performance on the job would be removed.

A third and final important implication for practitioners in personnel selection is the finding from the current dissertation that the factor structures of personality inventories are relatively unaffected by the item format. At the same time, it was found that forced-choice questionnaires do reduce the mean score differences between selection contexts and contexts in which less socially desirable responding is present. Practically, based on these results, this could mean that some practitioners might favor forced-choice questionnaires over traditional Likert questionnaires. The decision between a FC or Likert questionnaire may also be informed by results from other studies on this topic. A recent meta-analysis has shown that the criterion validities of FC questionnaires tend to be higher than those found for Likert questionnaires (Salgado et al., 2015). However, part of this higher level of predictive validity might be attributable to cognitive abilities since FC questionnaires are cognitively more demanding because they require the comparison of multiple statements at once (e.g., Vasilopoulos et al., 2006), and cognitive abilities are known to be a relatively strong predictor of job performance (Schmidt & Hunter, 1998). In addition, FC questionnaires have previously been found to be judged more negatively by applicants than Likert questionnaires (Converse et al., 2008). Predictive validity should in principle be the most important criterion for using a certain selection method indicating that FC questionnaires should be preferred, yet, practitioners will most likely keep all aforementioned factors in mind when choosing between a FC or Likert questionnaire for a selection procedure.

Suggestions for Future Research

The current dissertation provides strong evidence for the criterion validity of the GFP. Yet, the data on which this evidence was based were collected under low-stakes settings. In the future, it would be interesting to investigate whether the obtained results change when data from high-stakes settings (i.e., selection) are used. Many studies have tried to answer this question for the Big Five domains and HEXACO factors, without a definite answer (e.g., Anglim et al., 2018; Blickle et al., 2009; Donovan et al., 2014; Griffith et al., 2007; Ones et al., 1996; Ziegler & Buehner, 2009). Given the artefactual interpretation as the GFP as a *social desirability as a bias* ('faking') factor, for the GFP to be considered a substantive construct that can be used for personnel selection, it is crucial to investigate whether the criterion validity of the GFP holds under high-stakes settings. If the GFP is indeed mostly due to the fact that people provide unrealistic inflated self-ratings, then relations with work-related outcomes should diminish considerably when the GFP is extracted from self-ratings of applicants. Testing such hypotheses should not be a very difficult task. From any published study that has reported a complete correlation matrix of the Big Five and criteria collected from applicants, a GFP and its relations with work outcomes can be extracted. Preferably then, these correlations are pooled in a meta-analysis to diminish the effects of idiosyncrasies in individual studies and to arrive at robust estimates of the criterion validity of the GFP under high-stakes settings. Based on the mixed evidence found for the Big Five, is it not easy to predict what the outcomes of such a study would be; however, given the findings from the current dissertation and previous studies that 1) the GFP is robust to test-taking context and 2) shows criterion validity in applicant contexts (Van der Linden et al., 2011), it appears more likely that the relations between the GFP and job performance will be maintained under high-stakes settings.

Related to this point, it could be investigated whether the type of measurement instrument influences on the relations found between the GFP and criteria. The current dissertation has shown that the construct validity of personality questionnaires in general, and the general social desirability factor specifically, remains intact (also across low and high-stakes situations) when an item format (forced-choice) designed to reduce applicant faking is used. As noted previously (see *Practical implications*), a recent meta-analysis has shown higher predictive validities for FC questionnaires compared to traditional Likert questionnaires. However, there are other methods for reducing the effect of socially desirable responding found in the literature, for example rapid response measurement (RRM; Meade, Pappalardo, Braddy, & Fleenor, 2018) or the use of warnings (Converse et al., 2008; Fisher, Robie, Christiansen, & Komar, 2018; Vasilopoulos, Cucina, & McElreath, 2005), for which the effects on the properties of the GFP and its relations with external criteria are still unknown and therefore worth uncovering.

Another area for further research would be to connect research on the GFP and trait emotional intelligence on the one hand and research on the ability to identify criteria (ATIC; Kleinman et al., 2011) in selection situations on the other hand. ATIC is a construct that has been developed specifically for the selection context. Although it is proposed to be more cognitive than

social and emotional in nature, there are clear aspects in which ATIC and trait EI overlap. For example, both relate to identifying the "correct" or appropriate behavior given certain situations; as such, both concepts can be subsumed under the umbrella of social effectiveness (Ferris et al., 2002; Klehe et al., 2012). Since emotional intelligence is defined as a rather broad set of competences and skills with consequences for a wide range of life domains, and ATIC is very specific to the selection context, it could be hypothesized that ATIC is a more narrow, contextualized form of emotional intelligence. The study in **Chapter 4** could be replicated with ATIC as a predictor, for example, to test this hypothesis. Another possibility would be to investigate whether ATIC predicts work-related outcomes over and beyond emotional intelligence. As ATIC is defined as a cognitive process (Kleinmann et al., 2011), outcomes which are cognitively loaded to different degrees should be included (such as task performance and OCB).

On a final note, although the focus of this dissertation explicitly was on the general factor in self-report personality questionnaires, in the future, new ways could be sought to capture this construct. Any construct measured by a single source suffers from biases associated with the method of measurement (McCrae, 2018). This is also true for the GFP; in **Chapter 5** it was found that the self-report GFP can partly be influenced by pressures to present oneself favorably, as indicated by higher GFP mean scores in the selection context compared to a career development context. At the same time, if a social desirability factor truly exists, i.e., if people truly differ in the extent to which they are socially effective, then this should be observable by others or possibly measured in other ways than self-report. One obvious alternative would be to use other-reports of personality to circumvent measurement issues related to self-reports. Previous studies using a combination of self and other-reports to extract a GFP have presented mixed results, probably because scholars disagree on the correct model to extract a GFP from such combined data (Chang et al., 2012; McCrae et al., 2008; Van der Linden et al., 2016). In research on the Big Five, several variance decomposition studies have been published in which method and trait variance are separated by use of self- and other-ratings (e.g., McCrae, 2015, 2018; McCrae, Mõttus, Hřebíčková, Realo, & Allik, 2018; Mõttus, McCrae, Allik, & Realo, 2014). Given the debate around whether the GFP is due to single-source method bias, such decomposition studies on the GFP in self-ratings and other-ratings are timely at this point. Preliminary results can be found in McCrae et al. (2018) who showed that the size of the general factor did not differ much when examined in single-rater data (around 38% of explained variance based on Estonian data and 32% in Czech data) or self-other ratings (33% and 29% in respectively the Estonian and Czech context). At the same time, the general factor also accounted for a sizeable amount of variance when only method variance was investigated (43% and 40% in respectively the Estonian and Czech context). This implies that the GFP contains a mix of both substantive and error variance (Davies et al., 2015; Dunkel et al., 2016).

A completely different and innovative approach would be to circumvent questionnaires altogether: the developments in artificial intelligence have made it possible to extract personality profiles from social media posts (Youyou, Kosinski, & Stillwell, 2015). We can assume that people also try to present themselves a certain way in their expressions on social media (Krämer & Winter,

2008), which may indicate that evaluative biases are perhaps even larger in social media outings than in personality questionnaires. Since companies nowadays increasingly check social media profiles of applicants in the selection procedure (Nikolaou, 2014; Roth, Bobko, Van Iddekinge, & Thatcher, 2016), it would be interesting to investigate to what extent the general factor from self-report measures converges with the general factor from social media profiles as scored by AI-models, and whether their relations with external criteria are similar.

Conclusion

Self-report personality measures are an often-used component in current selection procedures. Getting or not getting the job or the desired promotion may depend on the scores on these measures. We therefore need to be sure that personality test scores reflect the personality of the applicant. An important question for scientists and practitioners therefore remains whether social desirability introduces bias to the measurement of personality in selection procedures, and subsequently unrightfully influences who gets hired, or whether social desirability validly predicts future job performance and other relevant work outcomes.

The current dissertation has shown that the socially desirable component in self-report personality questionnaires relates in a meaningful way to work-related outcomes such as job performance and organizational citizenship behavior. In addition, this component relates to general indicators of social effectiveness such as relationship quality and less interpersonal conflict, which in turn leads to higher levels of well-being. In addition, the social desirability factor facilitates better affective reactions to social setbacks such as an interpersonal conflict. This factor proved to be a stable construct across methods of analysis, and robust to differences in the motivation and opportunity to provide self-enhancing answers on personality measures. Together, the results from this dissertation make it plausible that the socially desirable component in self-report personality questionnaires reflects the extent to which one *is* socially effective. This notion is further supported by the finding that providing the desirable responses in selection procedures requires socio-emotional skills underlying social effectiveness in and beyond the workplace.

Based on this dissertation, we would recommend scientists and practitioners to take a more nuanced look at social desirability and acknowledge that self-presentation is an integral part of one's personality. Self-presentation requires social skills which will presumably be also important on the job; as such, social desirability scores may actually be used as a selective tool to determine who gets hired and who does not.

References

References

Aghababaei, N. (2013). Between you and God, where is the general factor of personality? Exploring personality-religion relationships in a Muslim context. *Personality and Individual Differences*, *55*, 196-198.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.

Ananiadou, K. & Claro, M. (2009). *21st Century Skills and Competences for New Millennium Learners in OECD Countries*. Paris: OECD Education Working Papers, No. 41.

Andrei, F., Siegling, A. B., Aloe, A. M., Baldaro, B., & Petrides, K. V. (2016). The incremental validity of the Trait Emotional Intelligence Questionnaire (TEIQue): A systematic review and meta-analysis. *Journal of Personality Assessment*, *98*, 261-276.

Anglim, J., Lievens, F., Everton, L., Grant, S. L., & Marty, A. (2018). HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts. *Journal of Research in Personality*, *77*, 11-20.

Anglim, J., Morse, G., De Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing job applicants to non-applicants using an item-level bifactor model on the HEXACO personality inventory. *European Journal of Personality*, *31*, 669-684.

Anguiano-Carrasco, C., MacCann, C., Geiger, M., Seybert, J. M., & Roberts, R. D. (2015). Development of a forced-choice measure of typical-performance emotional intelligence. *Journal of Psychoeducational Assessment*, *33*, 83-97.

Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among Big Five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology*, *97*, 1142-1156.

Argyle, M. (1969) *Social Interaction.* Chicago: Aldine.

Argyle, M. (2001). *The psychology of happiness* (2nd ed.). New York: Routledge.

Arias, V. B., Jenaro, C., & Ponce, F. P. (2018). Testing the generality of the general factor of personality: An exploratory bifactor approach. *Personality and Individual Differences*, *129*, 17-23.

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*, 1531-1544.

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*, 327-353.

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher order factors of personality: Do they exist? *Personality and Social Psychology Review*, *13*, 79 -91.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397-438.

Bäckström, M. (2007). Higher-order factors in a five-factor personality inventory and its relation to social desirability. *European Journal of Psychological Assessment*, *23*, 63-70.

Bäckström, M., & Björklund, F. (2013). Social desirability in personality inventories: Symptoms, diagnosis, and prescribed cure. *Scandinavian Journal of Psychology*, *54*, 152-159.

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five factor inventories have a major higher order factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, *43*, 335-344.

Bäckström, M., Björklund, F., & Larsson, M. R. (2012). Social desirability in personality assessment: Outline of a model to explain individual differences. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 201-213). New York, NY: Oxford University Press.

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, *69*, 49-56.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, *44*, 1-26.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, *81*, 261-272.

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.

Barta, W. D., Tennen, H., & Litt, M. D. (2013). Measurement reactivity in diary research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 108-123). New York: Guilford.

Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, *15*, 263-272.

Beal, D. J. (2015). Esm 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 383-407.

Bentler, P. M. (1990), Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

Berry, C. M., Ones, D. S., Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*, 410-424.

Bertl, B., Pietschnig, J., Tran, U. S., Stieger, S., & Voracek, M. (2017). More or less than the sum of its parts? Mapping the Dark Triad of personality onto a single Dark Core. *Personality and Individual Differences*, *114*, 140-144.

Biderman, M. D., McAbee, S. T., Job Chen, Z., & Hendy, N. T. (2018). Assessing the evaluative content of personality questionnaires using bifactor models. *Journal of Personality Assessment*, *100*, 375-388.

Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of informant. *Journal of Personality*, *72*, 845-876.

References

Bing, M. N., Kluemper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, *116*, 148-162.

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: a replication and extension. *Journal of Applied Psychology*, *89*, 150-157.

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.

Blickle, G., Meurs, J. A., Zettler, I., Solga, J., Noethen, D., Kramer, J. & Ferris, G. R. (2008) Personality, political skill, and job performance. *Journal of Vocational Behavior*, *72*, 377-387.

Blickle, G., Momm, T., Schneider, P. B., Gansen, D., & Kramer, J. (2009). Does acquisitive self-presentation in personality self-ratings enhance validity? Evidence from two experimental field studies. *International Journal of Selection and Assessment*, *17*, 142-153.

Blickle, G., Wendel, S., & Ferris, G. R. (2010). Political skill as moderator of personality-job performance relationships in socioanalytic theory: Test of the getting ahead motive in automobile sales. *Journal of Vocational Behavior*, *76*, 326-335.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187-215.

Bolger, N. & Schilling, E. A. (1991). Personality and the Problems of Everyday Life: The Role of Neuroticism in Exposure and Reactivity to Daily Stressors. *Journal of Personality*, *59*, 355-386.

Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology*, *69*, 890-902.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579-616.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, *89*, 901–910.

Bou Malham, P., & Saucier, G. (2016). The conceptual link between social desirability and cultural normativity. *International Journal of Psychology*, *51*, 474-480.

Breevaart, K., Bakker, A. B., Demerouti, E., & Hetland, J. (2012). The measurement of state work engagement: A multilevel factor analytic study. *European Journal of Psychological Assessment*, *28*, 305-312.

Brown, A. (2008, July). *The impact of questionnaire item format on ability to 'fake good'*. Paper presented at the International Test Commission conference, Liverpool.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*, 460-502.

Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, *44*, 1135-1147.

Brown, A., & Maydeu-Olivares, A. (2018). Ordinal factor analysis of graded-preference questionnaire data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 516-529.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance*, *24*, 358-372.

Burns, G. N., Morris, M. B., Periard, D. A., LaHuis, D., Flannery, N. M., Carretta, T. R., & Roebke, M. (2017). Criterion-related validity of a Big Five general factor of personality from the TIPI to the IPIP. *International Journal of Selection and Assessment*, *25*, 213-222.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … & Moffitt, T. E. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders?. *Clinical Psychological Science*, *2*, 119-137.

Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait-multimethod approach. *Journal of Personality and Social Psychology, 102*, 408-426.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 464-504.

Chen, Z., Watson, P., Biderman, M., & Ghorbani, N. (2016). Investigating the properties of the general factor (M) in Bifactor models applied to Big Five or HEXACO data in terms of method or meaning. *Imagination, Cognition and Personality*, *35*, 216-243.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Cheung, M. W. L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling*, *9*, 55-77.

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*, 40-64.

Chiaburu, D. S., Oh, I.-S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, *96*, 1140 -1166.

Choi, S., Kluemper, D. H., & Sauley, K. S. (2011). What if we fake emotional intelligence? A test of criterion validity attenuation. *Journal of Personality Assessment*, *93*, 270-277.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*, 267-307.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Connelly, B. S., & Chang, L. (2016). A meta-analytic multitrait multirater separation of substance and style in social desirability scales. *Journal of Personality*, *84*, 319-334.

Connelly, B. S., & Hülsheger, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of Personality*, *80*, 603-631.

References

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*, 1092-1122.

Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing Personality Test Formats and Warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, *16*, 155-169.

Cook, M. (2016). *Personnel selection: Adding value through people - A changing picture*. New York, NY: Wiley & Sons.

Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*, 653-665.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1989). *The NEO PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research*, *24*, 315-318.

Côté, S., & Miners, C. T. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly*, *51*, 1-28.

Cramer, A. O., Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*, 414-431.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Danay, E., & Ziegler, M. (2011). Is there really a single factor of personality? A multirater approach to the apex of personality. *Journal of Research in Personality*, *45*, 560-567.

Darcy, M., & Tracey, T. J. (2003). Integrating abilities and interests in career choice: Maximal versus typical assessment. *Journal of Career Assessment, 11*, 219-237.

David, J. P., Green, P. J., Martin, R., & Suls, J. (1997). Differential roles of neuroticism, extraversion, and event desirability for mood in daily life: an integrative model of top-down and bottom-up influences. *Journal of Personality and Social Psychology*, *73*, 149-159.

Davies, S. E., Connelly, B. S., Ones, D. S., & Birkland, A. S. (2015). The General Factor of Personality: The "Big One," a self-evaluative trait, or a methodological gnat that won't go away?. *Personality and Individual Differences*, *81*, 13-22.

Day, A. L., & Carroll, S. A. (2008). Faking emotional intelligence (EI): comparing response distortion on ability and trait-based EI measures. *Journal of Organizational Behavior*, *29*, 761-784.

De Vries, R. E. (2011). No evidence for a General Factor of Personality in the HEXACO Personality Inventory. *Journal of Research in Personality, 45*, 229-232.

De Vries, R. E., Zettler, I., & Hilbig, B. E. (2014). Rethinking trait conceptions of social desirability scales: Impression management as an expression of honesty-humility. *Assessment*, *21*, 286-299.

DeLongis, A., & Holtzman, S. (2005). Coping in context: The role of stress, social support, and personality in coping. *Journal of Personality*, *73*, 1633-1656.

Denissen, J. J., & Penke, L. (2008a). Neuroticism predicts reactions to cues of social inclusion. *European Journal of Personality*, *22*, 497-517.

Denissen, J. J., & Penke, L. (2008b). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*, 1285-1302.

Denissen, J. J., Geenen, R., Van Aken, M. A., Gosling, S. D., & Potter, J. (2008c). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, *90*, 152-157.

Denissen, J. J., Penke, L., Schmitt, D. P., & Van Aken, M. A. (2008). Self-esteem reactions to social interactions: evidence for sociometer mechanisms across days, people, and nations. *Journal of Personality and Social Psychology*, *95*, 181.

Derks, D., Bakker, A. B., Peters, P., & van Wingerden, P. (2016). Work-related smartphone use, work-family conflict and family role performance: The role of segmentation preference. *Human Relations*, *69*, 1045-1068.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*, 1138-1151.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health?. *Personality and Individual Differences*, *33*, 533-552.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*, 542-575.

Diener, E., & Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105-1117.

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., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, *48*, 209-225.

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, *16*, 81-106.

Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, *29*, 479-493.

References

Douglas, E. F., McDaniel, M. A. & Snell, A. E. (1996, August). *The validity of non-cognitive measures decays when applicants fake*. Paper presented at the annual meeting of the Academy of Management, Cincinnati, OH.

Drenth, P. J. D., P. W. C. van Wieringen, & Hoolwerf, G. (2001), *Drenth Testserie Hoger Niveau: Handleiding*. Lisse, the Netherlands: Swets & Zeitlinger.

Dueber, D. M., Love, A. M., Toland, M. D., & Turner, T. A. (2019). Comparison of single-response format and forced-choice format instruments using Thurstonian item response theory. *Educational and Psychological Measurement*, *79*, 108-128.

Dunkel, C. S. (2013). Evidence for the role of the General Factor of Personality (GFP) in enculturation: The GFP and self-construal in Japanese and American samples. *Personality and Individual Differences*, *55*, 417-421.

Dunkel, C. S., & Van der Linden, D. (2014). Evidence for the general factor of personality as social-effectiveness. *Personality and Individual Differences*, *64*, 147-151.

Dunkel, C. S., Nedelec, J. L., & Van der Linden, D. (2018). Using monozygotic twin differences to examine the relationship between parental affection and personality: a life history account. *Evolution and Human Behavior*, *39*, 52-58.

Dunkel, C. S., Van der Linden, D., & Kawamoto, T. (2019). Early childhood social responsiveness predicts the general factor of personality in early adolescence. *Infant and Child Development*, Advance online publication. https://doi.org/10.1002/icd.2123.

Dunkel, C. S., Van der Linden, D., Beaver, K. M., & Woodley, M. A. (2014). Using a prison sample to assess the association between the general factor of personality and general intelligence. *Intelligence*, *47*, 72-82.

Dunkel, C. S., Van der Linden, D., Brown, N. A., & Mathes, E. W. (2016). Self-report based general factor of personality as socially-desirable responding, positive self-evaluation, and social-effectiveness. *Personality and Individual Differences*, *92*, 143-147.

Dunnett, S., Koun, S. & Barber, P. J. (1981). Social desirability in the Eysenck Personality Inventory. *British Journal of Psychology*, *72*, 19-26.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology*, *15*, 13-24.

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking?. *Human Performance*, *16*, 1-23.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, *37*, 90-93.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden Press.

Edwards, A. L., & Walsh, J. A. (1963). The relationship between the intensity of the social desirability keying of a scale and the correlation of the scale with Edwards' SD scale and the first factor loading of the scale. *Journal of Clinical Psychology*, *19*, 200-203.

Edwards, A. L., Diers, C. J., & Walker, J. N. (1962). Response sets and factor loadings on sixty-one personality scales. *Journal of Applied Psychology, 46,* 220-225.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51-100.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*, 1577-1613.

Ellingson, J. E., & McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: An application of VIE theory. *Human Performance*, *24*, 322-337.

Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: insights into response distortion. *Journal of Applied Psychology*, *92*, 386-395.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*, 155-166.

Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, *86*, 122-133.

Erdle, S., & Rushton, J. P. (2010). The general factor of personality, BIS-BAS, expectancies of reward and punishment, self-esteem, and positive and negative affect. *Personality and Individual Differences*, *48*, 762-766.

Erdle, S., Irwing, P., Rushton, J. P., & Park, J. (2010). The general factor of personality and its relation to self-esteem in 628,640 internet respondents. *Personality and Individual Differences*, *48*, 343-346.

Eysenck, H. J. (1970). *The structure of human personality.* London: Methuen.

Eysenck, H. J., & Eysenck, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*. London: University of London Press.

Fallaw, S. S., & Kantrowitz, T. M. (2013). 2013 Global Assessment Trends Report. Retrieved from: https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/documents/gatr_2013_us.pdf.

Fan, X., & Sivo, S. A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16*, 54-69.

Ferguson, E., Chamorro-Premuzic T., Pickering, A., Weiss A. (2011). Five into one doesn't go: A critique of the General Factor of Personality. In T. Chamorro-Premuzic, S. von Stumm, A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences* (pp. 162-186). West Sussux, UK: Wiley-Blackwell.

References

Ferris, G. R., Perrewé, P. L., & Douglas, C. (2002). Social effectiveness in organizations: Construct validity and research directions. *Journal of Leadership & Organizational Studies*, *9*, 49-63.

Ferris, G. R., Witt, L. A., & Hochwarter, W. A. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology*, *86*, 1075-1082.

Figueredo, A. J., & Rushton, J. P. (2009). Evidence for shared genetic dominance between the general factor of personality, mental and physical health, and life history traits. *Twin Research and Human Genetics*, *12*, 555-563.

Figueredo, A. J., Vásquez, G., Brumbach, B. H., & Schneider, S. M. R. (2004). The heritability of life history strategy: The K-factor, covitality, and personality. *Social Biology*, *51*, 121-143.

Fisher, P. A., & Robie, C. (2019). A latent profile analysis of the Five Factor Model of personality: A constructive replication and extension. *Personality and Individual Differences*, *139*, 343-348.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sour sources. *Journal of Abnormal and Social Psychology, 44,* 329-344.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, *56*, 82-92.

Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology*, *9*, 46-55.

Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology*, *78*, 1135-1149.

Galton, F. (1887). Good and bad temper in English families. *Fortnightly Review*, *42*, 21-30.

Geiger, M., Olderbak, S., Sauter, R., & Wilhelm, O. (2018). The "g" in faking: Doublethink the validity of personality self-report measures for applicant selection. *Frontiers in Psychology*, *9*, 1-15.

Gnambs, T. (2013). The elusive general factor of personality: The acquaintance effect. *European Journal of Personality*, *27*, 507-520.

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology*, *50*, 151-160.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96.

Gray, J. A. (1987). The *psychology of fear and stress* (2nd ed.). New York: Cambridge.

Graziano, W. G., Jensen-Campbell, L. A., & Hair, E. C. (1996). Perceiving interpersonal conflict and reacting to it: the case for agreeableness. *Journal of Personality and Social Psychology*, *70*, 820-835.

Griffith, R. L., & Converse, P. D. (2011). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. McCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessments* (pp. 34-52). Oxford, UK: Oxford University Press.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*, 341-355.

Guenole, N., Brown, A. A., & Cooper, A. J. (2016). Forced-choice assessment of work-related maladaptive personality traits: Preliminary evidence from an application of Thurstonian item response modeling. *Assessment*, 1-14.

Gurtman, M. B. (1995). Personality structure and interpersonal problems: a theoretically guided item analysis of the inventory of interpersonal problems. *Assessment*, *2*, 343-361.

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual Review of Psychology*, *68*, 155-186.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639-683.

Hayes, A. F. (2013). Introduction to Mediation, Moderation and Conditional Process Analysis. New York, New York: Guilford Press.

He, J., & Van de Vijver, F. J. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences*, *55*, 794-800.

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, *91*, 9-24.

Heller, D., Watson, D., & Ilies, R. (2004). The role of person versus situation in life satisfaction: a critical examination. *Psychological Bulletin*, *130*, 574-600.

Henderson, S., Byrne, G. O., & Duncan-Jones, P. (1981). Neurosis and social environment. Sidney, Australia: Academic Press.

Hendriks, A. A. J. (1997). *The construction of the Five-Factor Personality Inventory (FFPI)*. Groningen, The Netherlands: Rijksuniversiteit Groningen.

Hendriks, A. J., Hofstee, W. K., & De Raad, B. (1999). The five-factor personality inventory (FFPI). *Personality and Individual Differences*, *27*, 307-325.

Hengartner, M. P., Van der Linden, D., Bohleber, L., & Wyl, A. (2017). Big Five personality traits and the General Factor of Personality as moderators of stress and coping reactions following an emergency alarm on a Swiss university campus. *Stress and Health*, *33*, 35-44.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, *74*, 167-184.

Hofstee, W. K. B. (2003). Structures of personality traits. In T. Millon & M. J. Lerner (Eds.), *Personality and Social Psychology*. In I.B.Weiner (Ed.). *Handbook of psychology* (Vol. 5) (pp. 231-254). Hoboken, NJ: Wiley.

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

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, *92*, 1270-1285.

Hogan, R. & Blickle, G. (2018). Socioanalytic Theory: Basic concepts, supporting evidences, and practical implications. In V. Zeigler-Hill, & T. K. Shackelford (Eds.), *The SAGE Handbook of Personality and Individual Differences*. Thousand Oaks, CA: SAGE.

Hogan, R. & Shelton, D. (1998) A socioanalytic perspective on job performance. *Human Performance*, *11*, 129-144.

Hogan, R. (1982). A socioanalytic theory of personality. In M. M. Page (Ed.), *1982 Nebraska symposium on motivation* (pp. 55-89). Lincoln: University of Nebraska Press.

Hogan, R. (1991). Personality and Personality Assessment. In: M. D. Dunnette & L. Hough, (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., pp. 873-919). Chicago: Rand McNally.

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

Hogan, R., & Shelton. D. (1998). A socioanalytic perspective on job performance. *Human Performance, 11*, 129-144.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, *51*, 469-477.

Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science*, *39*, 184-201.

Holden, R. R. (2008). Underestimating the effects of faking on the validity of self-report personality scales. *Personality and Individual Differences*, *44*, 311-321

Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology*, *37*, 379-388.

Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences*, *49*, 446-450.

Holland, A. S., & Roisman, G. I. (2008). Big Five personality traits and relationship quality: Self-reported, observational, and physiological evidence. *Journal of Social and Personal Relationships*, *25*, 811-829.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*, 53-60.

Hopwood, C. J.,Wright, A. G. C., & Donnellan, M. B. (2011). Evaluating the evidence for the general factor of personality across multiple inventories. *Journal of Research in Personality*, *45*, 468-478.

Hough, L. (1992). The "Big Five" personality variables - construct confusion: Description versus prediction. *Human Performance*, *5*, 139-155.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*, 581-595.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Huffcutt, A. I., Roth, P., & McDaniel, M. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, *81*, 459–473.

Humphrey, R. H., Pollack, J. M., & Hawver, T. H. (2008). Leading with emotional labor. *Journal of Managerial Psychology*, *23*, 151-168.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2015). Shall we continue or stop disapproving of self-presentation? Evidence on impression management and faking in a selection context and their relation to job performance. *European Journal of Work and Organizational Psychology*, *24*, 420-432.

Irwing, P. (2013). A general factor of personality: Substance or artefact? *Personality and Individual Differences*, *55*, 234-242.

Irwing, P., Booth, T., Nyborg, H., & Rushton, J. P. (2012). Are g and the General Factor of Personality (GFP) correlated?. *Intelligence*, *40*, 296-305.

Ispas, D., Iliescu, D., Ilie, A., Sulea, C., Askew, K., Rohlfs, J. T., & Whalen, K. (2014). Revisiting the relationship between impression management and job performance. *Journal of Research in Personality*, *51*, 47-53.

Ixly. (2012). *Werkgerelateerde Persoonlijkheidsvragenlijst (WPV) Handleiding*. Utrecht; Ixly B.V.

Ixly. (2014). *Handleiding WPV Compact*. Utrecht; Ixly B.V.

Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, *55*, 243-252.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution?. *Human Performance*, *13*, 371-388.

Jahng, S., Trull, T. J., Wood, P. K., Tragesser, S. L., Tomko, R., Grant, J. D., . . . Sher, K. J. (2011). Distinguishing general and specific personality disorder features and implications for substance dependence comorbidity. *Journal of Abnormal Psychology*, *120*, 656-669.

Jenkins, M., & Griffith, R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology*, *19*, 255-269.

Jensen, A. R. (1998). The g factor: The science of mental ability. Westport, CT: Praeger.

Jensen-Campbell, L. A., & Malcolm, K. T. (2007). The importance of conscientiousness in adolescent interpersonal relationships. *Personality and Social Psychology Bulletin*, *33*, 368-383.

References

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York: Guilford Press.

Johnson, J. A. (2011, June). *Development of a short form of the IPIP-NEO Personality Inventory*. Poster presented at the 2nd Biennial Meeting of the Association for Research in Personality, Riverside, CA.

Johnson, J. A., & Hogan, R. (2006). A socioanalytic view of faking. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 209-231). Greenwich, CT: Information Age.

Johnson, R. E., Rosen, C. C., & Chang, C.-H. (2011). To aggregate or not to aggregate: Steps for developing and validating higher-order multidimensional constructs. *Journal of Business and Psychology*, *26*, 241-248.

Jonason, P. K., Li, N. P., Webster, G. D., & Schmitt, D. P. (2009). The dark triad: Facilitating a short-term mating strategy in men. *European Journal of Personality*, *23*, 5-18.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). London: Sage.

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, *95*, 54-78.

Joseph, D. L., Jin, J., Newman, D. A., & O'Boyle, E. H. (2015). Why does self-reported emotional intelligence predict job performance? A meta-analytic investigation of mixed EI. *Journal of Applied Psychology*, *100*, 298-342.

Joubert, T., Inceoglu, I., Bartram, D., Dowdeswell, K., & Lin, Y. (2015). A comparison of the psychometric properties of the forced choice and Likert scale versions of a personality instrument. *International Journal of Selection and Assessment*, *23*, 92-97.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits - self-esteem, generalized self-efficacy, locus of control, and emotional stability - with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*, 80-92.

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

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the Five-Factor Model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*, 875-925.

Just, C. (2011). A review of literature on the general factor of personality. *Personality and Individual Differences*, *50*, 765-771.

Kantrowitz, T. M., Tuzinski, K. A., & Raines, J. M. (2018). 2018 Global Assessment Trends Report. Retrieved from: https://www.shl.com/wp-content/uploads/2018-global-assessment-trends-report-en.pdf.

Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, methods, and research. *Psychological Bulletin*, *118*, 3-34.

Kenny, D. A. (2015, November 24). *Measuring model fit*. Retrieved from http://davidakenny.net/cm/fit.htm.

Klehe, U. C., Kleinmann, M., Hartstein, T., Melchers, K. G., König, C. J., Heslin, P. A., & Lievens, F. (2012). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. *Human Performance*, *25*, 273-302.

Klehe, U. C., Kleinmann, M., Nieß, C., & Grazi, J. (2014). Impression management behavior in assessment centers: Artificial behavior or much ado about nothing? *Human Performance*, *27*, 1-24.

Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review*, *1*, 128-146.

Kluemper, D. H. (2008). Trait emotional intelligence: The impact of core-self evaluations and social desirability. *Personality and Individual Differences*, *44*, 1402-1412.

Kokkonen, M., & Pulkkinen, L. E. A. (2001). Extraversion and neuroticism as antecedents of emotion regulation and dysregulation in adulthood. *European Journal of Personality*, *15*, 407-424.

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte-Carlo investigation. *Journal of Applied Psychology*, *93*, 140-154.

König, C. J., Klehe, U. C., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, *18*, 17-27.

König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U. C. (2007). Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment*, *15*, 283-292.

König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U. C. (2006). The relationship between the ability to identify evaluation criteria and integrity test scores. *Psychological Test and Assessment Modeling*, *48*, 369-377.

Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality*, *20*, 549-566.

Krämer, N. C., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology*, *20*, 106-116.

Kuncel, N. R., & Tellegen, A. (2009). A conceptual and empirical reexamination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. *Personnel Psychology*, *62*, 201-228.

Kurtz, J. E., Tarquini, S. J., & Iobst, E. A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences*, *45*, 22-27.

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. New York: McGraw-Hill/Irwin.

Lai, X., Li, F., & Leung, K. (2013). A Monte Carlo study of the effects of common method variance on significance testing and parameter bias in hierarchical linear modeling. *Organizational Research Methods*, *16*, 243-269.

Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: recommendations and strategies. *Journal of Business and Psychology*, *28*(3), 251-261.

Larsen, R. J. (1992). Neuroticism and selective encoding and recall of symptoms: Evidence from a combined concurrent retrospective study. *Journal of Personality and Social Psychology*, *62*, 480-488.

Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology*, *61*, 132-140.

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology*, *68*, 518-530.

Lee, K., Ashton, M. C., & de Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. *Human Performance*, *18*, 179-197.

Lee, P., Joo, S. H., & Lee, S. (2019). Examining stability of personality profile solutions between Likert-type and multidimensional forced choice measure. *Personality and Individual Differences*, *142*, 13-20.

Lee, P., Lee, S., & Stark, S. (2018). Examining validity evidence for multidimensional forced choice measures with different scoring approaches. *Personality and Individual Differences*, *123*, 229-235.

Lee, P., Mahoney, K. T., & Lee, S. (2017). An application of the exploratory structural equation modeling framework to the study of personality faking. *Personality and Individual Differences*, *119*, 220-226.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*, 241-293.

Li, A. & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, *14*, 131-141.

Loehlin, J. C. (2012). How general across inventories is a general factor of personality? *Journal of Research in Personality*, *46*, 258-263.

Loehlin, J. C., Bartels, M., Boomsma, D. I., Bratko, D., Martin, N. G., Nichols, R. C., & Wright, M. J. (2015). Is there a genetic correlation between general factors of intelligence and personality?. *Twin Research and Human Genetics*, *18*, 234-242.

Longua, J., DeHart, T., Tennen, H., & Armeli, S. (2009). Personality moderates the interaction between positive and negative daily events predicting negative affect and stress. *Journal of Research in Personality*, *43*, 547-555.

Lönnqvist, J. E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality*, *75*, 291-322.

Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I., & Salovey, P. (2004). Emotional intelligence and social interaction. *Personality and Social Psychology Bulletin*, *30*, 1018-1034.

MacCann, C. (2013). Instructed faking of the HEXACO reduces facet reliability and involves more Gc than Gf. *Personality and Individual Differences*, *55*, 828-833.

MacCann, C., Matthews, G., Zeidner, M., & Roberts, R. D. (2003). Psychological assessment of emotional intelligence: A review of self-reported and performance-based testing. *The International Journal of Organizational Analysis*, *11*, 247-274.

MacCann, C., Pearce, N., & Jiang, Y. (2017). The General Factor of Personality is stronger and more strongly correlated with cognitive ability under instructed faking. *Journal of Individual Differences*, *38*, 46-54.

MacCann, C., Ziegler, M., & Roberts, R. D. (2011). Faking in personality assessment: Reflections and recommendations. In M. Ziegler, C. MacCann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 309-329). New York, NY: Oxford University Press.

Mahar, D., Cologon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences*, *18*, 605-609.

Marcus, B. (2009). 'Faking' From the applicant's perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment*, *17*, 417-430.

Marcus, B., Taylor, O. A., Hastings, S. E., Sturm, A., & Weigelt, O. (2013). The structure of counterproductive work behavior a review, a structural meta-analysis, and a primary study. *Journal of Management, 20*, 1-31.

Marinova, S. V., Moon, H., & Kamdar, D. (2013). Getting ahead or getting along? The two-facet conceptualization of conscientiousness and leadership emergence. *Organization Science*, *24*, 1257-1276.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85-110.

Marshall, M. B., De Fruyt, F., Rolland, J. P., & Bagby, R. M. (2005). Socially desirable responding and the factorial stability of the NEO PI-R. *Psychological Assessment*, *17*, 379-384.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires?. *Personality and Individual Differences*, *32*, 247-256.

References

Martins, A., Ramalho, N., & Morin, E. (2010). A comprehensive meta-analysis of the relationship between emotional intelligence and health. *Personality and Individual Differences*, *49*, 554 - 564.

Mavroveli, S., Petrides, K. V., Rieffe , C., & Bakker, F. (2007). Trait emotional intelligence, psychological well-being and peer-rated social competence in adolescence. *British Journal of Developmental Psychology*, *25*, 263 - 275.

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*, 97-112.

McCrae, R. R. (2018). Method biases in single-source personality assessments. *Psychological Assessment*, *30*, 1160-1173.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, *51*, 882-888.

McCrae, R. R., & Costa, P. T., (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 139-153). New York: Guilford.

McCrae, R. R., Mõttus, R., Hřebíčková, M., Realo, A., & Allik, J. (2018). Source method biases as implicit personality theory at the domain and facet levels. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12435.

McCrae, R. R., Yamagata, S., Jang, K. L., Riemann, R., Ando, J., Ono, Y., ... & Spinath, F. M. (2008). Substance and artifact in the higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *95*, 442-455.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*, 812-821.

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology*, *36*, 979-1016.

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, *77*, 531-551.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437-455.

Meade, A. W., Pappalardo, G., Braddy, P. W., & Fleenor, J. W. (2018). Rapid Response Measurement: Development of a faking-resistant assessment method for personality. *Organizational Research Methods.* Advance online publication. https://doi.org/10.1177/1094428118795295.

Melchers, K. G., Klehe, U. C., Richter, G. M., Kleinmann, M., König, C. J., & Lievens, F. (2009). "I know what you want to know": The impact of interviewees' ability to identify criteria on interview performance and construct-related validity. *Human Performance*, *22*, 355-374.

Mersman, J. L., & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences*, *24*, 217-227.

Mesmer-Magnus, J., & Viswesvaran, C. (2006). Assessing response distortion in personality tests: A review of research designs and analytic strategies. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 85-114). Greenwich, CT: Information Age.

Michel J. S., Viswesvaran C., Thomas J. (2011). Conclusions from meta-analytic structural equation models generally do not change due to corrections for study artifacts. *Research Synthesis Methods*, *2*, 174-187.

Mikolajczak, M., Luminet, O., Leroy, C., & Roy, E. (2007). Psychometric properties of the Trait Emotional Intelligence Questionnaire: Factor structure, reliability, construct, and incremental validity in a French-speaking population. *Journal of Personality Assessment*, *88*, 338-353.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246-268.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683-729.

Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, *58*, 583-611.

Morillo, D., Leenen, I., Abad, F. J., Hontangas, P., de la Torre, J., & Ponsoda, V. (2016). A Dominance Variant Under the Multi-Unidimensional Pairwise-Preference Framework: Model Formulation and Markov Chain Monte Carlo Estimation. *Applied Psychological Measurement*, *40*, 500-516.

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 116-139.

Moscoso, S., & Salgado, J. F. (2004). "Dark side" personality styles as predictors of task, contextual, and job performance. *International Journal of Selection and Assessment*, *12*, 356-362.

Mõttus, R., McCrae, R. R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality*, *52*, 47-54.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, *11*, 145-165.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher-order dimensions of the big five personality traits and the big six vocational interest types. *Personnel Psychology*, *58*, 447-478.

Mroczek, D. K., & Almeida, D. M. (2004). The effect of daily stress, personality, and age on daily negative affect. *Journal of Personality*, *72*, 355-378.

References

Mueller-Hanson, R. A., Heggestad, E. D., & Thornton, G. C. (2006). Individual differences in impression management: An exploration of the psychological processes underlying faking. *Psychology Science*, *48*, 288-312.

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

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, *41*, 1213-1233.

Musek, J. (2017). The General Factor of Personality: Ten years after. *Psihologijske teme*, *26*, 61-87.

Muthén, L. K., & Muthén, B. O. (1998-2010). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event-and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771-785.

Nezlek, J. B., & Allen, M. R. (2006). Social support as a moderator of day-to-day relationships between daily negative events and daily psychological well-being. *European Journal of Personality*, *20*, 53-68.

Nezlek, J. B., & Gable, S. L. (2001). Depression as a moderator of relationships between positive daily events and day-to-day psychological adjustment. *Personality and Social Psychology Bulletin*, *27*, 1692-1704.

Nezlek, J. B., Hampton, C. P., & Shean, G. D. (2000). Clinical depression and day-to-day social interaction in a community sample. *Journal of Abnormal Psychology*, *109*, 11-19.

Nikolaou, I. (2014). Social networking web sites in job search and employee recruitment. *International Journal of Selection and Assessment*, *22*, 179-189.

O'Boyle, E. H., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior*, *32*, 788-818.

O'Brien, T. B. & DeLongis, A. (1996). The interactional context of problem-, emotion-, and relationship-focused coping: The role of the Big Five personality factors. *Journal of Personality*, *64*, 775-813.

O'Connell, M. S., Kung, M. C., & Tristan, E. (2011). Beyond impression management: Evaluating three measures of response distortion and their relationship to job performance. *International Journal of Selection and Assessment*, *19*, 340-351.

Oltmanns, J. R., Smith, G. T., Oltmanns, T. F., & Widiger, T. A. (2018). General factors of psychopathology, personality, and personality disorder: across domain comparisons. *Clinical Psychological Science*, *6*, 581-589.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, *17*, 609-626.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, *60*, 995–1027.

Ones, D. S., Viswesvaran, C., & Reis, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679.

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*, 679-703.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

Paulhus, D. L. (1991) Measurement and control of response bias. In J. P. Robinson, P. R. Shaver and L. S. Wrightsman (Eds), *Measures of personality and social psychology attitudes* (pp. 17-59). New York: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Braun & D. N. Jackson (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*, 1025-1060.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556-563.

Pauls, C. A., & Crost, N. W. (2005). Cognitive ability and self-reported efficacy of self-presentation predict faking on personality measures. *Journal of Individual Differences*, *26*(4), 194-206.

Pavlov, G., Maydeu-Olivares, A., & Fairchild, A. J. (2018). Effects of applicant faking on forced-choice and Likert scores. *Organizational Research Methods*. Advance online publication. doi: 10.1177/1094428117753683.

Peabody, D. (1967). Trait inferences: evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, *7*, 1-18.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57*, 552-567.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*, Erlbaum, Hillsdale, NJ.

Peeters, M. A., Van Tuijl, H. F., Rutte, C. G., & Reymen, I. M. (2006). Personality and team performance: a meta-analysis. *European Journal of Personality*, *20*, 377-396.

Pekaar, K. A., Bakker, A. B., Born, M. P., & Van der Linden, D. (2018). The consequences of self-and other-focused emotional intelligence: Not all sunshine and roses. *Journal of occupational health psychology*. Advance online publication. https://doi.org/10.1037/ocp0000134.

Pekaar, K. A., Van der Linden, D., Bakker, A. B., & Born, M. P. (2017). Emotional intelligence and job performance: The role of enactment and focus on others' emotions. *Human Performance*, *30*, 135-153.

Pelt, D. H. M., Van der Linden, D., Dunkel, C. S., & Born, M. Ph. (2017). The General Factor of Personality and job performance: Revisiting previous meta-analyses. *International Journal of Selection and Assessment*, *25*, 333-346.

References

Pérez-González, J., & Sanchez-Ruiz, M. (2014). Trait emotional intelligence anchored within the big five, big two and big one frameworks. *Personality and Individual Differences*, *65*, 53-58.

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, *24*, 270-290.

Petrides, K. V. & Furnham, A. (2006). The role of trait emotional intelligence in a gender-specific model of organizational variables. *Journal of Applied Social Psychology*, *36*, 552-569.

Petrides, K. V. (2011). Ability and trait emotional intelligence. In T. Chamorro-Premuzic, A. Furnham, & S. von Stumm (Eds.), *The blackwell-wiley handbook of individual differences* (pp. 656-678). New York: Wiley.

Petrides, K. V., Frederickson, N., & Furnham, A. (2004). The role of trait emotional intelligence in academic performance and deviant behavior at school. *Personality and Individual Differences*, *36*, 277-293.

Petrides, K. V., Furnham, A., & Mavroveli, S. (2007a). Trait emotional intelligence: Moving forward in the field of EI. In G. Matthews, M. Zeidner, & R. Roberts (Eds.), *Emotional intelligence: Knowns and unknowns* (pp. 151-166). Oxford, UK: Oxford University Press.

Petrides, K. V., Pérez-González, J. C., & Furnham, A. (2007b). On the criterion and incremental validity of trait emotional intelligence. *Cognition and Emotion*, *21*, 26-55.

Petrides, K. V., Vernon, P. A., Schermer, J. A., Ligthart, L., Boomsma, D. I., & Veselka, L. (2010). Relationships between trait emotional intelligence and the Big Five in the Netherlands. *Personality and Individual Differences*, *48*, 906-910.

Pettersson, E., Turkheimer, E., Horn, E. E., & Menatti, A. R. (2012). The general factor of personality and evaluation. *European Journal of Personality*, *26*, 292-302.

Pinheiro J., Bates D., DebRoy S., Sarkar D., & R Core Team (2016). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-128

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of applied psychology, *88*, 879-903.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322-338.

Prediger, D. J. (1998). Is interest profile level relevant to career counseling? *Journal of Counseling Psychology, 45*, 204-211.

Presser, S., & Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review*, 137-145.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raymark, P. H. & Tafero, T. L. (2009). Individual differences in the ability to fake on personality measures. *Human Performance*, *22*, 86-103.

Raymark, P. H., Shilobod, T. L., & Steffensmeier, J. (April, 2004). *An examination of job-relevant pattern faking*. Paper presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667-696.

Revelle, W. (2017) *psych*: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych [Version = 1.7.8].

Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, *47*, 493-504.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145-154.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*, 137-150.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenström, T., Gjerde, L. C., Krueger, R. F., Aggen, S. H., Czajkowski, N. O., Gillespie, N. A., ... & Ystrom, E. (2018). Joint factorial structure of psychopathology and personality. *Psychological Medicine*, 1-10. Advance online publication. https://doi.org/10.1017/S0033291718002982

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*, 634-644.

Roth, M., & Altmann, T. (2019). A multi-informant study of the influence of targets' and perceivers' social desirability on self-other agreement in ratings of the HEXACO personality dimensions. *Journal of Research in Personality*, *78*, 138-147.

Roth, P. L., Bobko, P., Van Iddekinge, C. H., & Thatcher, J. B. (2016). Social media in employee-selection-related decisions: A research agenda for uncharted territory. *Journal of Management*, *42*, 269-298.

Rotundo, N., & Spector, P. E. (2010). Counterproductive work behavior and withdrawal. In J. L. Farr & N. T. Tippens (Eds.), *Handbook of Employee Selection* (pp.489-511). Routledge: Taylor & Francis.

Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review*, *6*, 145-170.

Rushton, J. P., & Erdle, S. (2010). No evidence that social desirability response set explains the general factor of personality and its affective correlates. *Twin Research and Human Genetics*, *13*, 131-134.

Rushton, J. P., & Irwing, P. (2008). A general factor of personality (GFP) from two meta analyses of the Big Five: Digman (1997) and Mount, Barrick, Scullen, and Rounds (2005). *Personality and Individual Differences*, *45*, 679-683.

Rushton, J. P., & Irwing, P. (2011). The general factor of personality: Normal and abnormal. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences*. West Sussux, UK: Wiley-Blackwell.

Rushton, J. P., Bons, T. A., & Hur, Y-M. (2008). The genetics and evolution of a general factor of personality. *Journal of Research in Personality*, *42*, 1173-1185.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, *50*, 540-547.

Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology*, *88*, 797-834.

Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment*, *21*, 74-84.

Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, *63*, 506-516.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality*, *36*, 1-31.

Scherbaum, C. A., Sabet, J., Kern, M. J., & Agnello, P. (2013). Examining faking on personality inventories using unfolding item response theory models. *Journal of Personality Assessment*, *95*, 207-216.

Schermer, J. A., & Goffin, R. D. (2018). A tale of two general factors of personality in relation to intelligence and validity measures. *Personality and Individual Differences*, *124*, 111-116.

Schermer, J. A., & MacDougall, R. (2013). A general factor of personality, social desirability, cognitive ability, and the survey of work styles in an employment selection setting. *Personality and Individual Differences*, *54*, 141-144.

Schermer, J. A., & Vernon, P. A. (2010). The correlation between general intelligence (g), a general factor of personality (GFP), and social desirability. *Personality and Individual Differences*, *48*, 187-189.

Schermer, J. A., Carswell, J., & Jackson, S. (2012). Correlations between a general factor of personality and employment measures. *Personality and Individual Differences*, *53*, 557-561.

Schermer, J. A., Holden, R. R., & Krammer, G. (2019a). The general factor of personality is very robust under faking conditions. *Personality and Individual Differences*, *138*, 63-68.

Schermer, J. A., Krammer, G., & Goffin, R. D. (2019b). The General Factor of Personality and faking: A cautionary note on the meaningfulness of the GFP under different response conditions. *Personality and Individual Differences*, *137*, 110-114.

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

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods*, *8*, 206-224.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, *78*, 966-974.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*, 607-620.

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, *91*, 613-621.

Schulte, M. J., Ree, M. J., & Carretta, T. R. (2004). Emotional intelligence: Not much more than *g* and personality. *Personality and Individual Differences*, *37*, 1059-1068.

Schutte, N. S., Malouff, J. M., Bobik, C., Coston, T. D., Greeson, C., Jedlicka, C., ... & Wendorf, G. (2001). Emotional intelligence and interpersonal relations. *The Journal of Social Psychology*, *141*, 523-536.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333-343.

Semadar, A., Robins, G., & Ferris, G. R. (2006). Comparing the effects of multiple social effectiveness constructs in the prediction of managerial performance. *Journal of Organizational Behavior*, *27*, 443-461.

Sevdalis, N., Petrides, K. V., & Harvey, N. (2007). Trait emotional intelligence and decision-related emotions. *Personality and Individual Differences*, *42*, 1347-1358.

Shaffer, J. A., & Postlethwaite, B. E. (2013). The validity of conscientiousness for predicting job performance: A meta-analytic test of two hypotheses. *International Journal of Selection and Assessment*, *21*, 183-199.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*, 543-554.

Shaver, P. R., & Brennan, K. A. (1992). Attachment styles and the "Big Five" personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin*, *18*, 536-545.

SHL. (2013). *OPQ32r technical manual version 1.0*. Thames Ditton, UK: SHL Group.

Şimşek, Ö. F. (2012). Higher-order factors of personality in self-report data: Self-esteem really matters. *Personality and Individual Differences*, *53*, 568-573.

Sitser, T. B., Van der Linden, D., & Born, M. P. (2013). Predicting six different sales performance criteria with personality measures: The use of the General Factor of Personality, the Big Five and narrow traits. *Human Performance*, *26*, 126-149.

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, *87*, 211-219.

Smith, D. B., & McDaniel, M. (2012). Questioning old assumptions: Faking and the personality-performance relationship. In M. Ziegler, C. MacCann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 53-70). New York, NY: Oxford University Press.

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology*, *86*, 304-315.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, *35*, 50-61.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, *9*, 219-242.

Spain, S. M., Harms, P., & LeBreton, J. M. (2014). The dark side of personality at work. *Journal of Organizational Behavior*, *53*, S41-S60.

Spector, P. E. (2006). Method variance in organizational research: truth or urban legend?. *Organizational Research Methods*, *9*, 221-232.

Steiger, J. H. (1990), Structural model evaluation and modification, *Multivariate Behavioral Research*, *25*, 214-12.

Stöber, J. (1999). The Social Desirability Scale-17 (SDS-17): Development and first results on reliability and validity. *Diagnostica*, *45*, 173-177.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta analytic review of longitudinal research. *Intelligence*, *35*, 401-426.

Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: A "multisaturation" perspective on faking as performance. *Human Performance*, *24*, 302-321.

Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Personality and Individual Differences*, *52*, 195-201.

Tisak, J., & Smith, C. S. (1994). Defending and extending difference score methods. *Journal of Management*, *20*, 675-682.

Tracey, T. J. G. (2012). Problems with single interest scales: Implications of the general factor. *Journal of Vocational Behavior, 81*, 378-384.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

Uziel L. (2010a). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, *5*, 243-262.

Uziel, L. (2010b). Look at me, I'm happy and creative: The effect of impression management on behavior in social presence. *Personality and Social Psychology Bulletin*, *36*, 1591-1602.

Uziel, L. (2014). Impression management ("lie") scales are associated with interpersonally oriented self-control, not other-deception. *Journal of Personality*, *82*, 200-212.

Van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & Van der Maas, H. L. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, *27*, 759-773.

Van der Linden, D. (2011). The general factor of personality and peer-rated social status: A rejoinder to de Vries (2011). *Personality and Individual Differences, 51,* 836-839.

Van der Linden, D., Bakker, A. B., & Serlie, A. W. (2011). The general factor of personality in selection and assessment samples. *Personality and Individual Differences*, *51*, 641-645.

Van der Linden, D., Dunkel, C. S., & Petrides, K. V. (2016). The General Factor of Personality (GFP) as social effectiveness: Review of the literature. *Personality and Individual Differences*, *101*, 98-105.

Van der Linden, D., Dunkel, C. S., Beaver, K. M., & Louwen, M. (2015). The unusual suspect: The General Factor of Personality (GFP), life history theory, and delinquent behavior. *Evolutionary Behavioral Sciences*, *9*, 145-160.

Van der Linden, D., Dunkel, C. S., Figueredo, A. J., Gurven, M., von Rueden, C., & Woodley of Menie, M. A. (2018). How universal is the General Factor of Personality? An analysis of the Big Five in forager farmers of the Bolivian Amazon. *Journal of Cross-Cultural Psychology*, *49*, 1081-1097.

Van der Linden, D., Figueredo, A. J., de Leeuw, R. N., Scholte, R. H., & Engels, R. C. (2012a). The General Factor of Personality (GFP) and parental support: testing a prediction from Life History Theory. *Evolution and Human Behavior*, *33*, 537-546.

Van der Linden, D., Oostrom, J. K., Born, M. P., Van der Molen, H. T., & Serlie, A. W. (2014a). Knowing what to do in social situations: The general factor of personality and performance on situational judgment tests. *Journal of Personnel Psychology*, *13*, 107-115.

Van der Linden, D., Pekaar, K. A., Bakker, A. B., Schermer, J. A., Vernon, P. A., Dunkel, C. S., & Petrides, K. V. (2017). Overlap between the general factor of personality and emotional intelligence: A meta-analysis. *Psychological Bulletin*, *143,* 36-52.

Van der Linden, D., Schermer, J. A., de Zeeuw, E., Dunkel, C. S., Pekaar, K. A., Bakker, A. B., Vernon, P. A., & Petrides, K. V. (2018). Overlap between the General Factor of Personality and Trait Emotional Intelligence: A genetic correlation study. *Behavior Genetics*, *48*, 147-154.

Van der Linden, D., Scholte, R. H. J., Cillessen, A. N. H., Te Nijenhuis, J., & Segers, E. (2010b). Classroom ratings of likeability and popularity are related to the Big Five and the General Factor of Personality. *Journal of Research in Personality*, *44*, 669-672.

Van der Linden, D., Te Nijenhuis, J., & Bakker, A. B. (2010a). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, *44*, 315-327.

Van der Linden, D., Te Nijenhuis, J., Cremers, M., Van de Ven, C., & Van der Heijden-Lek, K. (2014b). The general factor of personality (GFP) relates to other ratings of character and integrity: Two validity studies in personnel selection and training of the Dutch armed forces. International Journal of Selection and Assessment, 22, 261-271.

Van der Linden, D., Tsaousis, I., & Petrides, K. V. (2012b). Overlap between General Factors of Personality in the Big Five, Giant Three, and trait emotional intelligence. *Personality and Individual Differences*, *53*, 175-179.

Van Geert, E., Orhon, A., Cioca, I. A., Mamede, R., Golušin, S., Hubená, B., & Morillo, D. (2016). Study protocol on intentional distortion in personality assessment: Relationship with test format, culture, and cognitive ability. *Frontiers in Psychology*, *7*, 1-8.

Van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*, 301-316.

Vasilopoulos, N. L., & Cucina, J. M. (2006). Faking on noncognitive measures: The interaction of cognitive ability and test characteristics. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 305-331). Greenwich, CT: Information Age.

Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability?. *Journal of Applied Psychology*, *90*, 306-322.

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, *19*, 175-199.

Veselka, L., Just, C., Jang, K. L., Johnson, A. M., & Vernon, P. A. (2012). The general factor of personality: A critical test. *Personality and Individual Differences*, *52*, 261-264.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, *48*, 865-885.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*, 197-210.

Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment*, *9*, 277-289.

Von Dras, D. D., & Siegler, I. C. (1997). Stability in extraversion and aspects of social support at midlife. *Journal of Personality and Social Psychology*, *72*, 233-241.

Waters, L. K. (1965). A note on the "fakability" of forced-choice scales. *Personnel Psychology*, *18*, 187-191.

Watson D., Clark L. A., Tellegen A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Watson, D., Clark, L. A., McIntyre, C. W., & Hamaker, S. (1992). Affect, personality, and social activity. *Journal of Personality and Social Psychology*, *63*, 1011-1025.

Webb, E. (1915). *Character and intelligence. An attempt at an exact study of character.* London: Cambridge University Press.

Whitman, D. S., Van Rooy, D. L., Viswesvaran, C., & Alonso, A. (2008). The susceptibility of a mixed model measure of emotional intelligence to faking: a Solomon four-group design. *Psychology Science*, *50*, 44-63.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395.

Witt, L. A., & Ferris, G. R. (2003). Social skill as moderator of the conscientiousness-performance relationship: convergent results across four studies. *Journal of Applied Psychology*, *88*, 809-820.

Wright, A. G., Hopwood, C. J., Skodol, A. E., & Morey, L. C. (2016). Longitudinal validation of general and specific structural features of personality pathology. *Journal of Abnormal Psychology*, *125*, 1120-1134.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*, 1036-1040.

Zautra, A. J., Affleck, G. G., Tennen, H., Reich, J. W., & Davis, M. C. (2005). Dynamic approaches to emotions and stress in everyday life: Bolger and Zuckerman reloaded with positive as well as negative affects. *Journal of Personality*, *73*, 1511-1538.

Zettler, I., Hilbig, B. E., Moshagen, M., & de Vries, R. E. (2015). Dishonest responding or true virtue? A behavioral test of impression management. *Personality and Individual Differences*, *81*, 107-111.

Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, *84*, 551-563.

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, *69*, 548-565.

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the Big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality*, *24*, 341-355.

Ziegler, M., MacCann, C., & Roberts, R. D. (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3-16). New York, NY: Oxford University Press.

Ziegler, M., MacCann, C., & Roberts, R. D. (2011b). *New perspectives on faking in personality assessment*. New York, NY: Oxford University Press.

Samenvatting (Summary in Dutch)

Samenvatting

Wanneer iemand solliciteert voor een nieuwe baan, dan is het goed mogelijk dat een zelfrapportage persoonlijkheidsvragenlijst onderdeel is van het selectieproces. Een bedrijf laat een toekomstige medewerker vaak een persoonlijkheidsvragenlijst invullen om te weten te komen of de sollicitant past bij de functie en het bedrijf, en hoe deze zich zal gaan gedragen op de werkvloer. Eerder onderzoek liet zien dat scores op zelfrapportage persoonlijkheidsvragenlijsten inderdaad een voorspellende waarde hebben voor werkprestaties en ander relevant werkgedrag, zoals contraproductief werkgedrag (Barrick & Mount, 1991; Schmidt & Hunter, 1998). Er zullen echter ook bedrijven zijn die liever geen persoonlijkheidsvragenlijst inzetten in het selectieproces, omdat ze vermoeden dat sollicitanten de vragenlijst niet helemaal naar waarheid invullen, maar sociaal wenselijke antwoorden geven.

Sociaal wenselijk antwoorden is een groot thema in wetenschappelijk onderzoek naar zowel persoonlijkheid als personeelsselectie (Paulhus, 2002; Ziegler et al., 2011b). Lange tijd werd aangenomen dat sociaal wenselijk antwoorden een verstorende invloed heeft op scores op persoonlijkheidsvragenlijsten omdat deze een te rooskleurig beeld zouden geven van de kandidaat. Door deze verkleuring zouden de scores geen goede weerspiegeling zijn van de 'ware' persoonlijkheid van de persoon. In selectiesituaties kan dit natuurlijk tot problemen leiden; als scores op zelfrapportage persoonlijkheidsvragenlijsten worden gebruikt om kandidaten te selecteren voor een baan, maar de scores niet weergeven hoe de persoon in werkelijkheid is, dan vergroot dit de kans dat de verkeerde persoon de baan krijgt met alle negatieve gevolgen van dien. In deze gedachtegang wordt sociale wenselijkheid gezien als een vorm van *bias*.

Echter, drie onderzoekslijnen uit zowel de persoonlijkheids- als personeelsselectieliteratuur suggereren dat sociale wenselijkheid niet noodzakelijkerwijs als verstorende factor gezien hoeft te worden, maar eerder als een betekenisvolle eigenschap van een persoon.

De eerste onderzoekslijn richt zich op de zogeheten algemene factor van persoonlijkheid (Figueredo et al., 2004; Musek, 2007). Deze algemene factor wordt gegeven als een verklaring voor de vaak gevonden correlaties tussen de persoonlijkheidskenmerken van het Vijf Factor Model van persoonlijkheid, dat wil zeggen tussen de kenmerken Openheid voor Ervaringen, Consciëntieusheid, Extraversie, Vriendelijkheid en Neuroticisme (Costa & McCrae, 1992; Goldberg, 1990). Deze algemene factor wordt gekenmerkt door een mix van sociaal wenselijke eigenschappen, waarbij mensen die hoger scoren op deze factor als meer ruimdenkend, ijverig, sociaal, vriendelijk en emotioneel stabiel omschreven worden. Empirisch onderzoek heeft aangetoond dat scores op deze algemene factor samenhangen met onder andere leiderschap (Pelt et al., 2017; Van der Linden et al., 2014a), crimineel gedrag (negatief verband; Van der Linden et al., 2015), populariteit (Van der Linden et al., 2010b), en belangrijk voor dit proefschrift, met succes in selectiesituaties (Van der Linden et al., 2014b). De gangbare inhoudelijke interpretatie van deze algemene factor is dan ook dat deze de sociale effectiviteit van een persoon weerspiegelt (Van der Linden et al., 2016).

De tweede onderzoekslijn richt zich op het bestuderen van sociale wenselijkheidsschalen. Dit soort schalen zijn ooit ontwikkeld om na te gaan of iemand sociaal wenselijke antwoorden heeft gegeven of juist gelogen heeft bij het invullen van de

persoonlijkheidsvragenlijst. Echter, eerder onderzoek liet zien dat scores op deze schalen meer informatief zijn over iemands persoonlijkheid dan over de mate waarin deze heeft geprobeerd de scores op de vragenlijst te vertekenen. Deze bevinding blijkt onder andere uit het feit dat de scores op sociale wenselijkheidsschalen van een persoon verkregen via zelfrapportage overeenkomen met de scores over deze persoon die zijn verkregen bij partners, kennissen en familieleden (Connelly & Chang, 2016; De Vries et al., 2014; Uziel, 2010a).

De derde onderzoekslijn heeft als aandachtsveld personeelsselectie en richt zich op de manier waarop mensen zich in selectieprocedures en vervolgens op het werk gedragen. Uit verschillende studies is gebleken dat het tonen van het juiste gedrag in selectiesituaties een verzameling van sociale, emotionele en cognitieve vaardigheden vereist (Kleinmann et al., 2011; Roulin et al., 2016); vaardigheden die logischerwijs ook ingezet kunnen worden op het werk en hierdoor op een positieve wijze bijdragen aan de arbeidsprestaties (Hogan & Shelton, 1998). Deze onderzoekslijn suggereert dan ook dat de manier waarop een kandidaat zich voordoet tijdens selectieprocedures een reflectie is van werkrelevante eigenschappen van deze persoon, en dat deze manier van presenteren niet zo zeer verschilt van hoe de persoon dit buiten de selectieprocedure zal doen.

In het licht van de drie hierboven beschreven onderzoekslijnen was het doel van dit proefschrift om meer onderbouwing aan te dragen voor de inhoudelijke interpretatie van het concept sociale wenselijkheid in selectiesituaties en daarbuiten, te weten op het werk en in het dagelijks leven. De vier empirische studies gepresenteerd in dit proefschrift komen voort uit de onderstaande hoofdvraag:

> *Hoofdvraag: Kan, ten behoeve van personeelsselectie, sociale wenselijkheid in zelfrapportage persoonlijkheidsvragenlijsten worden beschouwd als een relevante factor?*

Hieronder worden de belangrijkste bevindingen van de empirische studies uit het proefschrift besproken. De eerste twee onderstaande onderzoeksvragen (en empirische studies) legden de nadruk op de voorspellende waarde van de algemene factor van persoonlijkheid, die wordt opgevat als operationalisatie van sociale wenselijkheid.

> *Onderzoeksvraag 1: Is sociale wenselijkheid in zelfrapportage persoonlijkheidsvragenlijsten voorspellend voor werkgerelateerde uitkomsten?*

Als de algemene factor van persoonlijkheid inderdaad een indicatie is van de sociale effectiviteit van een persoon, dan kan worden verwacht dat deze ook tot uiting komt op de werkvloer. Met andere woorden, mensen met hogere scores op deze algemene factor zullen waarschijnlijk betere beoordelingen krijgen van collega's en leidinggevenden, omdat zij beter weten hoe ze zich in verschillende sociale situaties dienen te gedragen. In de eerste empirische studie is daarom de relatie tussen de algemene persoonlijkheidsfactor en een aantal werkgerelateerde uitkomstmaten (werkprestatie, leiderschapskwaliteiten en -stijl, *organizational citizenship behavior* en contraproductief werkgedrag) onderzocht. Hiervoor werden correlatiematrices uit

vooraanstaande gepubliceerde meta-analyses met elkaar gecombineerd, om vervolgens door middel van structurele modellen relaties te schatten tussen de algemene persoonlijkheidsfactor en de uitkomstmaten.

Een belangrijke bevinding was dat de algemene persoonlijkheidsfactor relatief sterke relaties laat zien met de genoemde uitkomstmaten, zeker in vergelijking met de relaties die gevonden zijn voor bijvoorbeeld conscientiëusheid – de factor die uit eerder onderzoek als meest voorspellende persoonlijkheidskenmerk voor arbeidsprestaties naar voren komt (Barrick & Mount, 1991). Bovendien bleek dat de relaties tussen de vijf persoonlijkheidsdimensies van het Vijf Factor Model en de uitkomstmaten voor een groot deel beschreven kunnen worden door de algemene persoonlijkheidsfactor. Uit hiërarchische regressieanalyse bleek namelijk dat in veel van de gevallen de vijf factoren weinig verklaarde variantie in de uitkomstmaten toevoegden ten opzichte van de algemene persoonlijkheidsfactor. Dit is een relevante bevinding voor zowel persoonlijkheids- als personeelsselectie-theorieën. Als de sociaal wenselijke component in persoonlijkheidsvragenlijsten (dus de algemene persoonlijkheidsfactor) slechts *bias* zou voorstellen, dan zou beargumenteerd kunnen worden dat deze bias verwijderd moet worden, bijvoorbeeld door er statistisch voor te controleren. Echter, de resultaten uit **Hoofdstuk 2** tonen aan dat veel van de relaties tussen persoonlijkheid en uitkomstmaten dan veranderen of verdwijnen. Het is dus de vraag hoe zinvol het is om dit soort correcties toe te passen.

De voorspellende waarde van de algemene persoonlijkheidsfactor verschilde tussen de uitkomstmaten. De gevonden relaties met leiderschapskwaliteiten waren bijvoorbeeld relatief hoog, wat ondersteuning biedt voor het idee dat mensen die sociaal wenselijk gedrag laten zien eerder gekozen en geaccepteerd zullen worden als leiders (Van der Linden et al., 2014a). De relatie tussen de algemene persoonlijkheidsfactor en werkprestatie van *professionals* was juist relatief laag. Deze relatief lage relatie lijkt goed te verklaren wanneer het type banen in ogenschouw wordt genomen dat zich in de categorie *professionals* bevond; dit waren met name technische- en IT-gerelateerde beroepen. In deze beroepen zal de mate waarin iemand de benodigde technische vaardigheden heeft wellicht meer van belang zijn voor prestaties op het werk. Het lijkt daarom aannemelijk dat sociale componenten in deze beroepsgroepen wat minder een rol spelen dan in andere beroepsgroepen. Daarom is te verwachten dat het tonen van sociaal effectief gedrag bij *professionals* minder van invloed is op werkprestaties dan in andere beroepstypen.

Onderzoeksvraag 2: Is sociale wenselijkheid in zelfrapportage persoonlijkheidsvragenlijsten gerelateerd aan (dagelijkse) metingen van sociale effectiviteit?

De eerste studie, zoals beschreven in **Hoofdstuk 2,** richtte zich specifiek op het domein van werk. Het effect van sociale wenselijkheid zal zich echter niet beperken tot de werkvloer, maar zal ook invloed hebben op andere levensdomeinen, bijvoorbeeld op het sociale leven van een persoon. Dat wil zeggen, als de algemene persoonlijkheidsfactor inderdaad een factor is die weerspiegelt hoe gemakkelijk iemand door het sociale verkeer navigeert, dan valt te verwachten

dat deze factor positief gerelateerd is aan de kwaliteit van sociale relaties, en, als gevolg hiervan, aan een hogere mate van welbevinden.

Voor het toetsen van deze verwachting is in **Hoofdstuk 3** gebruikgemaakt van secundaire data uit de *Berlin Diary Study* (2005 - 2008) van Prof. Dr. Jaap Denissen en collega's. In deze studie werden volwassenen in Duitsland gevraagd om 25 dagen achter elkaar dagelijks een vragenlijst in te vullen over hun welbevinden, gemoedstoestand en hun sociale relaties met hun partner (indien aanwezig), een vriend en een familielid met wie men op dat moment het meeste contact had. Vóór de dagboekstudie werden een algemene vragenlijsten afgenomen, waaronder twee persoonlijkheidsvragenlijsten waaruit twee algemene persoonlijkheidsfactoren werden geëxtraheerd ten behoeve van onze studie. De antwoorden van 1223 personen (1055 vrouwen, 86%) werden gebruikt om drie hypothesen te toetsen. Ten eerste verwachtten we dat scores op de algemene persoonlijkheidsfactor samen zouden hangen met een aantal indicatoren van sociale effectiviteit (gerapporteerde relatiekwaliteit, aantal interpersoonlijke conflicten, en de indruk die men had achtergelaten op anderen). De tweede verwachting was dat het verband tussen de algemene factor enerzijds en negatief affect, positief affect en het gevoel van eigenwaarde anderzijds, gemedieerd zou worden door bovenstaande indicatoren. Ten derde verwachtten we dat scores op de algemene factor de relatie tussen dagelijkse sociale ervaringen en dagelijks welbevinden zouden modereren. De verwachting was bijvoorbeeld dat een conflict een grotere (negatieve) impact zou hebben op het ervaren welbevinden op een dag voor mensen met lagere scores op de algemene factor dan voor mensen met hogere scores op deze factor.

De resultaten wezen uit dat er een zwakke negatieve relatie was tussen algemene persoonlijkheidsfactor en het aantal ervaren conflicten. Ook waren er middelgrote (positieve) relaties te zien tussen de algemene persoonlijkheidsfactor enerzijds en relatiekwaliteit en de gemaakte indruk op anderen anderzijds. Verder bleek het verband tussen de algemene persoonlijkheidsfactor enerzijds en negatief affect, positief affect en het gevoel van eigenwaarde anderzijds gemedieerd te worden door relatiekwaliteit en de gemaakte indruk op anderen. Tot slot bleek, zoals voorspeld, dat mensen die hoger scoorden op de algemene factor minder aangeslagen waren na een conflict (blijkend uit hun gerapporteerde staat van negatief affect en hun gevoel van eigenwaarde) dan mensen die lager scoorden op de algemene factor.

De resultaten gepresenteerd in **Hoofdstuk 3** bieden een nieuw stukje bewijs voor de gedachte dat de algemene persoonlijkheidsfactor de mate van sociale effectiviteit van een persoon reflecteert. Hogere scores op de algemene persoonlijkheidsfactor duiden op meer sociale effectiviteit blijkens de betere sociale relaties die men ervaart. Bovendien zorgen deze kwalitatief hoogwaardiger relaties voor een hogere mate van welbevinden. De minder sterke affectieve reacties op lastige sociale situaties (bijvoorbeeld op een conflict) onder mensen die hoger scoren op de algemene factor is in overeenstemming met het idee dat sociale vaardigheden samenhangen met deze factor ervoor zorgen dat mensen beter kunnen reageren wanneer deze situaties zich voordoen, leidend tot een algeheel betere aanpassing aan de sociale omgeving. Voor de selectiepraktijk is de bevinding zinvol dat de algemene persoonlijkheidsfactor relatief sterk gerelateerd is aan de indruk die men achter laat op anderen. Deze bevinding ondersteunt het idee dat de algemene factor een sociale wenselijkheidsfactor

is, maar dan in inhoudelijke zin. Dat wil zeggen, het lijkt aannemelijk dat hoe iemand zich presenteert aan anderen een belangrijk onderdeel vormt van diens persoonlijkheid. Het is te verwachten dat hoe goed iemand zich kan presenteren van invloed is op diens sociale relaties en welbevinden, maar ook op diens succes in het werkende leven. Deze gedachtegang is in overeenstemming met de interpretatie van de bevindingen uit **Hoofdstuk 2** over de relatie tussen de algemene factor en werkprestaties, en met bevindingen van andere studies die hebben aangetoond dat deze factor gerelateerd is aan betere beoordelingen in sollicitatieprocedures (Van der Linden et al., 2014b).

Uit de onderzoeken beschreven in **Hoofdstuk 2** en **Hoofdstuk 3** bleek dat de sociaal wenselijke component in zelfrapportage vragenlijsten (de algemene persoonlijkheidsfactor) gebruikt kan worden om uitkomsten op en buiten het werk te voorspellen. De vraag is echter hoe deze component precies tot stand komt. Over het algemeen wordt aangenomen dat sociale wenselijkheid voor een deel sociale vaardigheden of -competenties vergt (Hogan & Shelton, 1998; Marcus, 2009). Dat wil zeggen, iemand kan pas sociaal wenselijk gedrag vertonen als hij of zij weet welk gedrag gepast is in welke (sociale) situatie. Bovendien moet iemand de capaciteiten hebben om deze kennis in te zetten om vervolgens ook het juiste, sociaal wenselijke, gedrag te laten zien.

Onderzoeksvraag 3: Is het geven van sociaal wenselijke antwoorden op zelfrapportage persoonlijkheidsvragenlijsten in selectiesituaties gerelateerd aan sociale competenties?

Selectieprocedures zijn bij uitstek situaties waarin het voor de hand ligt te onderzoeken of sociaal wenselijk antwoorden gebaseerd is op de sociale competenties van een kandidaat, omdat het in dit soort situaties belangrijk is dat de kandidaat zich van de beste kant laat zien en er dus een beroep gedaan wordt op deze sociale competenties. Het is vaak namelijk niet geheel duidelijk waar het bedrijf waar iemand bij solliciteert precies naar op zoek is of wat men precies van een sollicitant willen horen. In het algemeen wordt daarom aangenomen dat het in selectiesituaties voor sollicitanten belangrijk is om de informatie die voor handen is (dit kunnen interviewvragen zijn, maar ook bijvoorbeeld items in een persoonlijkheidsvragenlijst) juist te interpreteren, en vervolgens te bedenken welk antwoord en gedrag gepast is om zo de best mogelijke indruk achter te laten. Sociale kennis en -competenties, ook wel aangeduid met emotionele intelligentie, kunnen bij bovenstaande processen een belangrijke rol spelen (Ferris et al., 2002).

Gebaseerd op deze gedachtegang werd een studie uitgevoerd (beschreven in **Hoofdstuk 4**) waarin werd nagegaan of het geven van de gewenste antwoorden in een sollicitatieprocedure samenhangt met de emotionele intelligentie van een persoon. Een groep van 129 bachelorstudenten vulde eerst een emotionele intelligentie vragenlijst in en een persoonlijkheidsvragenlijst, waarbij de studenten geïnstrueerd werden zo eerlijk mogelijk te antwoorden, dus in overeenstemming met hoe ze in werkelijkheid zijn. Vervolgens werd ze nog twee keer gevraagd de persoonlijkheidsvragenlijst in te vullen, maar te bedenken dat ze solliciteerden voor respectievelijk de baan van archiefmedewerker en van advocaat, en te proberen zodanig te antwoorden dat hun kans om aangenomen te worden zo groot mogelijk

zou zijn. Op deze manier werden dus twee sollicitatieprocedures gesimuleerd, waarbij elke proefpersoon in totaal dus drie condities doorliep: eerlijk, archiefmedewerker en advocaat.

De twee banen (archiefmedewerker en advocaat) werden gekozen omdat voor deze banen de persoonlijkheidsprofielen sterk uiteenlopen. Voor de baan van archiefmedewerker werd bijvoorbeeld verwacht dat de proefpersonen hoger zouden scoren op de schaal Meegaandheid, terwijl voor de baan van advocaat juist werd verwacht dat proefpersonen lager op dit persoonlijkheidsfacet zouden scoren. De keuze van deze twee banen maakte *faken* dus moeilijker, waardoor nog beter kon worden ingeschat in hoeverre mensen de juiste antwoorden kunnen geven in een sollicitatieprocedure, en wat het effect van emotionele intelligentie hierop is.

Er werd een aantal – ook voor de selectiepraktijk – relevante bevindingen gedaan. Allereerst werden de gemiddelde scores op de facetten vergeleken tussen de eerlijke conditie enerzijds, en de archiefmedewerker- en advocaat-conditie anderzijds. Uit deze verschillen bleek dat de studenten zeer goed in staat waren om op de wenselijke manier te reageren gegeven de vooraf gedefinieerde persoonlijkheidsprofielen van de twee functies: wanneer een functie om hogere scores op een bepaald facet vroeg, dan ging de gemiddelde score van de studenten op dit facet omhoog in de sollicitatieconditie, en wanneer een lagere score behaald diende te worden voor een baan, dan ging de gemiddelde score van de studenten op dit facet omlaag. Deze resultaten repliceerden de bevindingen uit eerdere onderzoeken (Furnham, 1990; Raymark & Tafero, 2009) Ook bleek dat mensen die de gewenste resultaten in de ene sollicitatieconditie gaven, dit in sterke mate ook in de andere sollicitatieconditie deden. Bovendien bleek dat emotionele intelligentie, zoals voorspeld, gerelateerd was aan het geven van de juiste antwoorden in zowel de archiefmedewerker- als advocaat-conditie. Hoewel relatief klein in absolute zin waren de gevonden effecten nagenoeg gelijk in beide condities. Bovendien bleef het effect van emotionele intelligentie op het geven van de gewenste antwoorden bestaan wanneer voor de effecten van intelligentie en de vijf persoonlijkheidskenmerken van het Vijf Factor Model werd gecontroleerd in de voorspelling.

De resultaten beschreven in **Hoofdstuk 4** passen bij eerdere studies (Klehe et al., 2012; Kleinmann et al., 2011) die aangaven dat 'weten wat te zeggen en te doen' in selectiesituaties, dat wil zeggen sociaal wenselijk gedrag tonen, gerelateerd is aan sociale en cognitieve vaardigheden. Uiteraard kunnen mensen deze vaardigheden ook aanwenden op de werkvloer; uit eerder onderzoek bleek bijvoorbeeld dat een hogere mate van emotionele intelligentie tot betere werkprestaties leidt (Joseph et al., 2015; Joseph & Newman, 2010; O'Boyle et al., 2011). Voor de selectiepraktijk betekent deze bevinding dat de manier waarop iemand zich voordoet niet altijd gezien hoeft te worden als een vertekening van het beeld van deze persoon, maar eerder als een indicatie van de wijze waarop deze zich in de toekomst zal presenteren.

De studies besproken in **Hoofdstuk 2** en **Hoofdstuk 3** boden ondersteuning voor een inhoudelijke interpretatie van sociale wenselijkheid, zoals gerepresenteerd door de algemene persoonlijkheidsfactor, door relaties met (externe) criteria aan te tonen. Echter, de vraag kan worden gesteld in hoeverre de algemene persoonlijkheidsfactor afhankelijk is van situationele

en contextuele invloeden. Door deze vraag te beantwoorden wordt meer inzicht verkregen in de begripsvaliditeit van de algemene persoonlijkheidsfactor.

> *Onderzoeksvraag 4: Varieert sociale wenselijkheid in zelfrapportage persoonlijkheidsvragenlijsten met de testsituatie en de responsschaal van de vragenlijst?*

Een belangrijke alternatieve verklaring voor de aanwezigheid van de algemene persoonlijkheidsfactor is dat het een factor betreft die sociale wenselijkheid als *bias* meet, dat wil zeggen de mate waarin men zich beter voordoet dan men in werkelijkheid is (Bäckström et al., 2009; Petterson et al., 2012). Daarom ligt het voor de hand om te onderzoeken of de kenmerken van de algemene factor veranderen in selectiesituaties. In selectiesituaties kunnen we immers verwachten dat sollicitanten hun best zullen doen om een zo goed mogelijk beeld van zichzelf neer te zetten. In een adviessituatie zal iemand daarentegen minder gemotiveerd zijn om goed over te komen maar eerder authentiek willen antwoorden. Hieruit volgt dat als er bijvoorbeeld bij persoonlijkheidsmetingen in selectiesituaties een veel duidelijkere algemene factor naar voren komt dan in adviessituaties, dit ondersteuning zou bieden voor het idee dat de algemene factor het resultaat is van zelfoverschatting (MacCann et al., 2017). Door de invloed na te gaan van de testsituatie op de scores op de algemene factor kan dus inzichtelijk worden gemaakt in hoeverre sociaal wenselijkheid een gevolg is van de motivatie om zich beter voor te doen dan iemand daadwerkelijk is, dus in hoeverre mensen hun scores op een vragenlijst *willen* vertekenen.

Een andere belangrijke vraag is in hoeverre mensen hun scores *kunnen* vertekenen. De meeste persoonlijkheidsvragenlijsten hanteren een zogenaamde Likertschaal (vaak een 5-puntsschaal lopend van *Zeer mee oneens* tot *Zeer mee eens*). Dit type responsschaal maakt het vrij gemakkelijk om sociaal wenselijk te antwoorden (door simpelweg steeds *Zeer mee eens* in te vullen op positief geformuleerde items). Er zijn ook itemtypes waar het moeilijker is sociaal wenselijk te antwoorden, zoals zogenaamde geforceerde keuze items waarbij de respondent tussen twee stellingen (die allebei een andere persoonlijkheidstrek meten) dient te kiezen die even sociaal wenselijk zijn. Dit type responsschaal beperkt dus de mate waarin mensen sociaal wenselijk *kunnen* antwoorden als ze dat zouden willen.

De scores van 3980 personen die een persoonlijkheidsvragenlijst hadden ingevuld in verschillende situaties (advies- of selectie) en in verschillende vormen (met een Likert-antwoordschaal of met een geforceerde keuze antwoordschaal) werden gebruikt om de begripsvaliditeit van de algemene persoonlijkheidsfactor te onderzoeken. Deze data waren afkomstig uit de database van een grote testuitgever uit Nederland; het betrof hier dus echte selectie- en assessmentkandidaten. Door middel van het 2x2 design werden vier groepen (advies Likert, selectie Likert, advies geforceerde keuze en selectie geforceerde keuze) met elkaar vergeleken waarvan kan worden aangenomen dat ze verschillen in de mate waarin ze zich van hun beste kant *willen* en *kunnen* laten zien. De verwachting werd getoetst dat wanneer de algemene factor wat betreft interne structuur en grootte sterk zou verschillen tussen de vier groepen, dit erop zou duiden dat deze factor een weerspiegeling is van de mate waarin mensen

zich beter voordoen op vragenlijsten. Als anderzijds geen verschillen zouden voorkomen tussen de vier groepen, dan zou dit erop kunnen duiden dat de algemene factor een consistente factor is en minder onderhevig aan situationele omstandigheden waarvan we weten dat deze de mate waarin men goed over wil komen beïnvloeden.

Drie belangrijke resultaten kwamen uit de analyses naar voren. Ten eerste bleek dat de scores op de facetten, factoren, en de algemene factor over het algemeen hoger waren in de selectiesituatie dan in de adviessituatie. Dit resultaat repliceert eerdere bevindingen (Anglim et al., 2017; Van der Linden et al., 2011). Ten tweede bleek dat dit verschil kleiner was als de geforceerde keuze antwoordschaal in plaats van de Likertschaal werd gebruikt. Deze eerste twee resultaten geven aan dat de algemene factor voor een deel een meting is van de mate waarin iemand een te rooskleurig beeld van zichzelf geeft. Het derde en meest belangrijke resultaat was dat de factorstructuur van de vragenlijst, en ook de algemene factor, nagenoeg gelijk waren tussen de vier groepen, ondanks de verschillen in scores tussen de vier groepen. Bovendien was de algemene factor even groot (blijkend uit hoeveel variantie de algemene factor verklaarde in de onderliggende facetten) in alle vier de groepen. Deze drie hoofdbevindingen samen laten zien dat het niet aannemelijk is dat de algemene factor eenvoudigweg het gevolg is van het feit dat mensen zich op een zelfrapportage persoonlijkheidsvragenlijst beter voordoen dan ze in werkelijkheid zijn. Gebaseerd op deze bevinding, en op basis van eerdere studies en de bevindingen uit de eerdere hoofdstukken van dit proefschrift lijkt het aannemelijk dat iemands score op de algemene persoonlijkheidsfactor diens sociale effectiviteit weergeeft. Dit is een belangrijke conclusie voor de selectiepraktijk, omdat het impliceert dat een score op de algemene factor (1) gebruikt kan worden door de selecteur om de juiste kandidaat te selecteren en (2) een nuttige toevoeging zou kunnen zijn aan de feedback die gegeven wordt aan een kandidaat na het invullen van een persoonlijkheidsvragenlijst. Hoewel sollicitanten waarschijnlijk graag gedetailleerde feedback zullen willen krijgen op hun scores of facet- en domeinniveau (de vijf persoonlijkheidskenmerken), zal het terugkoppelen van de score op de algemene persoonlijkheidsfactor voor de kandidaat ook nuttig zijn, aangezien dit snel een algemeen beeld geeft van de sociale effectiviteit van de persoon in kwestie.

Conclusie

In scores op zelfrapportage persoonlijkheidsvragenlijsten is een sociale wenselijkheidscomponent zichtbaar. In dit proefschrift is onderzocht of er ondersteuning te vinden is voor een inhoudelijke, betekenisvolle interpretatie van deze component als een relevant individueel verschil tussen personen voor personeelsselectie. De resultaten in dit proefschrift laten zien dat de sociale wenselijkheidsfactor een redelijk sterke samenhang laat zien met werkgerelateerde uitkomsten zoals werkprestatie en *organizational citizenship behavior*. Bovendien blijken scores op deze factor gerelateerd te zijn aan meer generieke indicatoren van sociale effectiviteit, zoals de kwaliteit van sociale relaties en het hebben van minder conflicten met anderen. Deze betere sociale relaties resulteren op hun beurt weer in een betere gemoedstoestand en een hogere mate van welzijn. Daarnaast zorgen hogere scores op deze factor ervoor dat mensen minder aangeslagen zijn na een interpersoonlijk conflict. De

interne structuur van de algemene factor in persoonlijkheidsvragenlijsten bleek voorts consistent te zijn tussen verschillende situaties die variëren in de mate waarin mensen sociaal wenselijke antwoorden zullen geven. Deze bevinding komt overeen met een ander relevant resultaat uit dit proefschrift, namelijk dat het geven van de gewenste antwoorden in selectieprocedures positief gerelateerd is aan de emotionele intelligentie van een persoon. Samengenomen maken de resultaten in dit proefschrift het aannemelijk dat de algemene sociale wenselijkheidsfactor in zelfrapportage persoonlijkheidsvragenlijsten reflecteert in welke mate iemand sociaal effectief is. Deze factor kan dus betekenisvol zijn voor personeelsselectie.

Curriculum Vitae

Curriculum Vitae

Dirk Pelt was born on the 23th of December of 1986. After graduating the Christelijk Gymnasium Utrecht in 2004, he studied Sociology at Utrecht University. In 2007, during his bachelor studies, he spent 6 months in San Diego to do four elective courses at San Diego State University. He obtained his bachelor's degree in 2008, after which he enrolled in the research master program Sociology and Social Research (SaSR). After receiving the master's degree in 2010, he spent time traveling in India and working in Berlin. In 2012, he started working at Ixly, a test development and publishing company based in Utrecht, The Netherlands. Ixly develops innovative online tests for the measurement of personality, vocational interests, intelligence, and many other constructs relevant for selection and assessment of personnel. At Ixly, he developed a strong interest and acquired skills in psychometrics, testing, personality measurement, and the field of work and organizational psychology at large. He used his acquired skills to develop an adaptive test for cognitive abilities, and adaptive versions of questionnaires for personality, work values and vocational interest.

When, in 2014, the opportunity came to do a PhD under supervision of Marise Born and Dimitri van der Linden the Erasmus University Rotterdam, he did not hesitate to take this chance: because of his interests for research and the skills obtained in his academic and professional career, this was a next logical step forward. During his PhD, while still working at Ixly, he has published his research in international peer-reviewed journals and presented at international conferences on personality and selection and assessment across the globe, for example in Edinburgh, Oslo, Dublin and Hanoi. His first dissertation paper on the relationship between the General Factor of Personality (GFP) and work-related outcomes was nominated for the "Best paper award" at the 2015 conference of the Dutch Association of Work & Organizational Psychology. He is a member of the Dutch Association of Work & Organizational Psychology (WAOP), European Association of Work and Organizational Psychology (EAWOP), Dutch-Flemish Network for Selection Research, and European Network of Selection Researchers (ENESER).

Publications

International Peer-Reviewed Publications

Pelt, D. H. M., Van der Linden, D., & Born, M. Ph. (2018). How Emotional Intelligence Might Get You the Job: The Relationship Between Trait Emotional Intelligence and Faking on Personality Tests. *Human Performance*, *31*, 33-54.

Pelt, D. H. M., Van der Linden, D., Dunkel, C. S., & Born, M. Ph. (2017). The general factor of personality and job performance: Revisiting previous meta-analyses. *International Journal of Selection and Assessment*, *25*, 333-346.

Book Chapters

Van der Linden, D., Pelt, D. H. M., Dunkel, C. S., & Born, M. Ph. (2017). Personality, personnel selection, and job performance. In V. Zeigler-Hill & T.K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1-10). Basel: Springer AG.

Manuscripts Submitted for Publication

Pelt, D. H. M., Van der Linden, D., Dunkel, C. S., & Born, M. Ph. (2019). The motivation and opportunity for socially desirable responding does not alter the general factor of personality.

Conference Contributions

Presentations (selection)

Pelt, D. H. M., Born, M. Ph., & Oostrom, J. K. (2019). *Using cognitive pretesting to explore language and cultural effects on personality item bias*. 3rd World Conference on Personality, Hanoi, Vietnam.

Pelt, D. H. M., Van der Linden, D., & Born, M. Ph. (2018*). Factorial stability and the emergence of a general factor in personality questionnaires: The influence of test-taking context and instrument type*. European Network of Selection Researchers (ENESER) Meeting, Edinburgh, Scotland.

Pelt, D. H. M., Van der Linden, D., & Born, M. Ph. (2016). *How emotional intelligence (EI) might get you the job: The relationship between trait EI and faking personality tests*. European Network of Selection Researchers (ENESER) Meeting, Amsterdam, Nederland.

Pelt, D. H. M., Van der Linden, D., & Born, M. Ph. (2016). *Factorial stability and the emergence of a general factor in personality questionnaires: The influence of test-taking context and*

instrument type. 11th Dutch-Flemish Research Meeting on Personnel Recruitment and Selection, Groningen, The Netherlands.

Pelt, D. H. M., Van der Linden, D., Dunkel, C. S., & Born, M. Ph. (2015). *The General Factor of Personality and job performance: Meta-analytic evidence*. Dutch Association of Work & Organizational Psychology (WAOP) Conference, Amsterdam, The Netherlands (Nominated for *Best Paper Award*).

Pelt, D. H. M., Van der Linden, D., & Born, M. Ph. (2015). *Does the General Factor of Personality (GFP) reflect emotional intelligence and social knowledge?.* 17th Conference of the European Association for Work & Organizational Psychology (EAWOP)*, Oslo, Norway.*

Born, M. Ph., Fong, Y. M., Oostrom, J. K., & Pelt, D. H. M. (2015). *Examining item bias: Attempting to disentangle language and cultural effects on personality test scores*. 17th Conference of the European Association for Work & Organizational Psychology (EAWOP), Oslo, Norway.

Symposium

Pelt, D. H. M., & Van der Linden, D. (2017). *Social desirability in the selection context: Substance or artefact?.* 19th Conference of the European Association for Work & Organizational Psychology (EAWOP), Dublin, Ireland.

Posters

Janssen, N., Pelt, D. H. M., Mannheim, B., Schrijen, S. P. J. W., & Van Zand, D. (2019). *Comparing person-fit statistics in proctored vs unproctored testing in high- and low-stakes test situations.* 20th Conference of the European Association for Work & Organizational Psychology (EAWOP), Turin, Italy.

Mannheim, B., Pelt, D. H. M., Schrijen, S. P. J. W., Janssen, N., & Van Zand, D. (2019). *Practicing for cognitive tests: Is computerized adaptive testing the answer?.* 20th Conference of the European Association for Work & Organizational Psychology (EAWOP), Turin, Italy.

Schrijen, S. P. J. W., Pelt, D. H. M., Mannheim, B., Janssen, N., & Van Zand, D. (2019). *Validation of a digital divergent thinking task*. 20th Conference of the European Association for Work & Organizational Psychology (EAWOP), Turin, Italy.

Dankwoord (Acknowledgements)

Dankwoord

Het is volbracht! Het is vreemd om iets waar je vijf jaar aan gewerkt hebt ineens voor je te zien liggen. Vijf jaar: het klinkt ontzettend lang, maar gek genoeg is de tijd ook voorbij gevlogen. Ik ben heel trots op het resultaat dat het harde werken van mij, maar zeker ook van mijn begeleiders Marise en Dimitri, heeft opgeleverd. Dit zorgt er ook voor dat ik met een goed gevoel dit hoofdstuk – letterlijk en figuurlijk – af kan sluiten; het combineren van een baan met een promotietraject is soms hard werken (ook in de avond en weekenden) en de wetenschappelijke wereld ook niet altijd even zacht voor de tere ziel (afwijzingen, revisies, nog meer afwijzingen). Ik heb dan ook zin om weer nieuwe dingen te gaan doen en ook om weer nieuwe dingen te onderzoeken. Maar voordat ik daaraan kan beginnen wil ik hieronder eerst een aantal personen bedanken, zonder wie dit proefschrift er niet geweest zou zijn.

Dimitri, ik waardeer je praktische en realistische kijk op complexe zaken. Het was soms wel eens slikken als ik een paar dagen aan moeilijke statistische analyses had gewerkt en jouw reactie hierop ongeveer luidde: "Ik weet het niet, het is wel moeilijk allemaal…". Van jou heb ik ook geleerd wat goed schrijven is. In mijn geval was dit vooral veel schrappen; zonder jou was dit proefschrift waarschijnlijk twee keer zo dik en twee keer zo onbegrijpelijk geweest. Ik heb bewondering voor hoe je met heel weinig woorden complexe zaken heel erg *to-the-point* kunt beschrijven en inzichtelijk kunt maken. Hier heb ik veel van geleerd. Ik heb ook echt genoten van de inhoudelijke discussies die we hadden over individuele verschillen, persoonlijkheid, en de GFP. Ook hierin kon je dingen waar ik de hele dag mijn hoofd over zat te breken in een paar zinnen uitleggen waardoor het ineens heel duidelijk voor mij was. Ook de bemoedigende woorden na een afwijzing bij een tijdschrift – zeker wanneer deze niet mals waren vanuit de anti-GFP hoek – hebben mij goed gedaan. Jouw "het-komt-wel-goed"-mentaliteit werkte aanstekelijk, en zorgde ervoor dat ik met frisse moed doorging. Ik hoop dat we in de toekomst nog meer onderzoek kunnen doen.

En dan natuurlijk Marise, ik heb grote bewondering voor jouw bijna encyclopedische kennis van het veld. Je had altijd nieuwe inzichten, of kwam met nieuwe literatuur die ook relevant was voor mijn onderzoek. Dit heeft de kwaliteit van mijn proefschrift enorm naar boven gehaald. Ook met het schrijven – wat voor mij soms wel een worsteling was – was je van grote waarde: van jou heb ik geleerd hoe belangrijk het is om secuur te zijn en de lezer bij de hand te nemen. Ik ben je dan ook zeer dankbaar voor de uitgebreide feedback die ik altijd kreeg. Ook gaf je me op de juiste momenten het gevoel dat ik zelf de stuurman van dit project was, dat het belangrijk was dat ik zelf kon aangeven welke richting we op zouden moeten. Ik ben je ook dankbaar voor de kans om te presenteren op de *World Conference on Personality* in Hanoi, wat een mooi avontuur was. Wat ik ook altijd gewaardeerd heb is dat je altijd stilstond bij de successen die we behaalden: na een publicatie of een goed verlopen presentatie kreeg ik altijd een (digitaal) schouderklopje, wat ervoor zorgde dat ik nog trotser was op wat we samen bereikt hebben. Ik hoop dat we in de toekomst hier nog vele successen aan toe kunnen voegen.

I would also like to thank the members of the committee for taking the time to read and evaluate my dissertation. Henk van der Molen, Cornelius König, Rob Meijer, Eva Derous, and Janneke Oostrom. I feel honored that my work has been evaluated by such knowledgeable scholars in the field.

Grote dank ben ik ook verschuldigd aan Diddo: zonder jou was dit proefschrift er letterlijk niet geweest. Hoewel je mijn eerste werkgever bent en ik dus weinig vergelijkingsmateriaal heb, denk ik dat er weinig leidinggevenden zijn die zo gedreven en tegelijkertijd zo relaxt zijn als jij. Ik noem gedrevenheid omdat je denk ik zelden iemand tegenkomt die zo veel nieuwe ideeën heeft en kansen ziet. En relaxt in de zin dat als er aan het thuisfront iets is, of als je om wat voor reden dan ook eens thuis moet werken of uren op moet nemen dit nooit een probleem is. Die flexibiliteit en het vertrouwen in werknemers heb ik altijd als een groot voordeel ervaren. Ik waardeer het dat je mij de kans hebt gegeven om mij in de psychometrie en het test-vak te ontwikkelen, en mij de ruimte daarin gaf om hier zelf zaken in uit te zoeken en te ontdekken. En natuurlijk de kans om te promoveren "in de baas zijn tijd" (en geld). Ook hierin liet je mij volledig vrij, wat zeker niet altijd het geval is bij andere werkgevers. Qua onderwerp en studies heb ik grotendeels mijn eigen pad kunnen bewandelen, en ik kreeg altijd de kans om cursussen te volgen en naar congressen te gaan, wat ervoor gezorgd heeft dat ik een wetenschappelijk netwerk op heb kunnen bouwen. Dank hiervoor! Ik geniet erg van de inhoudelijke discussies die we hebben over nieuwe tests, moeilijke modellen, rekenregels en statistische methodes. Ik ben heel trots op de adaptieve testen die ik onder jouw supervisie heb kunnen ontwikkelen, en vind het te gek dat deze inmiddels gretig aftrek vinden in de markt. Ik hoop dat je zo groot blijft denken zoals je nu doet, want ik denk dat dit een grote drijvende kracht achter Ixly is. En hopelijk kan de Ixly-band ook weer ergens een optreden doen binnenkort!

Graag bedank ik ook de participanten die mee hebben gedaan aan mijn onderzoek; zonder jullie was dit proefschrift natuurlijk ook niet tot stand gekomen. In een deel van mijn studies heb ik gebruik gemaakt van data van derden; ook hen wil ik bedanken voor de mogelijkheid om deze data te gebruiken voor mijn onderzoek.

Ik wil hier ook twee groepen mensen noemen waarvan ik veel opgestoken heb in de afgelopen jaren. Dit zijn het Dutch-Flemish Network for Selection Research en het European Network of Selection Researchers (ENESER). De momenten waarop we samenkomen heb ik altijd als zeer inspirerend, leerzaam, maar ook erg leuk ervaren. Goed om te zien dat selectie- en assessmentonderzoekers elkaar opzoeken om ideeën uit te wisselen. Ik maak ook in de toekomst graag nog deel van deze twee netwerken uit. In het bijzonder wil ik hier nog even Reinout de Vries noemen; hoewel het bekend is dat jij niet erg gecharmeerd bent van de GFP, vond ik onze inhoudelijke discussies leuk en interessant. Het dwong mij om nog beter en dieper over het onderwerp na te denken, waarvoor dank. Ook Janneke wil ik speciaal nog even noemen: ik heb de samenwerking vanuit Ixly met jou altijd als heel prettig ervaren, en heb ook bewondering voor jouw onderzoekskwaliteiten en kennis van het veld.

Als buitenpromovendus heb je het voordeel dat je je in twee kringen tegelijk kunt bewegen: de wereld van het werk en de academische wereld. De mensen op beide plekken wil ik dan ook graag bedanken. Allereerst mijn collega's bij Ixly: hoewel er in de afgelopen jaren mensen zijn gekomen en gegaan, heb ik mij altijd helemaal thuis gevoeld bij Ixly. De sfeer is altijd goed, wat er toe bijdraagt dat we samen mooie en goede tests en software op de markt zetten. De gezellige (en enigszins chaotische) gezamenlijke lunches dragen daar sterk aan bij. Ook wil ik jullie bedanken voor de keren dat ik mijn presentaties op jullie mocht oefenen (ik hoop dat jullie er iets van begrepen hebben). Mijn R&D collega's wil ik nog even specifiek bedanken

voor de getoonde interesse in mijn project, en natuurlijk voor de champagne in het vliegtuig naar Turijn (voor EAWOP) om te vieren dat mijn proefschrift goedgekeurd was. Ook de overige uitstapjes naar congressen met het R&D team naar Münster, Oslo en Dublin waren leerzaam maar vooral ook heel gezellig. Met heel veel plezier denk ik ook terug aan de uitjes met heel Ixly naar Istanbul, Valencia, Lissabon en Gent. Ik vind het bijzonder dat collega's het ook zo goed met elkaar kunnen vinden buiten werktijd, en dit draagt dan ook heel sterk bij aan het werkplezier dat ik heb bij Ixly.

Ook wil ik de mensen van de A&O vakgroep aan de EUR – mijn tweede leefwereld – bedanken. Met name de PhD-kandidaten met wie we op donderdag ons onderzoek bespraken: hoewel ik er niet al te vaak bij kon zijn, hebben ze mij altijd erg geïnspireerd en ervoor gezorgd dat ik weer op een andere manier naar mijn eigen onderzoek ging kijken door reacties van anderen. Ook schept het een band om met andere uitverkorenen/lotgenoten even te kunnen schelden over wat die *reviewers* nu weer voor commentaar hadden. Arnold, die de PhD-meetings begeleidde, jou wil ik ook bedanken voor de kritische vragen die je stelde, jouw modelmatige manier van denken en de nieuwe inzichten die je mij gaf. Dit heb ik als zeer waardevol ervaren. In het bijzonder wil ik hier Keri nog even noemen: niet alleen hadden wij grotendeels dezelfde begeleiders, voor een deel overlapten ook de onderwerpen van onze projecten, waardoor ik het gevoel had dat we vaak "in hetzelfde schuitje" zaten. Jouw schuitje was wel aanzienlijk groter en mooier dan dat van mij, gezien het aantal en de kwaliteit van de studies die in jouw proefschrift belandden. Hiervoor werd jij helemaal verdiend cum laude onderscheiden, een topprestatie! Ik vind het knap dat je zo'n snelle denker bent, en leuk dat we door middel van de Rotterdam Emotionele Intelligentie Schaal (REIS) die we uitgeven bij Ixly nog samen kunnen werken.

Gelukkig is er in het leven meer dan werk, en heb ik een heleboel vrienden om leuke dingen mee te doen. De heren van de *MaarssenPosse* uiteraard: Tom, Tim, Daniël, Sharif, Roel, Olivier en Daan. Hoewel het jaarlijkse uitje er vorig jaar niet van is gekomen wegens laksheid van een aantal mensen (Tom en Sharif), weet ik zeker dat we elkaar altijd zullen blijven zien – en dat er altijd enorm gelachen zal blijven worden. Ook de FIFA-club wil ik bedanken voor de leuke potjes (en sorry voor de stukgegooide controllers). Geert-Jan en Jelte: nu ik klaar ben met dit verhaal, zullen we dan weer de gitaren oppakken en rocksterren worden? Het is wel tijd voor een comeback dacht ik zo. Geert-Jan, jij nog extra bedankt voor het beantwoorden voor al mijn PhD-gerelateerde vragen, en voor de gezelligheid van het afgelopen jaar. Je bent altijd welkom. Ik moet hier toch ook kort Tom en Godfried noemen, die andere rocksterren van Vriendje van Ferry. Vijf jaar geleden, toen ik hiermee begon, waren we op de top van onze roem. Ik had ook voor het DJ-en kunnen kiezen en misschien dat mijn leven er nu dan heel anders uit had gezien. Al denk ik dat een privéjet naar Ibiza er eerder inzit in de muziekwereld dan in de academische wereld (heb je toch de juiste keuze gemaakt, Tom). Tot slot Lion (en Loren en Pip natuurlijk), te gek dat mijn neef zo'n goede vriend van mij is. Ik dank jullie allemaal voor de interesse in mijn project, maar vooral voor de weekendjes weg, etentjes, festivals, feesten en gezelligheid in de afgelopen jaren.

Natuurlijk wil ik ook mijn familie bedanken. Mijn opa en oma, die regelmatig vroegen hoe het met "de studie" ging. Natuurlijk ook mijn broers (en paranimfen!), Daan en Joost; ik ben

blij dat jullie achter mij staan (letterlijk op de dag van de verdediging), en blij dat ik jullie ook tot mijn beste vrienden mag rekenen. Joost, goed dat je weer in Utrecht komt wonen zodat we elkaar wat vaker kunnen zien. Hilde (en natuurlijk Eri) en Jes: dank voor jullie meeleven de afgelopen vijf jaar! En de gezelligheid natuurlijk, onder andere in San Francisco en het pittoreske Susteren. Ik wil ook graag mijn vader bedanken: hoewel ik je helaas niet lang heb mogen kennen, heb ik wel een goed stel hersens van je gekregen, waar ik je dankbaar voor ben. Bovendien houd ik van normaalverdelingen en factoranalyses, maar vind ik het nog veel leuker om mensen aan het lachen te krijgen, en dat scheen jij ook graag gedaan te hebben. De belangrijkste persoon in dit rijtje is mijn moeder: niet te geloven dat jij er alleen voor kwam te staan en drie jongens (waarvan inmiddels twee door het leven gaan als doctor) in je eentje groot gebracht hebt. Ik vind het mooi om te zien en het ontroert me hoe trots je op ons alle drie bent. En hoe je altijd roept "zie je nou wel, ik zei het toch!" als ik weer eens veel te gestrest was voor een presentatie die uiteraard weer goed verliep. Bedankt voor dit geloof in mij. Ook jou wil ik bedanken voor het goede stel hersens: je refereert altijd naar een studie waaruit zou blijken dat intelligentie erfelijk is via de kant van de moeder, maar dit artikel heb ik tot op heden nog niet gevonden. Maar ik wil best geloven dat het waar is.

En tot slot wil ik hier uiteraard Roosmarijn en Ezra in het zonnetje zetten. Roosmarijn, dank voor je liefde, begrip en geduld de afgelopen vijf jaar. Voor de keren dat ik zei dat ik "er zo aan zou komen", om vervolgens om 2 uur 's nachts met vierkante ogen in bed te kruipen na het urenlang doen van moeilijke analyses (of in jouw woorden: "heb je weer te lang moeilijke sommetjes zitten doen?"). Voor de keren dat ik mijn presentaties op je mocht oefenen, waardoor je de term GFP inmiddels wel kunt dromen. Dit project heeft soms ook best wat van jou gevraagd, en ik ben heel dankbaar voor jouw flexibiliteit en onvoorwaardelijke steun die je mij altijd geeft. Wat een mooie reizen hebben we gemaakt in de afgelopen tien jaar (!), en ik hoop en weet zeker dat er nog heel veel gaan volgen. Ik heb zin om de wijde wereld met je in te trekken. En ik moet je natuurlijk ook bedanken voor het cadeautje dat al anderhalf jaar bij ons is: de liefste Ezra. Dit boekje is mooi, maar jij bent nog veel mooier. Wat een feest sinds je in ons leven bent. Ongelooflijk hoe snel jij je ontwikkelt – je praat al de oren van ons hoofd – en ik ben heel nieuwsgierig naar alle volgende stappen die je gaat zetten. Ik ben zo ontzettend trots op je, en ben erg benieuwd welk pad jij gaat bewandelen in het leven. Ik kan niet wachten om je nog meer van de wereld te laten zien en kijk uit naar de avonturen die we samen met je liefste moeder gaan beleven.