

# Improving personality and Interest measurement for purposes of selection and assessment

Djurro Holtrop





# **Improving personality and interest measurement for purposes of selection and assessment**

Djurre Holtrop

Cover design: Sander Cornet (sander.cornet@gmail.com)

Printing: ...

About the cover: Personality consists of six dimensions: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to experience. Vocational interests consist of three dimensions: Data versus Ideas, People versus Things, and Prestige interests. The characters on the cover represent six combinations of a personality trait and an interest: The creative chef, the non-integer manager, the precise banker, the neurotic teacher, the extraverted painter, and the friendly maid. Which personality traits and vocational interests do you think these characters combine?



VRIJE UNIVERSITEIT

Improving personality and interest measurement for purposes of selection and  
assessment

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan  
de Vrije Universiteit Amsterdam,  
op gezag van de rector magnificus  
prof.dr. V. Subramaniam,  
in het openbaar te verdedigen  
ten overstaan van de promotiecommissie  
van de Faculteit der Gedrags- en Bewegingswetenschappen  
op dinsdag 17 mei 2016 om 13.45 uur  
in de aula van de universiteit,  
De Boelelaan 1105

door

Djurre Joachem Holtrop  
geboren te Rotterdam

promotoren: prof.dr. M.Ph. Born  
prof.dr. R.E. de Vries  
copromotor: dr. R.H. van den Berg

The research presented in this dissertation was supported by Stichting NOA.

# Table of contents

Chapter 1: Improving personality and interest measurement for purposes of selection and assessment: An introduction .....	7
Chapter 2: A matter of context: A comparison of two types of contextualized personality measures.....	23
Chapter 3: Predicting performance with contextualized inventories, no Frame-of-Reference effect? .....	39
Chapter 4: Differential predictive validity of personality instruments for academic performance: Person-situation interaction for ethnic minorities .....	49
Chapter 5: Relating the Spherical representation of vocational interests to the HEXACO personality model .....	75
Chapter 6: Perceptions of vocational interests: Self-other agreement, assumed similarity, similarity, and reciprocity in student-parent dyads.....	97
Chapter 7: Improving personality and interest measurement for purposes of selection and assessment: Summary and general discussion .....	121
Chapter 8: Nederlandse samenvatting (Dutch summary) .....	143
References .....	165
Dankwoord.....	175



# Improving personality and interest measurement for purposes of selection and assessment: An introduction



1



In 1795, Napoleon Bonaparte formally introduced the metric system to allow his administration to communicate—with absolute clarity—about distance, weight, and volume. The golden standard for a meter, a platinum and iridium bar, is kept in a controlled environment in the *Bureau International des Poids et Mesures* in Paris. Measurement in psychological science, however, does not adhere to such a golden standard; there is no inscribed tablet, with a golden standard inventory, safely tucked away in a pressurized environment. In psychological science, measures are continually improved and, as such, the standard continually evolves. In this dissertation, the tradition of improving psychological measures is continued in two major domains of non-cognitive individual differences. In the domain of personality, we tested how the inclusion of a situational description to existing measures improves the measures' predictive validity and participant reactions. In the domain of vocational interests, we validated a new measure of vocational interests and explored how 'others' can contribute to the measurement of vocational interests.

Personality and vocational interests are often measured with self-report inventories. These inventories consist of a number of items, such as *'I want to be the best'* for personality and *'I would like to defend people in court'* for vocational interests. Participants respond on a Likert scale by indicating to which degree the item applies to them, or how much they like the activity. In the domain of selection and assessment, these measures are then mostly used to predict career choice (e.g., Lent, Brown, & Hackett, 1994) and academic (e.g., Richardson, Abraham, & Bond, 2012) or work outcomes (e.g., Schmidt & Hunter, 1997), such as performance or counterproductive behavior.

In this introductory chapter a short overview of the purpose of the studies in this dissertation is provided. Second, research is summarized which investigated how the inclusion of a situation (also called a Frame-of-Reference or FoR) to a personality measure affects its predictive validity. Third, research on the structure and measurement of vocational interests is presented. Lastly, this introduction concludes with a brief summary of the studies in the present dissertation. This dissertation then features five empirical chapters that each describe a study into the structural and predictive validity of personality and vocational interests. Figure 1 gives an overview of these studies and their main topics.

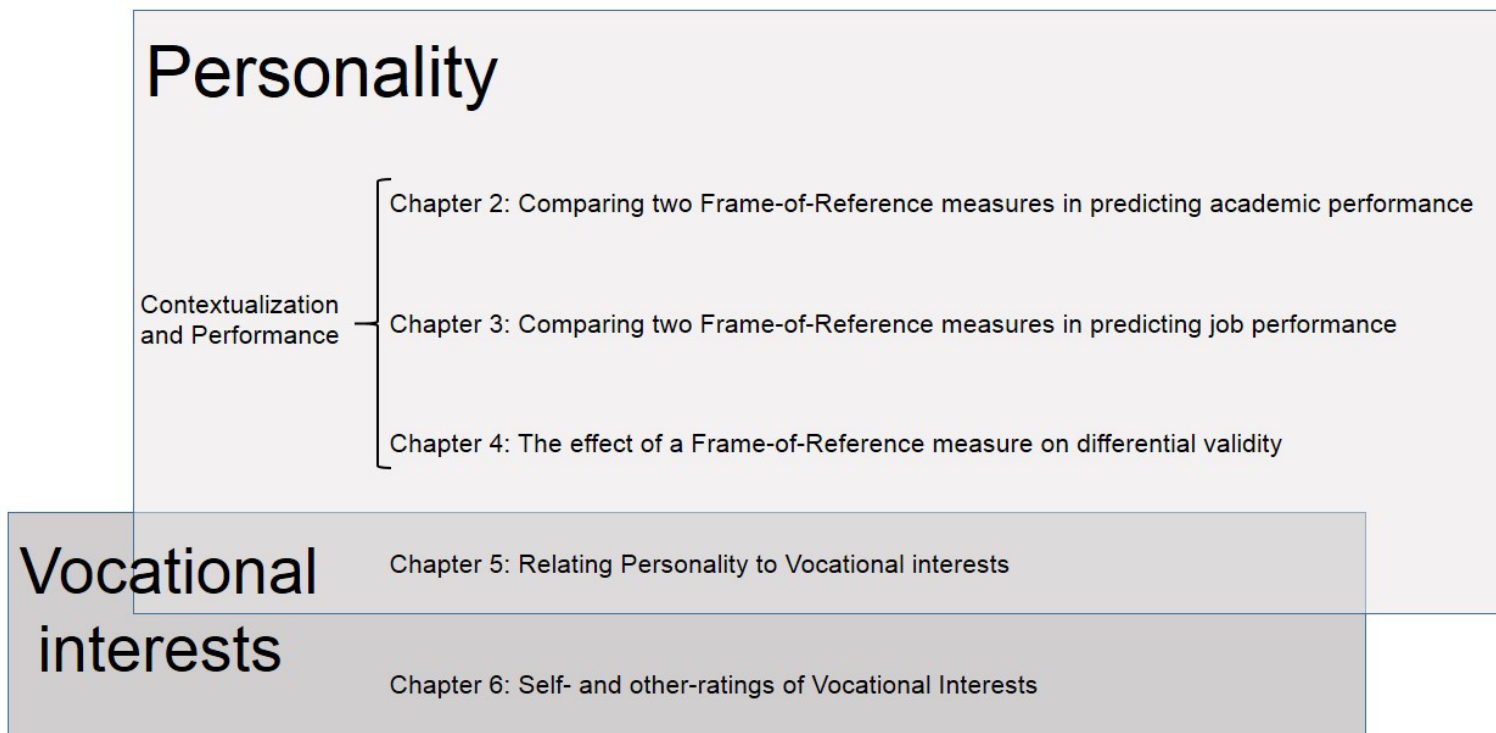


Figure 1. Overview of the empirical chapters in this dissertation.

Chapters two, three, and four in this dissertation address the contextualization of personality inventories. Contextualization is the process of adding a situation (or FoR) which is relevant to the criterion that is being predicted, to an otherwise situation-less personality inventory (e.g., Mount, Barrick, & Strauss, 1994; Schmit, Ryan, Stierwalt, & Powell, 1995). If, in this process, an academic situation is added to the previous example *'I want to be the best'* this item may then be changed to *'I want to be the best at school'*, or *'I want to be better than other students'*. The addition of a FoR is hypothesized to improve the criterion validity of a personality inventory (e.g., Shaffer & Postlethwaite, 2012). This improved criterion validity is called the Frame-of-Reference effect (FoR effect; Schmit et al., 1995). Chapter two and three explore which method of contextualization has the largest FoR effect and chapter four if the FoR effect reduces differential predictive validity. Differential validity refers to the phenomenon that some tests are more predictive of outcomes, such as performance, for one group (usually the majority) than for another (usually a minority).

Next, chapter five describes the relation between a recent measure of personality and a recent measure of vocational interests. During the last two decades, a revised model emerged in both domains. For personality this is the HEXACO model (Lee & Ashton, 2004), and for vocational interests the Spherical representation of interests (Tracey, 2002; Tracey & Rounds, 1996).



The HEXACO model distinguishes itself from previous models by the addition of the Honesty-Humility trait, which is similar to integrity. The Spherical representation's distinguishing feature from earlier vocational interest measures is the addition of the Prestige interest dimension. This dimension refers to the job level a person is interested in, and to how much a person desires to be challenged and put in effort. Honesty-Humility and Prestige interests seem conceptually (negatively) related to each other. In chapter five the relation between these two revised models is investigated.

Finally, chapter six describes how other-ratings can be used to measure vocational interests. So far, very little research has looked at the use of other-ratings of vocational interests (exceptions are Nauta, 2012; Nelling, Kandler, & Riemann, 2015), whereas other-ratings have often been used in personality research (e.g., Costa, McCrae, & Dye, 1991; Kenny, 2004; Zettler, Lang, Hülshager, & Hilbig, 2015). These studies show that several other-reports of personality have a higher predictive validity than self-rated personality (Connelly & Ones, 2010). Possibly other-rated vocational interests can predict successful vocational choice more accurately than self-rated interests. However, before other-ratings of vocational interests can be used in practice more knowledge is required about their fundamental attributes. Therefore we compared self- and other-ratings of vocational interests within parent-child dyads.

## **1. Measurement of Personality: The effect of a Frame-of-Reference**

### *1.1 Personality models*

Personality is commonly measured using the Big Five dimensions: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (Schneider, 2007). This structure of personality was uncovered by several psycholinguistic studies analyzing a large number of personality-relevant adjectives (e.g., Costa & McCrae, 1992; Goldberg, 1990). However, as several researchers have indicated (Lee & Ashton, 2004; Schneider, 2007), Integrity (or: Honesty-Humility) complements these five factors. Based on subsequent lexical studies, that have used the same approach as the earlier Big Five studies, six personality dimensions have been distinguished in more than twelve countries (e.g., Ashton et al., 2004; Ashton, Lee, & Goldberg, 2004; Ashton, Lee, Marcus, & De Vries, 2007; Lee & Ashton, 2008; Szarota, Ashton, & Lee, 2007; Wasti, Lee, Ashton, & Somer, 2008). These six dimensions are known by the acronym HEXACO, which stands for: Honesty-Humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A),

Conscientiousness (C), and Openness to experience (O; Lee & Ashton, 2004), in which Honesty-Humility closely resembles Integrity (Marcus, Lee, & Ashton, 2007). The Honesty-Humility scale measures the tendency to be sincere, fair, modest, and to avoid greed. Another change, compared to the Big Five model, is that in the HEXACO model the personality dimensions Emotionality (known as Emotional Stability in the Big Five model) and Agreeableness are rotated. This means that content associated with temper and irritability has moved from Big Five Emotional Stability to HEXACO Agreeableness, and content associated with sentimentality from Big Five Agreeableness to HEXACO Emotionality.

Personality is known to predict important job and academic outcomes, especially through the dimensions Conscientiousness and Honesty-Humility. For example, Schmidt and Hunter (1998) meta-analytically showed that, among personality dimensions, Conscientiousness and Integrity provide most incremental validity over intelligence measures for predicting work performance. Moreover, Poropat (2009) meta-analytically showed that Conscientiousness is an equally important predictor of academic performance as Intelligence is. Other studies have shown that Honesty-Humility offered incremental validity over the other five personality dimensions in the explanation of behavior and performance at work and in academe (Ashton & Lee, 2008; Johnson, Rowatt, & Petrini, 2011; Oh, Le, Whitman, Kim, Yoo, Hwang, & Kim, 2014).

As the above-mentioned and other studies have offered better support for six rather than five dimensions of personality, personality is measured as a six dimensional model in this dissertation. We measured the six dimensional model with two inventories: The Dutch HEXACO Personality Inventory Revised (HEXACO-PI-R; De Vries, Ashton, & Lee, 2009; Lee & Ashton, 2004) and the Multicultural Personality Test - Big Six (MPT-BS; De Vries, De Vries, & Born, 2010; NOA, 2009). Previous research has shown good convergent validity and reliability of these two inventories (De Vries et al., 2010; NOA, 2009).

### *1.2 Personality x Situation: The Frame-of-Reference effect*

To further improve the criterion validity of personality measures, researchers have started to investigate contextualized inventories (e.g., Bing, Whanger, Davison, & VanHook, 2004; Hunthausen, Truxillo, Bauer, & Hammer, 2003; Lievens, De Corte, & Schollaert, 2008; Reddock, Biderman, & Nguyen, 2012; Robie, Schmit, Ryan, & Zickar, 2000; Schmit et al., 1995;

Shaffer & Postlethwaite, 2012). Contextualization occurs when a relevant context is added to a personality inventory. There are three commonly applied methods to add context to a personality inventory: 1) Instructional contextualization, 2) Tagged contextualization, and 3) Complete contextualization. Instructional contextualization asks a participant to think of a certain situation (e.g., school) while filling out a generic questionnaire. Tagged contextualization modifies generic personality statements with an added tag. For example, the statement '*I am a busy person*' can be modified to '*I am a busy person at school*'. Complete (or full) contextualization occurs when an item is completely redesigned to match a context. For example, the statement '*People think I show a lot of effort*' can be modified to '*People think I study hard*'.

Regardless of the contextualization method, the added situation is often called a Frame-of-Reference (FoR; Mount, et al., 1994; Schmit et al., 1995). The idea behind adding a FoR is that personality is more consistent within one situation than when it is aggregated across several situations (Mischel, & Shoda, 1995). Moreover, measuring personality within a relevant situation improves the criterion validity of personality inventories for criteria relevant to that situation (e.g., personality at school predicts school performance). This improved criterion validity is called the FoR effect (Schmit et al., 1995).

The FoR effect was first proposed by Mount et al. (1994). They argued that behavior can be more accurately predicted if observers have a criterion-relevant context in mind when filling out a self-report personality questionnaire. Schmit et al. (1995) empirically tested the FoR hypothesis for employees and students. To this end they modified a measure of the Big Five personality model, the NEO-FFI (Costa & McCrae, 1989). For the employees, they added an 'at work' tag added behind every generic item to reflect a work setting, and, for the students, they added an 'at school' tag. Since then, several studies have contextualized personality inventory items by *tagging* them with a certain setting (Bing et al., 2004; Lievens et al., 2008; Robie et al., 2000). Some studies applied the FoR with a different method. For example, Hunthausen et al. (2003) *instructed* participants to consider how they are at work (i.e., instructional contextualization), and Pace and Brannick (2010) completely *rewrote* all items to reflect a work setting (i.e., complete contextualization). Independent from each other, these studies showed that contextualization of personality inventories improves the criterion validity for behaviors in that setting. Shaffer and Postlethwaite (2012) meta-analytically investigated 86 studies and confirmed the FoR effect.

The studies discussed above used tagged contextualization, instructional contextualization, or complete contextualization to add a FoR to an item. However, it is not yet clear which contextualization method is preferable. Tagged and complete contextualization are the strongest contextualization methods because each item reminds a participant about the relevant situation. Lievens et al. (2008) suggested that adding more content to each item (i.e., complete contextualization) may improve the criterion validity more than only tagging an item. However, so far, these two methods have not been directly compared in terms of their predictive validity. Therefore, in chapter two and three the expectation is tested that completely contextualized inventories outperform tagged inventories in terms of predictive validity. For this purpose participants filled out a generic (as a baseline measure), tagged, and contextualized personality measure, which were compared on their predictive validity. The design process for the contextualized inventories is presented in chapter two.

**Research Question 1:** Is a completely contextualized personality measure more predictive of academic and work performance than a tagged personality measure?

### *1.3 The Frame-of-Reference-effect and participant reactions*

In addition to investigating the criterion validity of tagged and completely contextualized inventories, we also looked at the participants reactions to these inventories. Adding a FoR may also improve participant reactions, because a participant may have an easier time filling out the inventory if there is no ambiguity about the situation (Wright & Mischel, 1987). To our knowledge, only Holtz, Ployhart, and Dominguez (2005) studied the effect of contextualization on participant reactions. Contrary to their expectations, they found no effect of contextualization on the affective feelings about the organization (to which the applicants applied). For example, Holtz et al. found that the participants who had filled out a contextualized measure, instead of a generic measure, were not more likely to recommend the organization to others. In our studies, different participant reactions were selected than those used by Holtz et al.. We focused on the reactions to the measures themselves and not on how these reflected on the administering organization. We investigated if contextualized measures change 1) how much the participants liked the test (e.g., Wiechman, & Ryan, 2003), 2) how relevant the participants found the test for their tasks (i.e., face validity; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993), and 3) how predictive the participants found the test for their performance (i.e., perceived predictive validity; Smither et al.,

1993). We expected that participants may find the (tagged and completely) contextualized measures more relevant and predictive. Additionally we expected that tagged measures may be liked less, because it is very repetitive if every item is tagged with “at work” or “at school”. In chapter two and three we further investigated our assumption that completely contextualized measures receive more positive participant reactions than tagged measures.

**Research Question 2:** Is a completely contextualized personality measure more positively perceived by participants than a tagged personality measure?

*1.4 Does the Frame-of-Reference-effect reduce differential validity for non-western minorities?*

Differential validity means that a psychological measure may be predictive of one group’s behavior or performance and less of another group’s. Differential validity has been found between different ethno-cultural groups, and usually implies that a measure is predictive for the ethnic majority and not (as much) for ethnic minorities. Most research into differential validity has investigated the measurement of cognitive individual differences (e.g., Katzell & Dyer, 1977; Te Nijenhuis & Van der Flier, 2000). However, some scholars have investigated and found differential validity of personality inventories (e.g., De Meijer, Born, Terlouw, & Van der Molen, 2008; De Vries, Born, & De Vries, 2012). For example, one such study (De Vries et al., 2012) found that Conscientiousness predicted academic performance among Dutch majority students, but not among non-western minority students. De Vries et al. proposed, as a possible explanation, that the behavior (and subsequent academic performance) of minority students may be mostly influenced by strong situations, such as their home situation, and not as much by their personality. In The Netherlands, 18.871 (14%) of all first year higher education students belong to a non-western ethnic minority (CBS, 2015a). Practically, this means that commonly used personality measures for selection and assessment may not be predictive for a significant portion of all students. For such a large group, the development and use of equally predictive measures seems crucial.

Some scholars have suggested that contextualization may actually reduce the differential validity of personality measures (e.g., Church, 2010; De Vries et al., 2012). Previous research has shown that people from non-western cultures perceive personality differently from westerners (e.g., Cousins, 1989) and that a situation has a much larger effect on how non-westerners describe their personality than on how westerners describe themselves. For example,

Asians may describe themselves different at work than with family, whereas Europeans may describe themselves more-or-less similar in these two settings. Cousins (1989) concluded that people from a non-western (in this specific study Japanese) culture may experience different—but equally valid—processes to construct their self-perception. Church suggested that, overall, contextualized measures may better capture the differential impact of situations on trait-relevant behaviors. Therefore, contextualized measures may show less differential validity across cultural groups. In chapter four we attempt to answer our final research question about contextualization.

**Research Question 3:** Does contextualization reduce the differential validity of personality measures across ethnic groups?

## **2. Measurement of Vocational Interests: Prestige interests, profile elevation, and other-ratings**

Chapters two to four investigate the effects of contextualization on personality measurement. Chapters five and six address the measurement of a different sort of individual differences, namely vocational interests.

### *2.1 Vocational interest models*

Vocational interest measures allow us to understand in which areas an individual would like to work or study. These measures usually ask people to indicate how much they like a number of jobs and activities and then return an interest score on several broad themes (e.g., Artistic interests and Social interests). Practically, vocational interest measures are mostly used for vocational counseling purposes. In this setting, these measures help people to structure their vocational preferences when they are undecided about which vocation to pursue, or which study program they would like to sign up for. For example, in The Netherlands, 54% of all aspiring students fills in a vocational interest measure (Markteffect, 2011).

Often, vocational interests are measured as six broad themes (or so-called ‘types’), which first have been proposed by John Holland (1959; 1997). Holland combined these themes into the acronym RIASEC: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. These themes represent six evenly distributed points on a circumplex, in the same order as in the acronym. Since the introduction of the RIASEC model, two major extensions have been proposed to its underlying structure.

First, Prediger (1982) proposed that two main dimensions underlie the RIASEC circumplex, namely Ideas/Data and People/Things. Idea tasks are

intrapersonal (i.e., occur within an individual's mind) and involve theories, knowledge, and creativity. Data tasks are impersonal and involve facts, numbers, and systematic procedures. People tasks are interpersonal (i.e., occur between people) and involve caring, persuading, and entertaining others. Things tasks are nonpersonal tasks that involve machines, materials, and tools. Figure 2 visualizes these main dimensions in respect to the RIASEC dimensions.

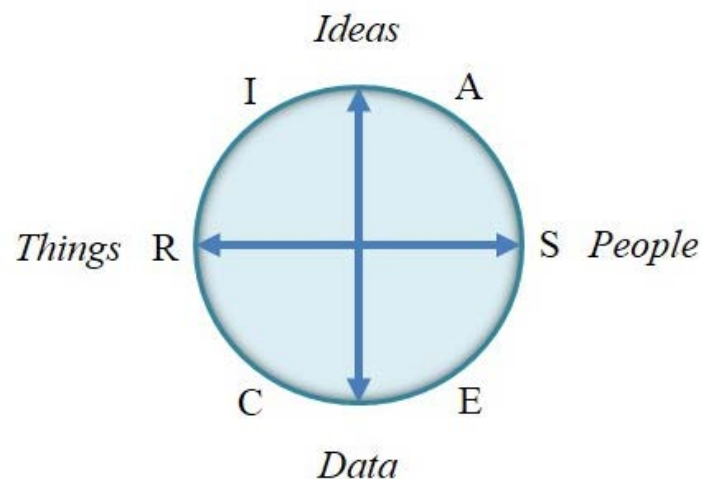
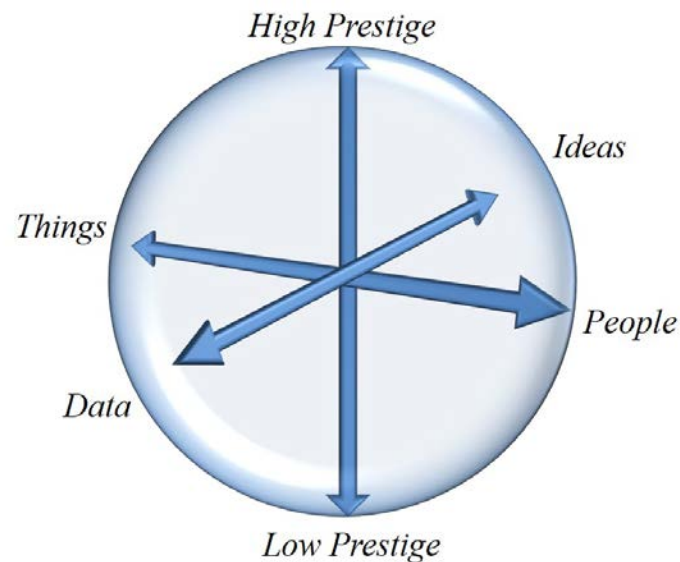


Figure 2. RIASEC dimensions (Holland, 1959) and Prediger (1982) dimensions of vocational interest. R = Realistic, I = Investigative, A = Artistic, S = Social, E = Enterprising, and C = Conventional.

Second, Tracey and Rounds (1996) proposed to include so-called Prestige interests as an additional dimension of vocational interest to the circular representation of vocational interest. Prestige interests represent the general difficulty, training, knowledge, education, and effort required of the activities a person is interested in (e.g., Roe, 1956; Sodano & Tracey, 2008). Some are interested in activities that require extensive training and/or a lot of effort, such as “defending people at court”. Others are interested in activities that require less training and/or less effort, such as “carrying or loading containers”. Tracey and Rounds showed that the Prestige interests dimension is orthogonal to the Ideas/Data and People/Things dimensions, changing the circular representation into a spherical representation. Figure 3 shows the resulting Spherical model visually. Prestige interests have also been discovered in previous research by, for example, Roe and Klos (1969), but have never been identified as a key dimension of vocational interests. Tracey and Rounds explained that Prestige interests were previously not recognized as a dimension of vocational interests, because most vocational interest measures are restricted to a limited range of occupational level and difficulty

(i.e., either assessed interests in lower or higher level activities). Because Prestige interests are strongly related to occupational level and general difficulty, most (range restricted) measures did not cover the full Prestige dimension, and therefore the complete spherical representation. Hence, most studies could not discover the three-dimensional structure of interest.



*Figure 3.* A visual representation of the three main dimensions of the Spherical representation of vocational interests (Tracey & Rounds, 1996).

To measure Prestige interests, Tracey (2002) developed the Personal Globe Inventory (PGI). The PGI is currently the only inventory that measures vocational interests as modeled by the Spherical representation. Using translated versions of the PGI, the Spherical representation's structure has been confirmed in several countries (e.g., Irish, Chinese, and Croatian; Darcy, 2005; Long, Adams, & Tracey, 2005; Sverko, 2008). In chapter five we investigated the relation between interests and personality and in chapter six we investigated other-ratings of vocational interests. To measure the complete range of vocational interests in these studies, as proposed in the Spherical representation, we translated the complete PGI to Dutch. The translation and validation of the full PGI is described in chapter five and the validation of the abbreviated version (PGI-short; Tracey, 2012) in chapter six.

## *2.2 Vocational interests' profile elevation: Trait or error?*

Scales in vocational interest measures always share a large overarching factor explaining approximately 35% of the variance in interest scores of adults and even more for children (Prediger, 1998). In this dissertation this overarching factor is interpreted as so-called profile elevation. Profile elevation refers to how high a person rates him/herself on all vocational interests on



average. Note that, because profile elevation is often used in applied settings (e.g., for vocational counseling purposes), we elected to use this approach to the general factor of interests, instead of more sophisticated approaches<sup>1</sup>. Surprisingly, whereas practitioners attach significant value to profile elevation, scholars still debate its meaning. Some scholars argue that it is a meaningful construct, others argue that it contaminates the measurement of vocational interests.

Tracey (2012) described three current interpretations of profile elevation. First, some scholars view profile elevation as 'substantive' (e.g., Darcy & Tracey, 2003; Fuller, Holland, & Johnston, 1999; Hirschi & Läge, 2007) and suggested that profile elevation is a meaningful factor that can be used in vocational counseling. Following this interpretation, some scholars have argued that profile elevation may be interpreted as interest flexibility (Darcy & Tracey, 2003), implying that people scoring high on profile elevation would be less hindered by a mismatch between the environment and their main interests. As a meaningful factor, profile elevation has found to be related to, for example, someone's activity in career planning and career exploration (Hirschi & Läge, 2007), and personality traits such as Openness (Fuller et al., 1999).

Second, some view profile elevation as a 'nuisance' or 'error' factor (e.g., Prediger, 1998) that causes systematic error and that its influence should be minimized when measuring vocational interests. As a way around the ambiguity of profile elevation, Prediger suggested to use alternative scoring procedures instead of raw or normed scores, such as high point scores. High point scores are a combination of the two or three highest interest scores (e.g., RIA would mean that a person is most interested in Realistic tasks, second in Investigative tasks, and third in Artistic tasks). High point scores are an effective method of summarizing vocational interests' profile shape/differentiation and this method appears largely unrelated to profile elevation. In essence high point scores represent vocational interests relative to each other within a participant.

Third, in line with the nuisance interpretation, some scholars view profile elevation as an 'artifact' (e.g., Tracey, 2012) and suggested that it is a systematic (self-)rater bias that influences all interest scales' relations to other

---

<sup>1</sup> An example of such an approach could be to compute the general factor scale for each participant based on the regression weights of individual items on the first factor of an unrotated principal component analysis. Profile elevation is strongly correlated, but not identical, to such a scale.

variables and that its influence can be reduced by means of ipsatization of the interest scale scores. Note that the latter two interpretations are highly similar to each other, in that they both consider profile elevation error caused by rater bias. The difference between both is that the ‘nuisance’ interpretation suggests that separate interests scale scores are less useful because they are contaminated with profile elevation, whereas the ‘artifact’ interpretation suggests that separate interest scales are useful, but that they need to be corrected for profile elevation.

In chapter five and six profile elevation is investigated and an attempt is made to establish whether it is substance or a nuisance/artifact. In chapter five, we followed the tradition of previous research (Fuller et al., 1999) and related profile elevation to personality dimensions. If profile elevation correlates to substantial traits then it may have substance. Subsequently, in chapter six, we described the first study into profile elevation with other-ratings. Substantial traits can be accurately perceived by others (Kenny, 1994), therefore we tested if other-rated profile elevation shows agreement with self-rated profile elevation, or if others cannot rate profile elevation accurately. If profile elevation shows high self-other agreement this would be in favor of the substance interpretation, whereas low self-other agreement would be in favor of the nuisance/artifact interpretation.

**Research Question 4:** Is profile elevation in interest measurement a substantive factor or a nuisance/artifact?

### *2.3 Vocational interest and personality*

It is important to precisely understand how personality and vocational interests measures are related, to more fully explain the behavior that follows (Armstrong & Anthony, 2009). A few years before Lee and Ashton (2004) introduced the HEXACO model, Tracey and Rounds (1996) introduced the Spherical representation of vocational interest. In chapter five, the relation between these two models is investigated. This investigation is particularly interesting because, compared to their predecessors, the additions to the revised personality (Honesty-Humility) and vocational interest model (Prestige interests) may seem conceptually (negatively) related to each other.

Several earlier studies have investigated the relations between personality measures and vocational interest measures. Holland (1959) even described the RIASEC vocational interest themes as personality typologies. Several studies (e.g., Armstrong & Anthony, 2009; Barrick, Mount, & Gupta, 2003; Larson, Rottinghaus, & Borgen, 2002) found a number of moderate

relations between several RIASEC scales and the Big Five personality dimensions. Recently, the HEXACO model of personality was also included in these investigations (McKay & Tokar, 2012; Pozzebon, Visser, Ashton, Lee, & Goldberg, 2010), and found to explain more variance in vocational interest than the Big Five. So far, no research has related Prestige interests to personality. If we understand which traits are related to Prestige interests we may be able to better explain the motives and behavior of people working in high and low prestigious jobs.

**Research Question 5:** Prestige vocational interests are related to which personality dimensions?

#### *2.4 Other-ratings of vocational interests*

Vocational interest measures are commonly measured with self-reports. However, asking observers to provide other-ratings may offer new and important information regarding a person's vocational interests. First, whereas it is sensible to first ask people themselves about their preferences, at times self-reports may not be informative. Self-reported vocational interests can be hard to interpret, because—for example—a candidate's profile does not show any differentiation between the interests measured. This latter phenomenon is called profile flatness (e.g., Sacket & Hansen, 1995). Profile flatness hinders the use of self-reported vocational interests because, if a candidate scores equally high on all interests, a counselor may be unable to single out the areas a candidate is most interested in. Hence, this profile flatness hinders finding a person's preferred interests. Therefore, for people who cannot differentiate between their own interests, it may be useful to measure interests with other-reports. Second, even for a person with differentiated interests, other-reports may yield new insights in vocational preferences. Take, for instance, a vocational interest measure that requires participants to indicate how much they like jobs. A participant may not fully understand the content of each of these jobs. Yet, some observers may fully understand the content of the job and also know the participant well, and may therefore be able to better understand to what degree the participant would like the job. Third, several other-reports may be more accurate than one self-report. For example, a meta-analysis (Connely & Ones, 2010) showed that several other-ratings of personality had higher criterion validity than one self-rating. Other-reports of vocational interests may thus help vocational counselors to gain more insight in the interests of their clients, especially when the clients' profiles do not differentiate between the interest scales.

However, not much is known about other-reports of vocational interests. So far, only two studies (Nauta, 2012; Nelling, Kandler, & Riemann, 2015) have investigated other-reports of vocational interests. Personality, on the other hand, is regularly measured with a combination of self- and other-ratings (e.g., De Vries, 2010; Watson & Clark, 1991). Nauta (2012) showed that the agreement between self- and other-ratings (i.e., the correlation between self- and other-ratings) is approximately similar for vocational interests and personality. Besides self-other agreement, there are several other possible combinations of self- and other-ratings. For instance, in a dyadic setting, personality research found significant 1) assumed similarity (i.e., the convergence of self- and other-ratings by one person, how similar a person thinks (s)he is to the other), 2) similarity (i.e., the convergence of two self-ratings, how similar two people are), and 3) reciprocity (i.e., the convergence of two other-ratings, how similar two people perceive each other). In chapter six, Nauta's research is extended and the fundamental characteristics of self- and other-ratings of vocational interests are investigated in a dyadic setting. If these characteristics are similar to those found in personality research, then research on other-ratings of vocational interests could tap into the wealth of knowledge about other-ratings of personality.

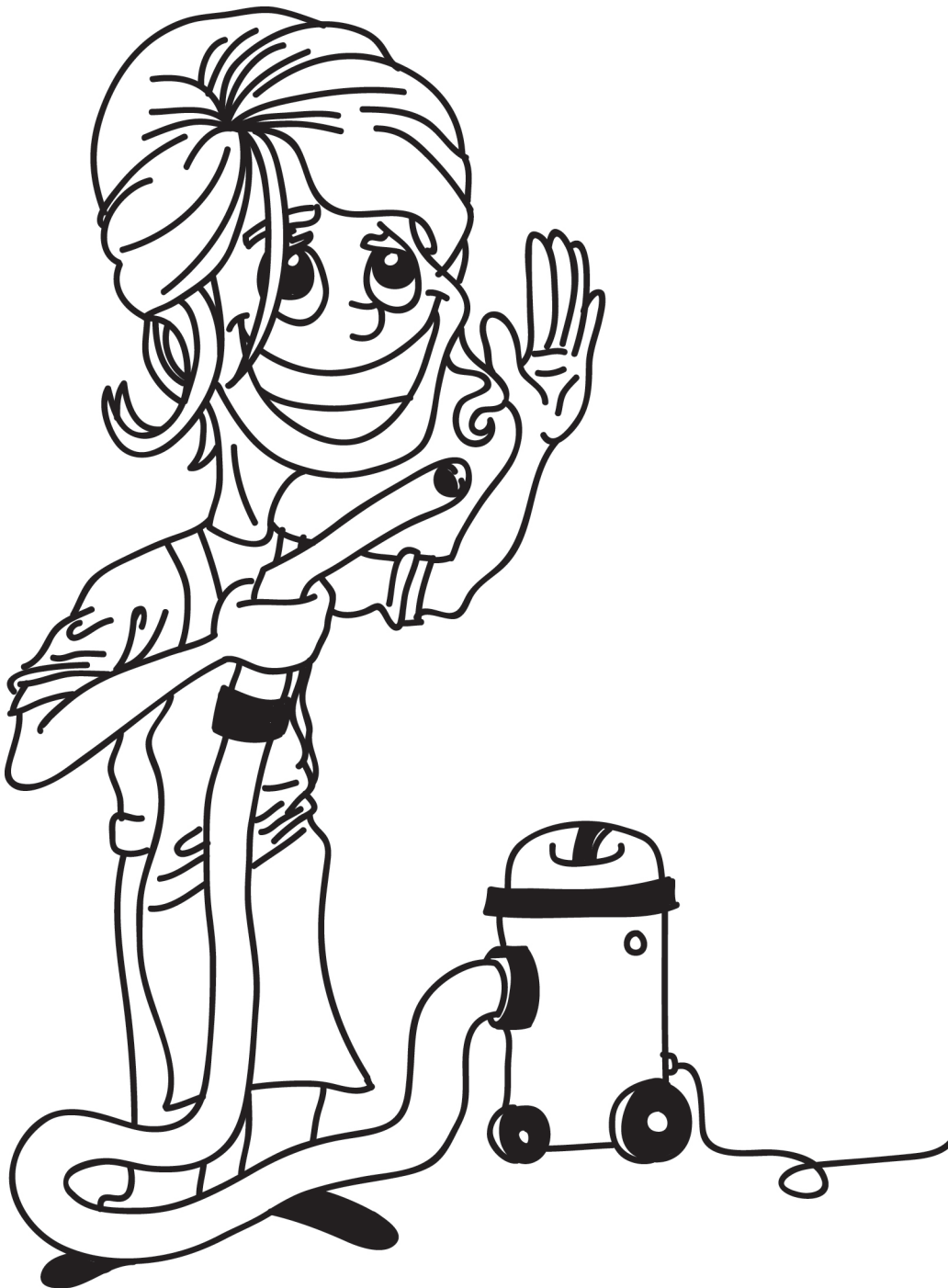
**Research Question 6:** Do other-ratings of vocational interests show self-other agreement, assumed similarity, similarity, and reciprocity?

### **3. Summary: Improving personality and interest measurement**

This dissertation studied the improvement of personality and vocational interest measures. First, the improvement of personality measures' predictive validity via contextualization was studied. We proposed and used a framework to design completely contextualized personality measures. Next, we used these newly designed measures to find out which contextualized personality measure (tagged or complete) yields the largest predictive validity and most positive participant reactions. Additionally, we investigated if the differential validity of personality measures can be reduced by the use of contextualized measures.

For the measurement of interests, we translated the PGI (Tracey, 2002) to Dutch and tested the translation's structural properties. Next, we used this vocational interest measure to investigate how profile elevation of interests and Prestige interests are related to personality measures. Lastly, the abbreviated PGI was used to investigate if other-ratings of vocational interests may be used as a measure of vocational interests.

**A matter of context: A comparison of two types of contextualized personality measures**



2

## Abstract

# A matter of context: A comparison of two types of contextualized personality measures

The present study compared the effect of two types of Frame-of-Reference modifications to each other and to a baseline generic measure. Generic personality scales, tagged scales with 'at school', and completely modified scales were compared in their prediction of academic performance, counterproductive academic behavior, and participant reactions. To this end the HEXACO-PI-R ( $n = 215$ ) and the MPT-BS ( $n = 316$ ) were filled out by students in a within-subject design. Results showed a significant increase in criterion validity from generic, to tagged, to completely contextualized personality scales. Face validity and perceived predictive validity improved with increasing contextualization. The current study indicates that completely contextualizing personality items increases criterion validity more than just adding a tag to items.

---

This chapter has been published as Holtrop, D., Born, M. Ph., De Vries, R. E., & De Vries, A. (2014). A matter of context: A comparison of two types of contextualized personality measures. *Personality and Individual Differences*, 68, 234-240. DOI:10.1016/j.paid.2014.04.029

Grateful acknowledgement is provided to Corine Sonke, and Remko van der Berg. Support was provided by Rotterdam University of Applied Sciences.

## 1. Introduction

Personality influences the way people perform in academic settings. To further improve the criterion validity of personality questionnaires, researchers have recently started to investigate contextualized questionnaires (e.g., Bing, Whanger, Davison, & VanHook, 2004; Lievens, De Corte, & Schollaert, 2008; Hunthausen, Truxillo, Bauer, & Hammer, 2003; Robie, Schmit, Ryan, & Zickar, 2000; Schmit, Ryan, Stierwalt, & Powell, 1995; Shaffer & Postlethwaite, 2012). Contextualization occurs when a relevant context is added to a personality questionnaire. There are three commonly applied methods 1) Instructional contextualization 2) Tagged contextualization and 3) Complete contextualization. Instructional contextualization asks a participant to think of a certain situation (e.g., school) when filling out questionnaire. Tagged contextualization modifies generic personality statements with an added tag, for example, by modifying statements such as 'I am a busy person' into 'I am a busy person *at school*'. Complete contextualization occurs when an item is completely redesigned to match a context. For example by changing 'People think I show a lot of effort' into 'People think I study hard'.

Regardless of the applied method, the added context is often called a Frame-of-Reference (FoR; Mount, Barrick, & Strauss, 1994; Schmit et al., 1995). The idea behind adding a FoR is that personality is more consistent within one meaningful situation than when it is aggregated across several situations. Measuring personality within a relevant situation is hypothesized to improve criterion validity of personality questionnaires for criteria relevant to that situation (e.g., personality at school predicts school performance), which is called the FoR-effect (Schmit et al., 1995). Shaffer and Postlethwaite (2012) investigated the FoR-effect in a recent meta-analysis and concluded that the criterion validity of contextualized measures is higher than that of generic measures. Lievens et al. concluded that the FoR-effect was mainly due to a reduction of within-person variability. We considered tagged scales likely to reduce within-person variability more than instructional contextualization, because a participant is constantly reminded which FoR to use. Also, tagged contextualization is the most common method to apply a FoR. Therefore, we believe that tagging, better than instructional contextualization, represents the current knowledge about the FoR-effect.

Several studies applied tagging to add a FoR to their personality questionnaire items (Bing et al. 2004; Lievens et al., 2008; Robie et al., 2000; Schmit et al., 1995). However, Lievens et al. (2008) suggested adding more content to each item, rather than solely adding a tag. So far, to our knowledge,

only three studies have investigated complete contextualization (Butter & Born, 2012; Murtha, Kanfer, & Ackerman, 1996; Pace & Brannick, 2010). Butter and Born compared a tagged Conscientiousness scale and the narrow trait Achievement, to an ecological scale 'Time management in a research context'. They found that the narrow trait and the ecological scale both predicted more variance than the general Conscientiousness scale in research progress, meeting deadlines, and the estimated probability to finish the PhD in time. Murtha et al. (1996) changed Conscientiousness and Agreeableness items to reflect several situations (e.g., work, school), and found that complete contextualization improves the reliability of personality inventory scales. Pace and Brannick (2010) changed generic Openness to Experience scales to completely contextualized work scales and found that the latter scales predicted supervisory rated creative work performance better than did generic scales. However, to our knowledge, no research has directly compared the predictive validity of completely contextualized scales with tagged scales for actual performance. In this study, we will compare the FoR-effect of two types of contextualization, tagged and complete contextualization, on study outcomes.

So far, most FoR research has focused on performance prediction. However, it is likely that it applies to (other) behaviors as well. Therefore the present study includes counterproductive academic behavior (CAB) as well as Grade Point Average (GPA) as criteria. CAB consists of behaviors such as cheating, plagiarism, and tardiness (Marcus, Lee, & Ashton, 2007). Both Marcus et al. (2007) and De Vries, De Vries, and Born (2011) found a negative relation between CAB on the one hand and generic Conscientiousness and Honesty-Humility on the other.

In our study, a within-person comparison between three types of personality scales was made. The personality dimensions Conscientiousness (Chamorro-Premuzic & Furnham, 2003; Poropat, 2009), Integrity (De Vries et al., 2011; Van Iddekinge, Taylor, & Eidson, 2005), and Emotional stability (Chamorro-Premuzic & Furnham, 2003) were selected for the present study as they are the most predictive personality dimensions for academic performance and for other study-related behaviors. The present study uses two different personality inventories, the HEXACO-PI-R (Lee & Ashton, 2004) and the MPT-BS (NOA, 2009), effectively repeating the experiment under the same conditions, therefore strengthening the findings.

Based on the above we hypothesize three differences in criterion validity between the personality inventories. We expect that more contextualized



inventories outperform other inventories in terms of criterion validity. We used the two previously mentioned criteria to estimate the criterion validity of three types of personality inventories: Study performance and CAB. The first two hypotheses replicate previous research. First, we expect that tagged inventories outperform generic inventories (H1). Second, we expect completely contextualized inventories to outperform generic inventories (H2). Lastly, we expect completely contextualized inventories to outperform tagged inventories (H3).

Adding a FoR may also improve participant reactions, because relatedness of a test to a task, or situation, positively influences perception of overall fairness of an assessment process (Gilliland, 1993). To our knowledge, only Holtz, Ployhart, and Dominguez (2005) studied the effect of contextualization on participant reactions. They included perceived job-relatedness, process-fairness, organizational attractiveness, and recommending the organization to others. Contrary to their expectations, no effect of contextualization on these participant reactions was found. Because the current study is conducted in an academic setting we cannot use the same participant reactions as Holtz et al., and therefore somewhat different participant reactions were selected. Three participant reactions were measured in this study: (1) *liking* of the test (Wiechman & Ryan, 2003); (2) *face validity*, the extent to which a participant perceives the test relevant for their tasks (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993); and (3) *perceived predictive validity*, the extent to which a participant perceives the test predictive for their performance (Smither et al., 1993). We expect more positive participant reactions for tagged personality inventories, and even more positive participant reactions for completely contextualized inventories (H4).

## 2. Method

### 2.1 Procedure and Participants

Our design included scales from two different personality inventories, the HEXACO-PI-R (Lee & Ashton, 2004) and the MPT-BS (NOA, 2009). Two FoR versions were created for both of these questionnaires, namely a tagged version and a completely contextualized version. Participants were asked to complete two survey sessions. During the first session they filled out the generic questionnaire and one (randomly determined) FoR version of this questionnaire. One week later, the remaining FoR questionnaire was filled out in the second session, to counteract possible carry-over of the FoR. This second session also included the CAB inventory. Immediately after filling out

each personality questionnaire, participant reactions related to that questionnaire were measured. Analyses showed that there was only one, out of six, small significant mean difference between session one and session two, for the randomized FoR questionnaires.

All participants ( $N = 531$ ) were second and third year students at an institute for higher vocational education in The Netherlands. Approximately 7000 students ( $\approx 50\%$  women) were approached by email for voluntary participation. A total of 695 students completed the first session. 531 of these students completed both sessions (23.60% attrition). Of these 531 students, 316 students completed both MPT-BS sessions ( $M(\text{age}) = 22.58$ ,  $SD = 5.22$ , 68.4% women) and 215 completed both HEXACO-PI-R sessions ( $M(\text{age}) = 23.90$ ,  $SD = 6.99$ , 59.2% women). Participants came from a variety of educational programs (e.g., 17% social, 13% teacher, 11% management, 9% construction).

## 2.2 Predictor Measures

### 2.2.1 HEXACO-PI-R

The Dutch HEXACO-PI-R (De Vries, Ashton & Lee, 2009; Lee & Ashton, 2004) consists of 200 statements measuring six personality dimensions: Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to experience. Only the dimensions Conscientiousness, Emotionality and Honesty-Humility were measured. Alpha reliabilities in present study were .89 for Honesty-Humility, Emotionality, and Conscientiousness.

### 2.2.2 Multicultural Personality Test – Big Six (MPT-BS)

The MPT-BS (NOA, 2009; De Vries et al., 2011) is a personality inventory that consists of 200 short statements, measuring six personality dimensions: Emotional stability, Conscientiousness, Extraversion, Agreeableness, Openness, and Integrity. The factor-level structure of the MPT-BS is based on the HEXACO model, but contains different subscales and operationalizes these scales independent from the HEXACO (NOA, 2009). Participants respond on a 5-point Likert scale, ranging from 'disagree strongly' to 'agree strongly'. An example item is: 'I dislike rules'. Alpha reliabilities in the present study were .81 for Integrity, .92 for Emotional stability, and .90 for Conscientiousness.

### 2.2.3 Tagged contextualization

Behind all personality inventory items, an 'at school' tag was added. If the tag grammatically did not fit after the last word of the item, it was placed elsewhere. Some items were dropped because they did not make sense with a tag, for example: '*I would like to live in a very expensive, high-class neighborhood*'. The HEXACO-PI-R Conscientiousness and Emotionality scales lost two items each, and Honesty-Humility lost five items. The fact that only HEXACO-PI-R items, and no MPT-BS items were dropped was because the HEXACO-PI-R items generally have more context than the MPT-BS items.

### 2.2.4 Complete contextualization

For complete contextualization, every item was completely revised. The steps are presented in Table 1. As a last step in the design process two experts on personality inventories performed a 'back-translation' to the facet level. Krippendorff's Alpha (10,000 bootstrap samples; Hayes & Krippendorff, 2007) was then calculated, showing acceptable inter-rater reliability: .85 for the HEXACO-PI-R ( $LB = .79$ ,  $UB = .90$ ) and .80 for the MPT-BS ( $LB = .74$ ,  $UB = .86$ ). Krippendorff's Alpha is a robust reliability coefficient that ranges from .00 'unreliable' to 1.00 'completely reliable'. It is unaffected by the amount of raters, or missing data, and functions independent of the metric used.

**Table 1.**  
*Complete contextualization process.*

Step	Activity
1	Generating examples: Seven people (four professors and three students) with experience in the relevant context thought of one contextualized example per generic item. This example explains how the item could be relevant in a school situation. Resulting in seven practical examples per generic items.
2	Developing a preliminary list of items: Test design team, consisting of three expert test developers, used the list of examples from step 1 to design one completely contextualized item per generic item.
3	Back-translation: One expert on personality inventories assigned the completely contextualized items to the facet scales used in the inventory.
4	Revision: All items assigned to the wrong facet scale were adjusted by the test design team (5 HEXACO-PI-R, 12 MPT-BS).
5	Final check: Two experts on personality inventories assigned the completely contextualized items to the facet scales used in the inventory. Reliability of these 'back-translations' was calculated using Krippendorff's Alpha.

## 2.3 Criterion Measures

### 2.3.1 GPA

Two months after this study, the average grade over the entire school career at the institution was obtained from the institution's database. The GPA criterion is therefore an objective measure of students' academic performance. GPA scores ranged from 1 to 10, with higher scores indicating better performance.

### 2.3.2 Counterproductive Academic Behavior

To measure CAB, a 25 item Dutch scale was used (De Vries et al, 2011; Marcus et al., 2007;  $\alpha = .84$ ), using a 6-point Likert scale, ranging from: '*Never even considered it*' to '*Did it three or more times*'. Participants were asked to think of the last five school years when filling out the CAB inventory. An example item is: '*Submitted a class paper or project that was not your own work*'.

### 2.3.3 Participant reactions

Three questions were designed to measure participant reactions, based on items from Smither et al. (1993) for perceived predictive validity and face validity, and based on items from Wiechman and Ryan (2003) for liking. The items were: '*The content of this questionnaire is clearly related to my study*' (face validity); '*With the results of this questionnaire my study performance can be predicted*' (perceived predictive validity), and '*I did not enjoy completing this questionnaire*' (liking; reverse coded). Participants reacted on a 7-point Likert scale, ranging from '*completely disagree*' to '*completely agree*'.

A supporting study was conducted to estimate single-item reliability (Wanous & Reichers, 1996) of the participant reaction items. 269 students, who also participated in the main study, filled out an abbreviated personality inventory and three scales designed to measure perceived predictive validity, face validity (both 5 items; Smither et al., 1993), and liking (4 items; Wiechman & Ryan, 2003). The single-item reliability was estimated following a procedure described by Wanous and Reichers, and was .64 for liking, .60 for perceived predictive validity, and .55 for face validity.

## 2.4 Data analyses

To investigate our hypotheses, several three-step hierarchical regression analyses were performed according to the method used by Pace and Brannick (2010). When inventory A significantly increases explained variance over inventory B, but not vice versa, then inventory A explains more

variance than B. Results of these analyses are presented in Table 3 for GPA and Table 4 for CAB. We always included one inventory version (generic/tagged/completely contextualized) in the second step, and one other version in the third step. Every column in Tables 3 and 4 shows two three-step hierarchical regression analyses; please note that the two values in the third step represent different analyses.

### 3. Results

The personality measures used in this study all showed adequate reliability, including the newly designed and modified personality scales (see Table 2 for descriptive statistics). Comparable reliabilities for the different inventory types indicated that the modification process did not affect the scales' quality. Correlations between the same personality scale dimensions in different FoR-versions ranged from  $r = .74$  to  $r = .89$ , indicating that the FoR modifications resulted in modest changes to the constructs investigated. A principal component factor-analysis with varimax-rotation was performed for each personality questionnaire, showing that the factor structure of the three scales remained the same.

Only scales for which the generic scale related significantly to the criteria were selected for the hierarchical regression analyses, namely Conscientiousness for GPA and CAB, and Honesty-Humility/Integrity for CAB. All but two correlations between Emotionality/Emotional stability and the criteria were non-significant. Consequently Emotionality/Emotional stability was left out of the hierarchical regression analyses.

First, the results for GPA showed a clear pattern. The hierarchical regression analysis with MPT-BS Conscientiousness showed that the completely contextualized questionnaire explained most variance in GPA, the tagged version significantly less and the generic version least. Differences in explained variance were small. These findings supported respectively H2 and H3. For HEXACO-PI-R Conscientiousness, tagged and completely contextualized questionnaires both explained more variance in GPA than generic, supporting H1 and H2. However, for HEXACO-PI-R Conscientiousness neither tagged scale nor completely contextualized scale explained more variance in GPA over the other, thus not supporting H3. Second, results for CAB showed a similar pattern. HEXACO-PI-R Conscientiousness and MPT-BS Integrity showed the hypothesized stepwise increasing pattern when predicting CAB, supporting H1 through H3. The

**Table 2.***Descriptive statistics for HEXACO-PI-R, MPT-BS and criterion variables.*

	<i>M(SD)</i>		1	2	3	4	5	6	7	8	9	10	11
<b>Predictor</b>													
1. (Generic) Emotional stability	3.10(0.49)	3.57(0.44)	.89/.92	.07	.10	.83**	.19**	.12	.81**	.15*	.13	.11	-.03
2. (Generic) Conscientiousness	3.45(0.49)	3.64(0.41)	.16**	.89/.90	.09	.05	.87**	.10	.14*	.88**	.16*	.26**	-.34**
3. (Generic) Integrity/Honesty-Humility	3.72(0.46)	3.45(0.37)	.34**	.25**	.89/.81	.00	.11	.83**	.00	.12	.84**	.01	-.33**
4. (Tagged) Emotional stability	2.84(0.47)	3.71(0.43)	.81**	.26**	.32**	.88/.93	.14*	.01	.86**	.12	-.01	.10	.05
5. (Tagged) Conscientiousness	3.50(0.49)	3.63(0.41)	.13*	.84**	.29**	.32**	.90/.91	.15*	.21**	.89**	.19**	.31**	-.36**
6. (Tagged) Integrity/Honesty-Humility	3.91(0.51)	3.64(0.36)	.36**	.29**	.82**	.42**	.35**	.90/.81	.03	.12	.84**	-.02	-.34**
7. (Compl.) Emotional stability	2.82(0.45)	3.63(0.40)	.78**	.20**	.29**	.85**	.20**	.35**	.87/.91	.19**	.04	.08	-.01
8. (Compl.) Conscientiousness	3.42(0.51)	3.54(0.40)	.07	.81**	.22**	.22**	.87**	.27**	.19**	.91/.89	.20**	.31**	-.39**
9. (Compl.) Integrity/Honesty-Humility	3.85(0.50)	3.57(0.35)	.34**	.24**	.74**	.39**	.32**	.77**	.40**	.27**	.91/.76	-.02	-.39**
<b>Criterion</b>													
10. GPA	6.86(0.57)		-.02	.41**	.10	.12*	.46**	.11*	.06	.48**	.10	-	-.25**
11. CAB	2.43(0.69)		-.05	-.38**	-.32**	-.13*	-.39**	-.36**	-.08	-.39**	-.42**	-.25**	.85

*Note.*  $n(\text{HEXACO}) = 215$ ,  $n(\text{MPT-BS}) = 316$ . For *M*,  $\alpha$ , and *SD* the first value relates to the HEXACO, the second value to the MPT-BS.  $\alpha$  is shown on the diagonal. Correlations above the diagonal relate to the HEXACO-PI-R, below the diagonal to the MPT-BS.

\*  $p < .05$ . \*\*  $p < .01$ .

**Table 3.***Hierarchical regression analyses results for Conscientiousness with GPA.*

		Conscientiousness $R^2(\Delta R^2)$			
		MPT-BS		HEXACO	
Step 1	Age, gender	.04		.09	
Step 2	Generic	.19(.15**)		.13(.05**)	
	Tagged	.22(.18**)		.16(.07**)	
	Complete	.24(.21**)		.16(.07**)	
Step 3	Generic	.22(.00)	.25(.00)	.16(.00)	.16(.00)
	Tagged	.22(.03**)	.25(.00)	.16(.02*)	.16(.00)
	Complete	.25(.06**)	.25(.03**)	.16(.02*)	.16(.00)

\*  $p < .05$ . \*\*  $p < .01$ 

HEXACO Honesty-Humility scale supported H2 and H3. However, it did not show an increase in criterion validity when the generic scale was compared to the tagged scale, not supporting H1. MPT-BS Conscientiousness also did not fit the predicted pattern, which was due to the generic version outperforming the tagged version and performing equally well, compared to complete contextualization.

In sum, the majority of the comparisons showed the hypothesized pattern for predicting both GPA and CAB. H1 was least often supported, in two out of six cases the tagged scale did not have a significantly higher criterion validity than the generic scale. In all but one case completely contextualized scales performed better than generic scales, supporting H2. Also, in all but one case completely contextualized scales outperformed tagged scales, supporting H3.

Last of all, we analyzed if contextualization improved participant reactions. Descriptive statistics for participant reactions are shown in table 5. Participants reacted generally positively to the questionnaires (lowest:  $M = 3.73$  on a 7-point scale). Several paired t-tests were performed. Descriptives,  $t$ -values and effect sizes of these analyses are presented in Table 5 and visually in Figures 1 and 2. First, face validity increased significantly when a tag was added and increased even more when scales were completely contextualized. Second, perceived predictive validity only increased for completely contextualized, but not for tagged questionnaires. Interestingly, liking was significantly lower for the tagged version and the completely

**Table 4.***Hierarchical regression analyses results for Conscientiousness and Integrity/Honesty-Humility with CAB.*

		Conscientiousness $R^2(\Delta R^2)$				
		MPT-BS		HEXACO		
Step 1	Age, gender	.09			.11	
Step 2	Generic	.22(.13**)			.21(.10**)	
	Tagged		.21(.12**)			.23(.12**)
	Complete			.23(.13**)		.24(.13**)
Step 3	Generic		.23(.02**)	.24(.01*)		.23(.00)
	Tagged	.23(.01)		.23(.00)	.23(.02*)	.24(.00)
	Complete	.24(.02*)	.23(.02**)		.24(.04**)	.24(.01*)
		Integrity/Honesty-Humility $R^2(\Delta R^2)$				
		MPT-BS		HEXACO		
Step 1	Age, gender	.09			.11	
Step 2	Generic	.17(.07**)			.18(.07**)	
	Tagged		.18(.09**)			.18(.07**)
	Complete			.22(.13**)		.21(.10**)
Step 3	Generic		.18(.00)	.22(.00)		.19(.01)
	Tagged	.18(.02*)		.22(.00)	.19(.01)	.21(.00)
	Complete	.22(.05**)	.22(.04**)		.21(.03**)	.21(.03**)

\*  $p < .05$ . \*\*  $p < .01$



**Table 5.***Descriptive statistics and paired t-tests for Participant reactions.*

		Generic				Tagged				Complete			
		<i>M</i>	<i>SD</i>	<i>t vs. Tagged</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>t vs. Complete</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>t vs. Generic</i>	<i>d</i>
Face validity	MPT-BS	4.08	1.58	-5.09**	-.31	4.57	1.63	-5.83**	-.35	5.1	1.37	10.66**	.69
	HEXACO	3.73	1.61	-4.98**	-.40	4.38	1.65	-6.20**	-.45	5.1	1.56	11.22**	.86
Perceived predictive validity	MPT-BS	4.03	1.61	-0.34	-.02	4.06	1.57	-3.73**	-.20	4.37	1.5	4.66**	.22
	HEXACO	3.87	1.54	-1.30	-.08	3.99	1.53	-2.94**	-.18	4.27	1.59	4.19**	.26
Liking	MPT-BS	5.28	1.52	5.89**	.28	4.84	1.66	-2.67**	-.14	5.07	1.55	-3.02**	-.14
	HEXACO	5.26	1.63	5.07**	.28	4.77	1.83	-1.81	-.11	4.97	1.67	-3.20**	-.18

\*  $p < .05$ . \*\*  $p < .01$

2

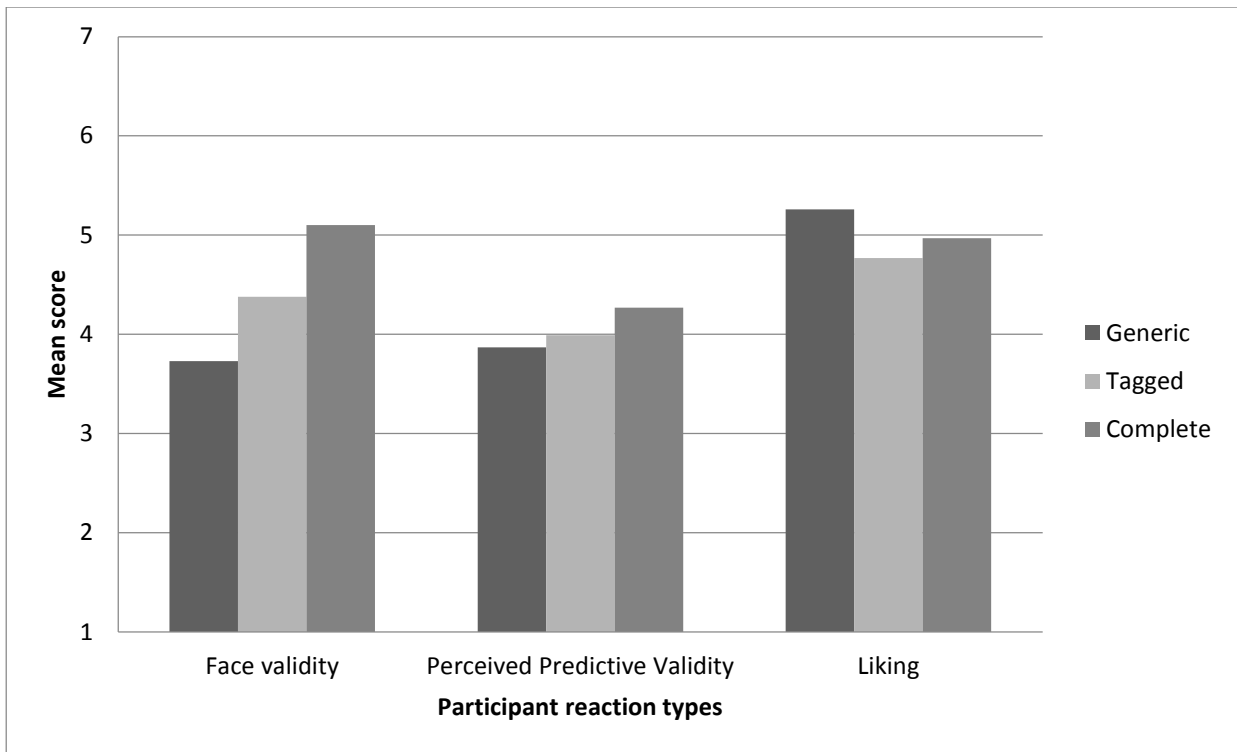


Figure 1. Participant reactions per FoR-type for the HEXACO-PI-R.

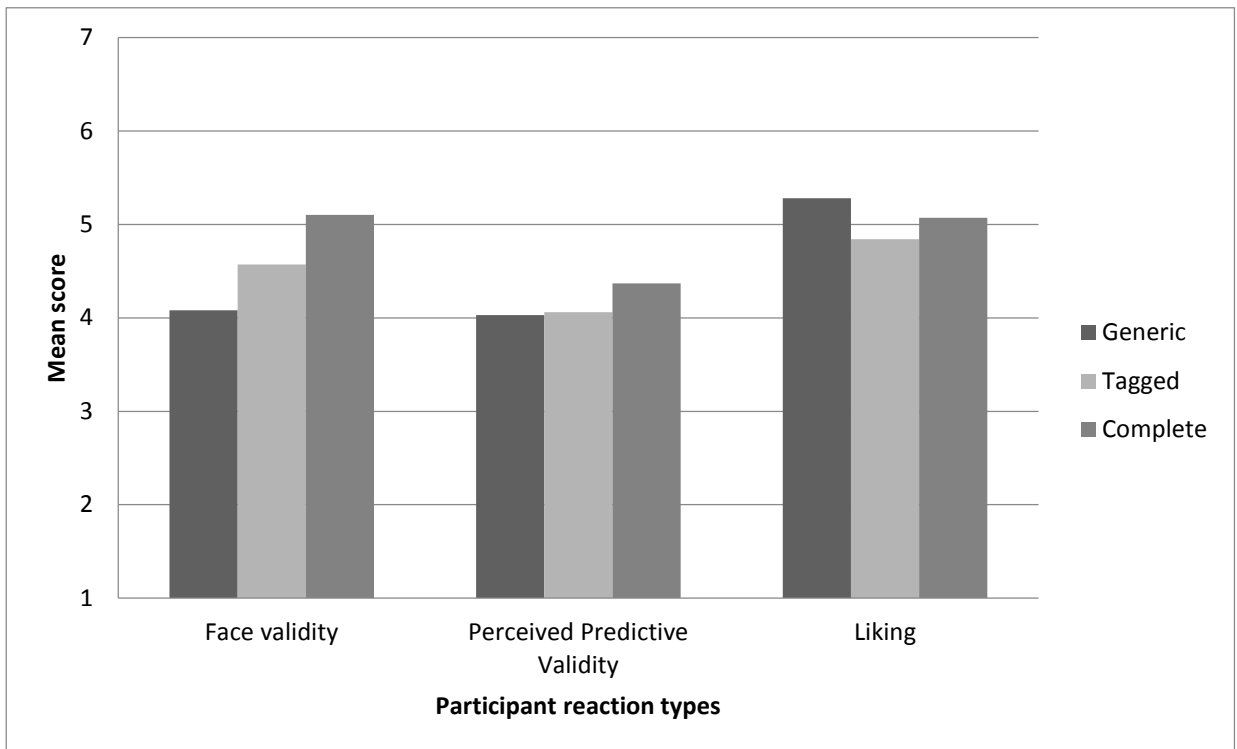


Figure 2. Participant reactions per FoR-type for the MPT-BS.

contextualized version when compared to the generic version. Participants liked the tagged questionnaire least, then the completely contextualized questionnaire, and they liked the generic questionnaire most. In sum, H4 was supported for face validity and perceived predictive validity, but not for liking.

#### 4. Discussion and conclusions

The first and most important conclusion of the present study is that different methods of contextualization yield different results; complete contextualization seems to have a larger FoR-effect than tagging. Tagged items and completely contextualized items both explained more variance in criteria than generic items. The finding that completely contextualized items outperform generic items in criterion validity is in line with Pace and Brannick's study (2010). The present study adds to their findings by directly comparing tagged and complete contextualization. In only one out of six comparisons, tagged items and completely contextualized items performed equally well. In sum, given conceptually related predictors and criteria, both tagged items and completely contextualized items have a higher criterion validity than generic items and completely contextualized items outperform tagged items.

The second major finding of this study is that participant reactions showed significant differences between different types of contextualization. According to Gilliland (1993), participants like contextualized inventories better than generic inventories because contextualized inventories relate more strongly to the task or situation than generic measures. In our study, both perceived predictive and face validity increased when a FoR was added. However, results also showed that participants liked the generic version best and the tagged version least. A possible reason is that tagged items (e.g., '*My most important purpose is to have a lot of money at school*') can sound artificial to participants. The tagged items were also somewhat tedious, because 'at school' was repeated in every item. A possible reason for the difference in liking between the generic and completely contextualized version is that participants may find that completely contextualized items restrict their response options to only one situation.

An important strength of the current study is that, compared to other studies, GPA was collected from objective data from the institution's database, and not substituted with estimated GPA. Therefore, the current study is not subject to a common source bias for the performance measure. Some limitations of the study should also be noted. First of all, participant reactions were measured with single items, with single-item reliability estimates below

the – generally accepted – .70 baseline. However, although these results should be interpreted with caution, the participant reaction findings are likely to underestimate the real effects. Additionally, our study was performed in a low stakes research situation. Our findings may not be directly applicable to selection situations. Lastly, the set-up of the current study may have invited order-effects. Participants always filled out the generic inventory first and then the two FoR-inventories, which were administered in a randomly determined order. We chose to do this to prevent carryover of the school-FoR to the generic scales.

Future research could focus on the behaviors to which the FoR-effect extends. A next step could be to combine personality traits with a taxonomy of situations (e.g. Ten Berge & De Raad, 1999) and try to replicate the FoR-effect with an inventory that measures personality traits in situations described in a taxonomy. However, situations in a taxonomy might nullify the FoR effect, because these situation may be too broad to increase predictive validity. Therefore future research might consider narrow versus broad situations and the effect of those on contextualization and the FoR-effect.

There seem to be meaningful differences between different types of contextualization that have not been included in the current lines of research. These results inform researchers to not generalize findings too readily with one type of contextualized inventories, across other types of contextualization. Practical consequences of our findings are not one-directional. Compared to completely contextualized items, tagged items are easier to construct and also increase criterion validity compared to generic questionnaires. Designing a completely contextualized questionnaire took roughly 65 hours, whereas designing a tagged personality questionnaire took roughly 3 hours. Well-designed completely contextualized items involve a lengthy process with many stages. At first glance, the relatively small added effect size of completely contextualized scales over tagged scales may not be encouraging enough to engage in this process. However, participants seem to consider completely contextualized personality inventories to be more relevant and more predictive for behaviors, in the added context, than tagged and generic versions, which may lower the chance of objections against predicting performance with a personality inventory. In sum, practitioners should design completely contextualized inventories to predict performance for larger groups in order to make the investment worthwhile; Researchers should not focus on tagged inventories to investigate the full FoR-effect.

**Predicting performance with  
contextualized inventories, no Frame-of-  
Reference effect?**



3

# **Abstract**

## **Predicting performance with contextualized inventories, no Frame-of-Reference effect?**

A recent meta-analysis (Schaffer & Postlethwaite, 2012) showed that contextualized personality inventories have incremental predictive validity over generic personality inventories when predicting job performance. This study aimed to investigate the differences between two types of contextualization of items: Adding an 'at work' tag versus completely modifying items. 139 pharmacy assistants from 29 pharmacies filled out a generic, a tagged and a completely modified personality inventory. The assistants also provided participant reactions for each of the personality inventories. Performance ratings were collected from the supervising pharmacists. We expected to find incremental criterion validity for both the tagged inventory and the completely modified inventory for predicting job performance. However, the results showed an unexpected decrease in predictive validity for the contextualized inventories. Contextualized inventories were liked less than the generic inventory, but evaluated somewhat more face valid and predictive by the participants.

---

This chapter has been published as Holtrop, D., Born, M. Ph., & De Vries, R. E. (2014). Predicting performance with contextualized inventories, no Frame-of-Reference effect? *International Journal of Selection and Assessment*, 22, 219-223. DOI:10.1111/ijsa.12071.

Grateful acknowledgement is provided to Clara Meijer, Ruud Spoor, and Thijs Slotman for their assistance in making the present study possible. Support for this research was provided by 'the Big Five from the Hoeksche Waard', 'Thio Pharma' and 'Pharmalead'.

## 1. Purpose

The current study investigates the Frame-of-Reference (FoR; Mount, Barrick, & Strauss, 1994) effect in a job context. The FoR effect implies that measuring personality within a relevant situation improves the criterion validity of personality questionnaires for criteria relevant to that situation (e.g., adding 'at work' to personality items predicts job performance better than generic personality items). A recent meta-analysis by Shaffer and Postlethwaite (2012) showed that the criterion validity of FoR personality measures is higher than that of generic measures.

The purpose of this study is to investigate what type of FoR modification is preferable: 1) tagged contextualization, which is easy to apply and not very time-intensive, such as adding 'at work' to items, or 2) full contextualization, rewriting every item to completely apply to the specific job. This research compares both types of FoR and a generic measure in terms of their predictive validity for job performance and job satisfaction, as well as on participant reactions. We hypothesized that the tagged version would perform better than the generic version, and that the fully contextualized version would perform even better than the tagged version.

## 2. Method

### 2.1 Sample and procedure

40 pharmacies were approached to participate in this study. 33 pharmacies agreed to participate. Roughly 200 pharmacy assistants, all female, were approached by their employers to participate in 'a study investigating personality at work'. The participants were informed they would be awarded a € 15,- voucher for a beauty store if they completed two online questionnaire sessions. They were also informed that all the information gathered during this study would be treated confidentially and that their employer would not receive the results.

156 pharmacy assistants (78%) completed both sessions. Subsequently, supervisory performance ratings could be collected for 139 participants. The ratings were filled out by the pharmacists in charge of the pharmacy assistants. Analyses were performed on 139 female pharmacy assistants ( $M(\text{age}) = 40.09$ ,  $SD = 11.35$ ). The participants came from 33 different pharmacies, ranging 1 to 12 assistants per pharmacy ( $M = 5$ ), with one pharmacist per pharmacy. Most participants were ethnic Dutch ( $n = 122$ ), and 17 participants were from various other ethnic groups.

All participants completed two online questionnaire sessions. The first session included a general personality inventory (the HEXACO, see below) and a randomly determined contextualized version, either tagged or fully contextualized. The second session included the remaining contextualized personality inventory and a work satisfaction measure, the AJIG (Abridged Job in General Scale; Bowling state, 2009). Most participants completed the inventories at home (session 1: 96.4%; session 2: 95%). After the participants completed the second session their supervisors were approached to fill out the performance measure. Data collection took place in 2011 and 2012 in the Netherlands.

## 2.2 Measures

### 2.2.1 Personality measures

The Dutch HEXACO-PI-R (De Vries, Ashton & Lee, 2009) was used to measure personality. For this study the three most predictive main dimensions of personality for work performance were selected (Honesty-Humility, Emotionality, and Conscientiousness) from the six main dimensions that are included in the HECACO model.

The tagged personality measure version was created by adding the tag 'at work' behind every HEXACO item. Six items could not sensibly be tagged and were therefore left out (e.g., I would like to live in a very expensive, high-class neighborhood). These items were also left out for the generic and fully contextualized version to enhance the accuracy of the comparison between the scales.

The fully contextualized version of the personality measure was created in several steps. First, a pharmacist was interviewed in depth about the job of a pharmacy assistant. Second, the authors designed new items based on the original HEXACO items. These items were then evaluated by a pharmacist and pharmacy assistant on their relevance for the job of a pharmacy assistant. Third, a personality inventory expert assigned the fully contextualized items to the facets (scales underlying the main dimensions) used in the inventory. All items assigned to the wrong facet (10.42 % of all items) were adjusted. As a final check, three undergraduate students assigned the fully contextualized items to the facets used in the inventory. The reliability of these 'back translations' was calculated using Krippendorf's  $\alpha$  (.73).

After the data collection the scores of the participants of the current study were compared to the female Dutch HEXACO norm group ( $n = 680$ ; De Vries et al., 2009) on the three generic main dimensions. The participants



scored significantly higher on the Conscientiousness and Honesty-Humility scales (Conscientiousness:  $M\Delta = 0.16$ ,  $t = 4.33$ ,  $p < .01$ ; Honesty-Humility:  $M\Delta = 0.17$ ,  $t = 3.97$ ,  $p < .01$ ).<sup>1</sup>

### 2.2.2 Participant reactions

Participant reactions were completed by 101 participants for each of the three personality inventory types, i.e. generic, tagged, and fully contextualized. Specifically, face validity (Smither et al., 1993; 5 items;  $\alpha$ (current study) = .81/.53/.88), perceived predictive validity (Smither et al., 1993; 5 items;  $\alpha$ (current study) = .87/.76/.92) and liking (Wiechman & Ryan, 2003; 4 items;  $\alpha$ (current study) = .71/.52/.76) were measured. Participants responded to all items on a 7-point Likert scale ranging from *completely disagree* to *completely agree*.

### 2.2.3 Job satisfaction

To measure work satisfaction participants filled out the Abridged Job In General scale (AJIG; 8 items; Bowling Green State University, 2009). The AJIG measures job satisfaction in general, participants indicate if keywords apply to their job (e.g., Enjoyable). The inventory uses statements that can be answered with 'yes', 'no', or '?' ( $\alpha = .82$ ).

### 2.2.4 Job performance.

A 20 item job performance inventory was distributed to the supervisors ( $n = 33$ ) of the participating pharmacy assistants. The inventory was based on a job performance form distributed by a large Dutch national pharmacy branch organization (Stichting Bedrijfsfonds Apotheken, SBA) and consists of short statements about aspects of the job. All items were answered on a 5-point Likert scale ranging from *completely disagree* to *completely agree*. Mean performance ratings were relatively high ( $M = 4.14$ ,  $SD = 0.57$ ,  $\alpha = .96$ ), suggesting restriction of range. Yet, average performance ratings from the pharmacists were still significantly different and large (pharmacist) rater effects were observed ( $min = 3.31$ ,  $max = 4.95$ ;  $F(28) = 5.57$ ,  $p < .01$ ,  $ICC(1) =$

<sup>1</sup> The scale scores on the personality dimensions only differed significantly on generic and tagged Honesty-Humility between the pharmacies. The other personality scale scores did not differ between pharmacies (generic Honesty-Humility:  $F(28) = 1.64$ ,  $p < .05$ ,  $ICC(1) = 0.13$ ,  $ICC(2) = 0.39$ ; tagged Honesty-Humility:  $F(28) = 2.12$ ,  $p < .01$ ,  $ICC(1) = 0.21$ ,  $ICC(2) = 0.53$ ; all other personality scales:  $p > .10$ ). Because within-subject variance explained more differences in personality than the between-pharmacy variance we chose for individual level analyses across pharmacies while correcting for any differences between pharmacists. Additionally, the dataset did not have sufficient power for multi-level research according to the 50/20 rule or the 30/30 rule.

**Table 1.**

*Descriptive statistics and correlations of the generic, tagged, and fully contextualized HEXACO and criterion variables.*

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
<b>Predictors</b>														
1. Emotionality	3.25	0.36	.83											
2. Conscientiousness	3.57	0.35	-.15	.84										
3. Honesty-Humility	3.97	0.36	-.02	.19*	.86									
4. (Tagged) Emotionality	2.96	0.39	.82**	-.14	-.12	.86								
5. (Tagged) Conscientiousness	3.65	0.33	-.16	.81**	.27**	-.22**	.85							
6. (Tagged) Honesty-Humility	4.12	0.38	-.05	.20*	.80**	-.10	.29**	.89						
7. (Fully) Emotionality	2.87	0.32	.77**	-.22**	-.10	.79**	-.26**	-.08	.81					
8. (Fully) Conscientiousness	3.72	0.30	-.17*	.58**	.37**	-.21*	.68**	.31**	-.35**	.81				
9. (Fully) Honesty-Humility	3.93	0.29	-.03	.16	.74**	-.07	.21*	.72**	-.02	.34**	.82			
<b>Criteria</b>														
10. Performance	4.15	0.56	.14	.09	.04	.10	.04	.06	.03	.11	-.04	.96 <sup>1</sup>		
11. Corrected Performance	0.00	0.36	.06	.29**	.06	-.04	.15	-.03	-.04	.15	-.02	.64**	-	
12. AJIG	2.28	0.52	.03	-.01	.19*	.06	.02	.24**	.05	.05	.19*	.09	-.04	.82

*Note.*  $n=139$ , <sup>1</sup> $n = 132$  due to missing values,  $\alpha$  coefficient on diagonal, AJIG 0=No, 1=?, 3=Yes (Russell et al., 2004), all other scales use a 5-point Likert scale. Corrected Performance is one's individual performance rating minus the average rating given by the pharmacist ( $M_{individual} - M_{pharmacist}$ ) to correct for rater tendencies.

\*  $p < .05$ . \*\*  $p < .01$ .

**Table 2.***Participant reactions to the generic, tagged, and fully contextualized HEXACO.*

Variable	Version	Reaction	<i>M</i>	<i>SD</i>	<i>t</i> (tag)	<i>t</i> (fully)	1	2	3	4	5	6	7	8	9
1	Generic	Face validity	4.37	1.14	1.21	-9.58**	.81								
2	Generic	Liking	4.15	0.96	3.71**	0.51	.69**	.71							
3	Generic	Perceived Predictive Validity	2.94	1.26	-3.03**	-1.59	.69**	.65**	.87						
4	Tagged	Face validity	4.50	0.95	-	-8.29**	.01	.18	-.02	.53					
5	Tagged	Liking	3.88	0.85	-	-2.86**	.12	.20*	.15	.51**	.52				
6	Tagged	Perceived Predictive Validity	3.30	1.11	-	1.60	.09	.28**	.14	.73**	.45**	.76			
7	Fully	Face validity	5.43	1.16	-	-	.47**	.38**	.32**	.16	.09	.18	.88		
8	Fully	Liking	4.11	1.02	-	-	.31**	.25*	.18	.02	.10	.04	.47**	.76	
9	Fully	Perceived Predictive Validity	3.10	1.33	-	-	.40**	.42**	.40**	.07	.03	.19	.53**	.43**	.92

*note.* *N* = 101, scales ranged from 1 to 7,  $\alpha$  coefficient on diagonal. *t*(tag) and *t*(full) show paired t-tests comparing the participant reactions for the different inventory types.

\**p* < .05. \*\**p* < .01.

0.52,  $ICC(2) = 0.82$ ). Because of these relatively large differences, job performance ratings were centered for each pharmacy to correct for any differences in rater tendency. Descriptives and correlations for both ratings (centered and uncentered) are shown in Table 1.

### 3. Results

Principal component factor analysis with Varimax rotation was performed to check the personality measure's structure. The facets underlying the main personality dimensions generally loaded on the correct main dimensions (component loadings  $> .50$ ). However, the facet Fairness shifted from Honesty-Humility -the intended dimension- to Conscientiousness for both contextualized versions. Correlations between the main dimensions of the generic and contextualized versions were high ( $.58 - .82$ ). Based on previous studies (e.g. Pace & Brannick, 2010) investigating the FoR-effect this indicates that, in this sample, the FoR-effect will be likely to be small to medium at best.

Table 1 shows the means, standard deviations, reliabilities, and correlations for the independent and criterion variables. As can be derived from this table, the results yielded no support for the FoR effect hypotheses. The only personality trait that showed a relation with the corrected work performance measure was generic Conscientiousness ( $r = .29, p < .01$ ). Similar relations were found for the facets of Conscientiousness: In case a generic facet was significantly related to work performance, its contextualized facet related less strongly than, or almost as strongly as, this generic facet to performance. The other personality traits did not show any significant relations with job performance. A Relative Weight Analysis, yielding relative weights ( $\epsilon$ , indicating the relative proportion of explained variance for each variable) for the facets of Conscientiousness (combined  $\epsilon = .13$ ) showed that the generic Conscientiousness facets Diligence ( $\epsilon = .06$ ) and Perfectionism ( $\epsilon = .06$ ) accounted for most of the explained variance (85.6%) in job performance, when all four generic facets of Conscientiousness were compared. The only other two facets that significantly correlated with job performance were tagged and fully contextualized Diligence (respectively  $r = .17, p < .05, \epsilon = .06$  and  $r = .26, p < .01, \epsilon = .06$ ).

All measures of Honesty-Humility significantly correlated with the job satisfaction (AJIG) scale. The fully contextualized and generic Honesty-Humility scales correlated at  $r = .19 (p < .05)$  with job satisfaction. To check if the tagged Honesty-Humility, with a correlation of  $r = .24 (p < .01)$ , correlated

stronger to job satisfaction than the generic scale, we performed a hierarchical regression analysis with age in the first step, generic Honesty-Humility in the second step, and tagged Honesty-Humility in the third step. The hierarchical regression analyses showed that the tagged scale did not relate stronger to job satisfaction than did the generic scale ( $\Delta R^2 = .03, p = .08$ ).

Lastly, participant reactions were analyzed. Correlations and descriptive statistics are shown in Table 2. The participant reactions within the same inventory correlated stronger with each other than across inventories. Respondents perceived the tagged inventory to be more predictive for job performance than the generic inventory ( $t = 3.71, p < .01$ ) but liked it less than the generic inventory ( $t = -3.03, p < .01$ ). The fully contextualized inventory was perceived to be the most face valid inventory by the participants (vs generic:  $t = -9.58, p < .01$ ; vs tagged:  $t = -8.29, p < .01$ ) and was liked better than the tagged inventory ( $t = -2.86, p < .01$ ). In general, we found that the contextualized inventories received more positive reactions from the participants, although the tagged inventory was liked less.

#### **4. Conclusions**

The results of the current study are in contrast with most research on the FoR effect. We found no relation between contextualized personality measures and job performance, for both the main dimensions and the facets of personality. However, we did find a significant relation between generic Conscientiousness and job performance. The personality dimension means are more positive for both contextualized versions compared to the generic version. This might indicate that participants filled out the inventory in terms of 'How the task should be done' instead of 'How I do the task'. The fully contextualized measure did receive more positive reactions from the participants in general. A potential limitation of the current study is that participants were not in a high stakes situation. Our findings can therefore not readily be generalized to a selection situation, as they do not fully capture the full effects of contextualization for selection purposes. Because contextualization could make it easier to identify desirable items for participants, future studies should consider the effects of contextualization on social desirability, especially in high stakes situations.



**Differential predictive validity of  
personality instruments for academic  
performance: Person-situation interaction  
for ethnic minorities**



4

# **Abstract**

## **Differential predictive validity of personality instruments for academic performance: Person-situation interaction for ethnic minorities**

The present study investigated the differential criterion validity of personality inventories for the prediction between Dutch majority students ( $n = 190$ ) and non-western minority students ( $n = 110$ ) in higher education. We hypothesized that the contextualization of a personality inventory could reduce its differential validity and improve participant reactions. To this end the students filled out a generic personality inventory and a contextualized personality inventory in a within-subject design. As a measure of academic success, Grade point average (GPA) was self-reported and obtained from institutional records. Replicating earlier studies, results showed a significant increase in the prediction of actual GPA for the majority students from generic Conscientiousness ( $r = .29$ ) to completely contextualized Conscientiousness ( $r = .39$ ). However, actual GPA was not correlated to generic ( $r = -.02$ ) nor to completely contextualized Conscientiousness ( $r = .05$ ) for the minority students. Additionally, results showed significant correlations between self-reported GPA and personality for majority students and also (contrary to actual GPA) for minority students. Self-reported GPA correlated highly to actual GPA for the majority ( $r = .80$ ) and significantly lower for the minority ( $r = .33$ ). Participant reactions improved for the contextualized inventory, but did not differ between the majority and minority students. In conclusion, personality inventories seem to show differential predictive validity between majority and minority students, and contextualization does not seem to solve this. Moreover, the use of self-reported GPA may mask important subgroup differences in a personality inventory's predictive validity.

---

Grateful acknowledgement is provided to The Hague University of Applied Sciences, Windesheim Flevoland University of Applied Sciences, and InHolland University of Applied Sciences for their collaboration in this project, and to Hagar van der Knijff, Maura van der Molen, and Paulette Teken for their assistance in collecting data.



## 1. Introduction

Whereas most research treats personality traits as universalistic predictors of academic performance (e.g., Paunonen & Ashton, 2013; Poropat, 2009), a small amount of research has indicated that personality traits may be subject to differential criterion validity (e.g., Church, 2010; De Meijer, Born, Terlouw, & Van der Molen, 2008; De Vries, Born, & De Vries, 2012). Differential criterion validity refers to the phenomenon that commonly accepted causal relations of personality traits, such as Conscientiousness, on performance may actually vary between (cultural) groups. In an academic setting, we investigated if the differential validity of personality inventories can be reduced by including person-situation interaction to the inventory. For this purpose we compared the criterion validity of a generic personality inventory to that of a contextualized personality inventory among majority and ethnic minority students.

This investigation is particularly relevant due to the increasing diversity in European countries, on the average 6.6% of the population in the European Union is born in non-EU countries and this percentage differs widely between the EU countries (Eurostat, 2014). For example, 14% of the students starting a Dutch higher education belongs to a non-western ethnic minority (CBS, 2015a). More than half of these students—we estimate—fills in one or more psychological inventories. It is thus imperative that these inventories are equally valid for majority and ethnic minority students. Even more so because differential predictive validity may disadvantage ethnic minority students to a greater extent, a group that achieves significantly less academic success than majority students (e.g., Roth & Bobko, 2000; Kuncel, Credé, & Thomas, 2005, Severiens & Wolff, 2008).

### 1.1 Predictive validity of personality and ethnic minorities

The most predictive personality trait of academic performance is Conscientiousness (e.g., De Vries et al., 2012; Paunonen & Ashton, 2013; Poropat, 2009). Additionally, some research reported weak to moderate relations with academic success for the personality traits Integrity (e.g., De Vries, De Vries, & Born, 2011) and Emotional stability (e.g., Chamorro-Premuzic & Furnham, 2003). However, the predictive validity of psychological instruments may differ between subgroups (e.g., Church, 2010; Cook, 2004). In the case of different predictive validity an instrument may be 1) predictive for a majority group and not at all predictive for minorities, this notion is called single group validity, or 2) predictive for a majority group and less (but still

significantly) predictive for a minority. This latter notion is called differential validity. Single group validity can be considered a stronger version of differential validity. In the 70s, several scholars (Boehm, 1977; Schmidt, Berner, & Hunter, 1973; Katzell & Dyer, 1977) meta-analytically investigated how validity coefficients of cognitive ability tests differed between black and white Americans. Their findings led Schmidt et al. (1973) to conclude that differential validity is probably illusionary. Overall, later dated research seems to support this conclusion for American minorities (e.g., Hunter & Schmidt, 1978). However, some researchers (e.g., De Meijer et al., 2008; Te Nijenhuis & Van der Flier, 2000) did show some differential validity of cognitive measures for European non-western minorities.

Whereas cognitive measures' differential validity in general does not seem a major problem, ethnic minorities still score consistently lower on these measures. This results in a higher selection rate for majority groups and is commonly referred to as the diversity-validity dilemma (Pyburn, Ployhart, & Kravitz, 2008). Ployhart and Holtz (2008) reviewed research on the diversity-validity dilemma. To reduce racioethnic and sex subgroup differences, they recommended to supplement cognitive measures with non-cognitive measures (such as personality inventories) to more completely measure the full range of cognitive and non-cognitive Knowledge, Skills, Abilities, and Other characteristics (KSAOs).

Of the KSAOs usually measured with inventories, the personality traits Integrity and Conscientiousness have been found to show most incremental validity to cognitive tests (Schmidt & Hunter, 1998). Therefore, personality inventories seemingly make an ideal supplement to cognitive measures to reduce racioethnic differences. However, some studies have shown that personality inventories may actually have differential criterion validity. To date, only a few studies have investigated possible differential predictive validity of personality measures for different ethnic groups (De Meijer et al., 2008; De Vries, Born, & De Vries, 2012, Te Nijenhuis & Van der Flier, 2000). De Meijer et al. (2008) studied the prediction of police officers' training performance with personality inventories and found some differential validity of Intellect (i.e., Openness to Experience) between ethnic minority and majority applicants. The researchers found that Intellect was not predictive of the training performance of the majority group, and was negatively predictive of the performance of the minority group. Another study (De Vries et al., 2012) found differential validity of Conscientiousness and Integrity between ethnic minority and majority students for predicting grade point average (GPA). This study

showed that Conscientiousness was less predictive and Integrity more predictive for non-western minorities. De Vries et al. subsequently hypothesized that the differential predictive validity of Conscientiousness and Integrity could be caused by different personality-situation interactions for ethnic minority and majority students.

### *1.2 Reducing differential validity with contextualization of personality measures*

After a review of perspectives on the study of personality across cultures, Church (2010) recommended that future cross-cultural research should investigate established causal relations of personality on outcomes, such as behavior or performance. He suggests the use of contextualized measures to better capture the differential impact of situations on trait-relevant behaviors. 'Contextualization' means that personality is measured within a specific (type of) situation and is often achieved by adding relevant situational cues to a personality inventory (Schmit, Ryan, Stierwalt, & Powell, 1995). For example, an item from a generic personality inventory could be "I keep to my planning". The contextualized version of this item could be "I keep to my study planning". These added cues are often called a Frame-of-Reference (FoR; Mount, Barrick, & Strauss, 1994; Schmit et al., 1995). The idea behind adding a FoR is that personality is more consistent within one meaningful situation than when it is aggregated across several situations. Previous studies (e.g., Bing, Whanger, Davison, & VanHook, 2004; Holtrop, Born, De Vries, & De Vries, 2014; Hunthausen, Truxillo, Bauer, & Hammer, 2003; Pace & Brannick, 2010; Shaffer & Postlethwaite, 2012) have shown that contextualized personality inventories have incremental criterion validity over generic personality inventories, which is called the FoR effect.

Initially, the FoR effect was attributed to a reduction of between-person variance (e.g., Schmit et al., 1995). When filling out a generic (i.e., not contextualized) personality inventory, participants were thought to choose one seemingly random FoR and use that throughout the inventory. Contextualized personality inventories offer situational cues which elicit responses about only the target-situation. Thus participants fill out all items with the same context in mind, which causes the between-person variance to decrease, and subsequently results in a higher criterion validity for the target-situation. However, Lievens, De Corte, and Schollaert (2008) showed that when an inventory is made up of generic items, participants did not just use *one* random FoR, but instead alternated between *several* seemingly random FoR's. They observed that when an inventory is made up from items with

conceptually similar situational cues (i.e., contextualized items) participants no longer alternated between FoR's when answering the items and concluded that contextualized inventories therefore also reduce within-person variance. Moreover they showed that the amount of conceptual overlap between the predictor scale and criterion is a key determinant for the FoR effect, meaning that the amount of items in a scale related to the target-situation largely determines its criterion validity. Thus, Lievens et al. showed that the increased criterion validity of contextualized inventories (the FoR effect) is caused by both a reduction of between-person variance and a reduction of within-person variance.

In the Cognitive-Affective Personality System (CAPS) theory, Mischel and Shoda (1995) describe how situational cues trigger cognitive paths. The cognitive-affective system differs between individuals in the amount of cognitive-affective units, the activation level of the units, and the connections between the units. When a situation presents itself to an individual, the (person-specific) cues of that situation activate (or deactivate) cognitive-affective units. These units then trigger other units and so forth. Because networks differ in their sensitivity to situational cues, the network associated with a target-situation will fire more easily for some people than for others (e.g., the network "school" is more easily activated for some students than for other students).

We expect that people have different dominant contexts when they fill out a generic inventory. That is, we expect that people do not randomly select a FoR when answering generic items. Some groups of people, for whom a situation is very accessible, may use this situation more often than others when filling out a generic inventory. For example, one (cultural) group may fill out most generic items with mostly a school situation in mind (Figure 1a) whereas another (cultural) group fills them in with mostly another context in mind (e.g., home or work; Figure 1b). In this example, the effect for contextualization would be stronger for the latter group, because their answers would benefit the most from the predictor-criterion alignment (e.g., compare Figure 1c to 1b). In short: The within-person variability reduction of contextualization may differ between groups.

Ethnic minority students may be a group for whom the school context is less salient than most students. A large amount of research (e.g., Arends-Toth & Van de Vijver, 2003; Berry, 1997; Ryder, Alden, & Paulhus, 2000) has shown that ethnic minorities adopt the resident culture in their cultural self-view to different degrees. The degree to which a person adopts the resident

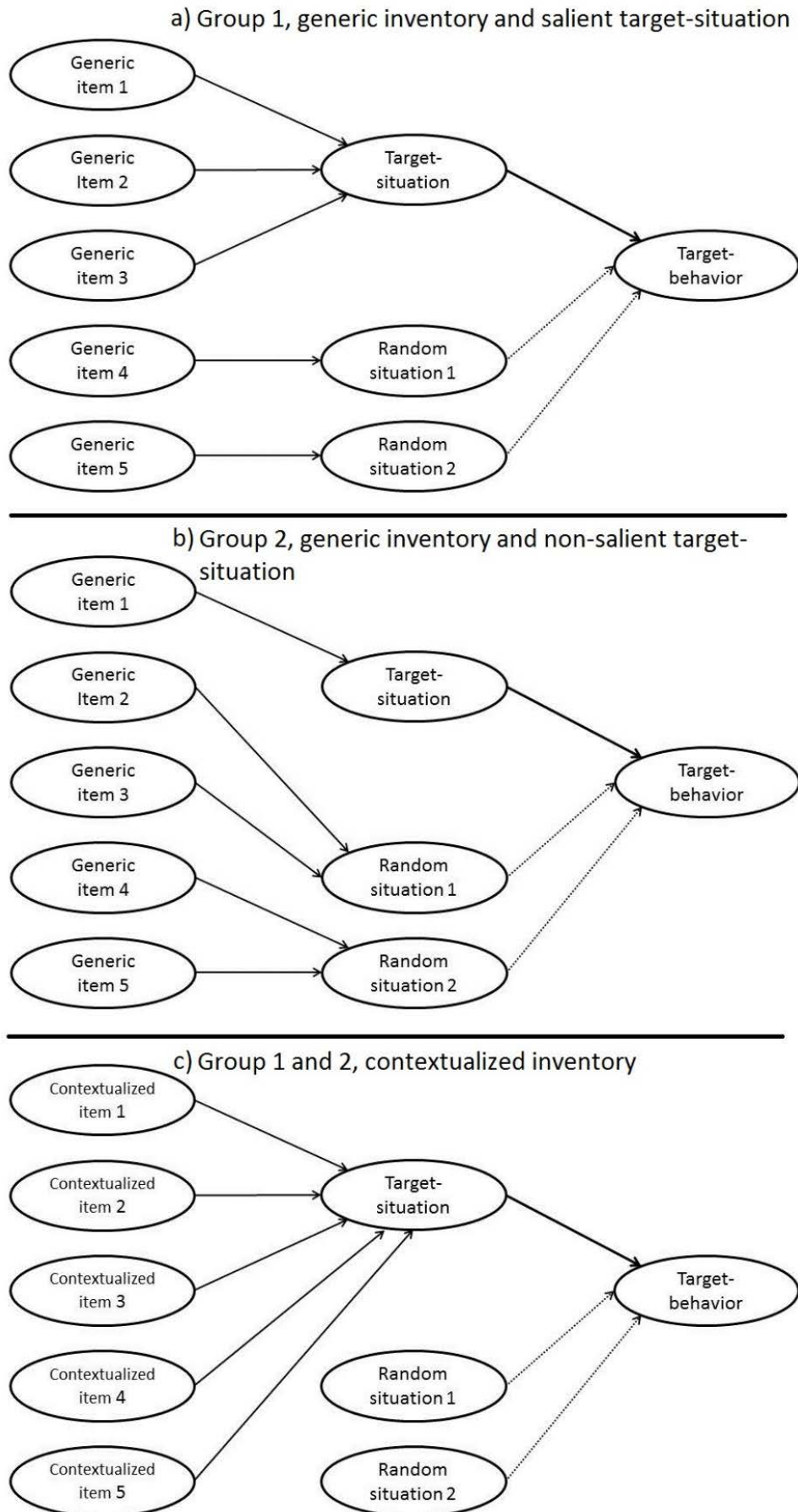


Figure 1. Example models of within-person variance, for two different groups, when filling in generic and contextualized inventories.

culture is commonly known as cultural Accommodation, a dimension of the bidimensional construct acculturation. The other acculturation dimension is called Maintenance, the degree to which a person maintains his/her paternal/maternal culture. Several researchers (Andriessen & Phalet, 2002; Hannover, Morf, Neuhaus, Rau, Wolfgramm, & Zander-Music, 2013) have reported that the degree to which students adopt the resident culture is positively related to their academic performance and that acculturation can vary between situations or life domains (e.g., Andriessen & Phalet, 2002; Arends-Toth & Van de Vijver, 2003, 2007). Subsequently, Hannover et al. (2013) found that only accommodation within a school context was related to academic performance and that accommodation in a home context was not. Acculturation in a relevant setting may thus contribute to the prediction of performance in that setting.

In conclusion, contextualized items force the same FoR on all participants and, as such, reduce the within-person variance which, in turn, improves criterion variance (i.e., the FoR effect). We expected that this reduction of within-person variability may differ between groups depending on the saliency of the context for the group. Previous research found that ethnic minority students—on the average—include a school situation to a lesser extent in their self-views (Hannover et al., 2013). Consequently, we expected the FoR effect to be stronger for ethnic minority students, compared to majority students, because a school situation may not be included in the minority's self-view. Moreover, we hypothesized that the FoR effect may be strongest for ethnic minority students who show low levels of cultural accommodation at school.

*Hypothesis 1.* The contextualized inventory will show incremental predictive validity over the generic inventory, but not vice versa.

*Hypothesis 2.* The incremental predictive validity of the contextualized inventory over the generic inventory will be significantly higher for ethnic minority students.

*Hypothesis 3.* The incremental predictive validity of the contextualized inventory over the generic inventory will be significantly higher for ethnic minority students with lower levels of accommodation.

The extent to which an instrument appears to be related to a task, or situation, improves perception of overall fairness of an assessment (Gilliland, 1993). If contextualized personality inventories indeed show less differential validity than generic personality inventories, then minority students may also

show more positive participant reactions towards contextualized inventories than to generic inventories (compared to majority students). Three participant reactions were measured in this study: 1) Liking of the inventory (Wiechman & Ryan, 2003), 2) Face validity, how relevant the students perceive the inventory for their role (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993), and 3) Perceived predictive validity, how predictive the students perceive the inventory for their performance (Smither et al., 1993).

*Hypothesis 4.* Compared to majority students, minority students will show more positive participant reactions towards the contextualized personality inventory than to the generic personality inventory.

### *1.3 The criterion: Self-reported and actual study performance*

High levels of correlation between GPA obtained from institutional records and self-reported GPA have been reported (e.g., Nofle & Robins, 2007; Schmitt, Oswald, Kim, Gillespie, Ramsay, & Yoo, 2003). However, GPA can show some notable differences depending on the ethnic group the participant belongs to (Kuncel et al., 2005). Kuncel et al. meta-analytically investigated the correlation between self-reported GPA and actual GPA. They found a higher correlation for white Americans ( $r_{\text{obs}} = .80$ ) than for minority students ( $r_{\text{obs}} = .66$ ). Overall, they found that the minority group reported their GPA less accurately (i.e., over- and under-reported GPA, compared to the white group). Kuncel et al. concluded that self-reported GPA showed substantial ethnic biases which may be tied to the overall lower academic performance of this group and may thus not be such a good criterion for minorities as some research portrayed it to be.

We aimed to further explore ethnic bias when using self-reported GPA in terms of differential criterion validity. In the present study, actual GPA (as obtained from the institutions' databases) and estimated GPA were obtained for all participants. First, it was investigated whether the finding of Kuncel et al. (2005) that minority students estimate their GPA less accurately, could be replicated. Second, the Frame-of-Reference effect and differential validity (as described in H2 and H3) were investigated using actual GPA and self-reported GPA as separate criteria.

## **2. Method**

### *2.1 Procedure and Participants*

Three subsamples were combined in the present study, which were collected at three large universities of applied sciences in The Netherlands (of

which the educational level could be described as higher vocational education). The students were in their second and third year from a wide array of four-year educational programs (e.g., economics, information technology, and pedagogy). All subsamples were invited to participate via a bulk e-mail, additionally the third subsample was also approached by research assistants. For the first subsample approximately 4.000 students were invited to participate, for the second approximately 2.000, and for the third approximately 700 (specifically targeted at non-western minority students). In the first subsample, 176 students completely filled out all questionnaires ( $M(\text{age}) = 23.19$ ,  $SD = 6.51$ ,  $\text{♂} = 32.4\%$ , non-western minority = 10.2%), in the second subsample, 48 students ( $M(\text{age}) = 22.56$ ,  $SD = 3.11$ ,  $\text{♂} = 29.2\%$ , non-western minority = 16.7%), and in the third subsample, 102 students ( $M(\text{age}) = 22.96$ ,  $SD = 4.27$ ,  $\text{♂} = 29.4\%$ , non-western minority = 82.4%). Next, we checked for subsample differences in students' self-reports. We found no significant differences between the subsamples on the personality measures (except for contextualized Openness on which subsample 3 scored significantly lower:  $F = 5.22$ ,  $p < .01$ ), acculturation, or participant reactions. Hence, it was decided to combine all subsamples. The complete sample consisted of 326 students ( $M(\text{age}) = 23.03$ ,  $SD = 5.47$ ,  $\text{♂} = 31\%$ ).

To establish if a student belonged to a minority we used the Statistics Netherlands (CBS, 2015b) definition of non-western ethnic minorities. This definition states that someone belongs to a non-western minority in The Netherlands if this person is or has at least one parent born in Africa, South America, Asia (excl. Indonesia and Japan), or Turkey. Given the size of our sample it was not possible to distinguish more homogeneous minority groups. This sample consisted of 190 (58.3%) Dutch ethnic majority students, 110 (33.7%) non-western ethnic minority students (of which approximately 60% with a Caribbean, Moroccan, Surinamese, or Turkish background, the largest non-Dutch ethnic groups in The Netherlands), and 26 (8%) western ethnic minority students. The western ethnic minority students were only included in analyses using the whole sample and were left out in the investigation of differential validity.

All students participated by filling in questionnaires on an online platform. Participation took approximately 45 minutes. First, the students filled out some personal details. Second, they filled out the generic personality inventory and third, they filled out the contextualized inventory. We chose not to randomize the personality questionnaires because we did not want the Frame-of-Reference of the contextualized inventory to carry over to the



generic inventory. Fourth and last, the students filled out an acculturation measure.

## 2.2 Predictor Measures

### 2.2.1 Generic personality measure: Multicultural Personality Test – Big Six (MPT-BS)

The MPT-BS (NOA, 2009; De Vries, De Vries, & Born 2011; Holtrop et al., 2014) is a personality inventory that consists of 200 short and easy-to-understand statements, measuring six personality dimensions: Integrity, Emotional stability, Extraversion, Agreeableness, Conscientiousness, and Openness. Its factor-level structure is based on the HEXACO personality model (Lee & Ashton, 2004), but it contains somewhat different facets (NOA, 2009). Our aim was to measure the full scope of personality, but to minimize attrition we reduced the items measuring Extraversion, Agreeableness, and Openness to 10 items per scale (for the generic and contextualized measure, reducing the length of both inventories by 35%). We chose to reduce the length of these scales, because previous research has shown that Conscientiousness (Chamorro-Premuzic & Furnham, 2003; Poropat, 2009), Integrity (De Vries et al., 2011; Van Iddekinge, Taylor, & Eidson, 2005), and Emotional stability (Chamorro-Premuzic & Furnham, 2003) are most predictive of academic performance. The items were selected based on their psychometric properties in unpublished datasets. In these datasets, the abbreviated scales showed very high correlations to the original main dimensions. Participants responded on a 5-point Likert scale, ranging from ‘*strongly disagree*’ to ‘*strongly agree*’. An example item of Conscientiousness (of the facet performance motivation) is: ‘*I want to be the best*’. In the present study, the alpha reliabilities of the full scales were .83 for Integrity, .94 for Emotional stability, and .91 for Conscientiousness, and for the abbreviated scales .86 for Extraversion, .81 for Agreeableness, and .82 for Openness.

### 2.2.2 Contextualized personality measure

The purpose of the contextualized personality inventory was to measure personality traits in a school setting. For this purpose, we used a contextualized personality measure previously designed by the authors (Holtrop, Born, De Vries, & De Vries, 2014). To construct this contextualized inventory, every item of the MPT-BS was completely revised with several field experts and psychological test developers. At the end of the design process, two experts on personality inventories performed a ‘back-translation’ to the facet level showing an acceptable inter-rater reliability of .80 (Krippendorff’s

alpha; Hayes & Krippendorff, 2007). The final contextualized inventory thus completely mirrored the generic inventory in terms of structure and item count and only differed in terms of item content, which was modified to reflect an academic setting. An example item of contextualized Conscientiousness is: '*I want to be better than other students*'. In the present study, alpha reliabilities of the contextualized scales were .77 for Integrity, .91 for Emotional stability, and .89 for Conscientiousness, and for the abbreviated contextualized scales .84 for Extraversion, .75 for Agreeableness, and .83 for Openness. The generic and contextualized scales correlated highly, from .66 for Agreeableness to .83 for Emotional stability.

### 2.2.3 Acculturation measure

Because an appropriate acculturation measure for our target group (students from various ethnic minorities in The Netherlands) was unavailable, we designed a general acculturation measure for the purpose of this study. In the process of designing the acculturation measure we paid particular attention to the recommendations proposed by Celenk and Van de Vijver (2011). In short, these recommendations are to 1) measure acculturation as a bidimensional construct, 2) consider the domain (private versus public), 3) consider the conditions for acculturation, 4) have a sufficient number of items per domain/dimension, and 5) aim for high alpha reliabilities.

The newly designed acculturation measure consisted of 28 items in total, 4 scales of 7 items each measured two dimensions of acculturation in the private (at home) and public domain (at school). Thus, we developed seven (very similar) items that measured Accommodation of the resident per domain, and we developed seven items for Maintenance of the original (maternal/fraternal) home culture per domain. An example item of Accommodation in a school setting is: "*How often do you eat Dutch food at school?*" An example item of Maintenance in a home setting is: "*How often do you watch television from your other culture at home?*" The items were answered on a seven-point scale ranging from "*Never*" (1) to "*Daily*" (7). We introduced this measure by asking the student if he/she felt that he/she belonged to another culture, besides the Dutch culture. Student who answered positively were asked which specific culture they meant and then they were presented with the acculturation items. On the page the acculturation items the students were instructed to "*Think of [insert the culture the student filled out] where the items below mention your other culture.*"

A Principal Component Analysis with Varimax Rotation<sup>1</sup> showed that the items for Accommodation and Maintenance, as intended, loaded highly on two separate scales. However, the item clusters showed no clear differences between the home and school domain. Therefore, it was decided to collapse the home and school domain in the two main dimensions. The alpha reliabilities of both dimensions were acceptable with .76 for accommodation and .83 for Maintenance.

### *2.3 Criterion Measures*

#### *2.3.1 Actual Grade Point Average*

Approximately two months after the students had participated in this study, their GPA over the entire school career at the institution was obtained from the institutions' records. The GPA criterion is therefore an objective measure of students' academic performance. GPA can range from 1 to 10, with higher scores indicating better performance and where 5.5 or higher is considered a sufficient grade.

#### *2.3.2 Self-reported Grade Point Average*

At the start of the questionnaire, the students were asked to fill out some personal details and to report their average grade at the applied university (1 to 10). The students were instructed to estimate their grade if they were not sure.

#### *2.3.3 Participant reactions*

Participant reactions were completed by all students for both personality inventories separately (i.e., generic and contextualized). Specifically, liking (Wiechman & Ryan, 2003; 4 items;  $\alpha(\text{current study}) = .71/.74$ ), face validity (Smither et al., 1993; 5 items;  $\alpha(\text{current study}) = .75/.75$ ), perceived predictive validity (Smither et al., 1993; 5 items;  $\alpha(\text{current study}) = .88/.86$ ) were measured. The alpha coefficients per ethnic group are shown in Table 5. Participants responded to all items on a 7-point Likert scale ranging from completely disagree to completely agree.

## **3. Results**

### *3.1 Descriptive statistics and preliminary analyses*

For the entire sample, the means, standard deviations, and correlations of the generic and contextualized personality scales, and of actual and self-reported GPA are shown in Table 1. Before turning to the main questions of

---

<sup>1</sup> Please contact the authors for the results of the PCA.

this study, we first took a closer look at the differences between ethnic groups in terms of academic achievement and self-reported personality. The non-western minority students, compared to the majority students, obtained ( $M_{\text{majority}} = 7.17$ ,  $SD = 0.56$ ,  $M_{\text{minority}} = 6.82$ ,  $SD = 0.70$ ,  $t = -4.82$ ,  $p < .01$ ) and reported ( $M_{\text{majority}} = 7.08$ ,  $SD = 0.64$ ,  $M_{\text{minority}} = 6.84$ ,  $SD = 0.53$ ,  $t = -3.40$ ,  $p < .01$ ) a significantly lower GPA. There were no significant mean differences on the personality scales. The means, standard deviations, and correlations, for the majority and non-western minority are shown in Table 2 and 3.

Next, the correlations between GPA and personality were investigated for the whole sample. For the generic personality measure, actual GPA correlated positively to generic Conscientiousness ( $r = .15$ ,  $p < .05$ ), and—unexpectedly—negatively to Agreeableness ( $r = -.14$ ,  $p < .05$ ; as shown in Table 1). Self-reported GPA did not correlate to Agreeableness ( $r = -.08$ ,  $p > .05$ ), but did correlate significantly to Integrity ( $r = .25$ ,  $p < .01$ ). Also, self-reported GPA correlated significantly stronger to generic Conscientiousness than actual GPA did ( $r_{\text{self-reportedGPA}} = .31$ ,  $r_{\text{actualGPA}} = .15$ ,  $z = -2.15$ ,  $p < .05$ ).

For the contextualized personality measure, actual GPA correlated positively to contextualized Conscientiousness ( $r = .24$ ,  $p < .01$ ), and to Integrity ( $r = .15$ ,  $p < .01$ ). Self-reported GPA also correlated significantly to Conscientiousness ( $r = .41$ ,  $p < .01$ ) and Integrity ( $r = .28$ ,  $p < .01$ ). Self-reported GPA correlated significantly stronger than actual GPA to contextualized Conscientiousness ( $r_{\text{self-reportedGPA}} = .41$ ,  $r_{\text{actualGPA}} = .24$ ,  $z = 2.43$ ,  $p < .01$ ) and Integrity ( $r_{\text{self-reportedGPA}} = .28$ ,  $r_{\text{actualGPA}} = .15$ ,  $z = 1.74$ ,  $p < .05$ ).

### 3.2 Hypothesis testing

To investigate the first hypothesis, several three-step hierarchical regression analyses were performed. Pace and Brannick (2010) previously used the same method to investigate the incremental criterion validity of contextualized personality inventories. If the contextualized inventory significantly increases explained variance over generic inventory, but not vice versa, then the contextualized inventory explains more variance than the generic (see Table 4). Every column in Table 4 shows a three-step hierarchical regression analysis. The hierarchical regression analyses were only performed for Conscientiousness, because Conscientiousness was the only trait for which the generic and contextualized measure correlated significantly to GPA. Generic Agreeableness ( $r = -.14$ ,  $p < .05$ ) and contextualized Integrity ( $r = .15$ ,  $p < .05$ ) also correlated weakly and significantly to actual GPA. However, contextualized Agreeableness and generic Integrity did not correlate

**Table 1.**

*Means, Standard Deviations, Reliabilities, Correlations of the generic and contextualized personality inventory (MPT-BS), actual, and self-reported GPA.*

	<i>M(SD)</i>		$\alpha$	1	2	3	4	5	6	7	8	
Personality scales												
1. Integrity	3.51(0.40)	3.64(0.36)	.83	.77	<b>.80**</b>	.35**	.08	.14*	.47**	.12*	.15**	.28**
2. Emotional Stability	3.51(0.49)	3.57(0.42)	.94	.91	.33**	<b>.83**</b>	.42**	.28**	.24**	.30**	.01	.21**
3. Extraversion	3.33(0.59)	3.38(0.58)	.86	.84	-.00	.46**	<b>.82**</b>	.16**	.06	.38**	-.05	.05
4. Agreeableness	3.83(0.45)	3.61(0.44)	.81	.75	.23**	.32**	.23**	<b>.66**</b>	.26**	.45**	-.04	.03
5. Conscientiousness	3.73(0.43)	3.65(0.42)	.91	.89	.41**	.21**	.05	.27**	<b>.83**</b>	.45**	.24**	.41**
6. Openness	3.55(0.53)	3.45(0.55)	.82	.83	.10	.44**	.46**	.42**	.32**	<b>.77**</b>	.03	.20**
Criterion												
7. Actual GPA	7.04(0.64)		-	.10	-.10	-.11	-.14*	.15**	-.05	-	-	-
8. Self-reported GPA	6.99(0.61)		-	.25**	.07	-.06	-.08	.31**	.06	.63**	-	-

*Note.*  $N = 326$ . For  $M$ ,  $\alpha$ , and  $SD$  the first value relates to the generic MPT-BS, the second value to the contextualized MPT-BS. Correlations between generic and contextualized scales are shown on the diagonal. Correlations below the diagonal relate to the generic MPT-BS scales to each other, above the diagonal to the contextualized MPT-BS. Self-reported personality ranged from 1 to 5. Actual and self-reported GPA ranged from 1 to 10.

\* $p < .05$ , \*\* $p < .01$ .

**Table 2.**

*Majority students: Means, Standard Deviations, Reliabilities, Correlations of the generic and contextualized personality inventory (MPT-BS), actual, and self-reported GPA.*

	<i>M(SD)</i>		$\alpha$	1	2	3	4	5	6	7	8	
Personality scales												
1. Integrity	3.50(0.38)	3.64(0.33)	.82	.76	<b>.79**</b>	.27**	-.03	.10	.45**	.12	.19**	.28**
2. Emotional Stability	3.49(0.50)	3.57(0.42)	.94	.91	.26**	<b>.82**</b>	.43**	.22**	.12	.33**	.08	.20**
3. Extraversion	3.31(0.59)	3.36(0.60)	.87	.86	-.12	.50**	<b>.85**</b>	.06	-.07	.32**	-.09	-.02
4. Agreeableness	3.81(0.47)	3.60(0.43)	.84	.76	.22**	.28**	.19**	<b>.65**</b>	.23**	.44**	-.09	-.03
5. Conscientiousness	3.72(0.44)	3.66(0.41)	.91	.89	.39**	.07	-.06	.19**	<b>.86**</b>	.32**	.39**	.42**
6. Openness	3.56(0.50)	3.47(0.51)	.81	.82	.06	.46**	.48**	.42**	.15*	<b>.78**</b>	.05	.13
Criterion												
7. Actual GPA	7.17(0.56)		-	.14	-.12	-.23**	-.27**	.29**	-.11	-	-	-
8. Self-reported GPA	7.08(0.64)		-	.27**	.01	-.14*	-.16*	.30**	-.04	.80**	-	-

*Note.*  $N = 190$ . For  $M$ ,  $\alpha$ , and  $SD$  the first value relates to the generic MPT-BS, the second value to the contextualized MPT-BS. Correlations between generic and contextualized scales are shown on the diagonal. Correlations below the diagonal relate to the generic MPT-BS scales to each other, above the diagonal to the contextualized MPT-BS. Self-reported personality ranged from 1 to 5. Actual and self-reported GPA ranged from 1 to 10.

\*  $p < .05$ , \*\*  $p < .01$ .

**Table 3.**

*Non-western minority students: Means, Standard Deviations, Reliabilities, Correlations of the generic and contextualized personality inventory (MPT-BS), actual, and self-reported GPA.*

	M(SD)		$\alpha$	1	2	3	4	5	6	7	8	9	10	
Personality scales														
1. Integrity	3.54(0.45)	3.65(0.40)	.85	.81	<b>.80**</b>	.43**	.23*	.11	.54**	.16	.08	.31**	-.04	-.14
2. Emotional Stability	3.55(0.47)	3.57(0.42)	.93	.90	.42**	<b>.84**</b>	.46**	.34**	.45**	.34**	-.13	.25**	.11	.08
3. Extraversion	3.37(0.59)	3.41(0.57)	.85	.81	.11	.41**	<b>.78**</b>	.35**	.25**	.46**	-.02	.20*	.02	.05
4. Agreeableness	3.86(0.44)	3.63(0.44)	.80	.73	.21*	.39**	.32**	<b>.67**</b>	.33**	.60**	-.01	.16	.06	.17
5. Conscientiousness	3.75(0.44)	3.62(0.45)	.91	.90	.48**	.46**	.22*	.40**	<b>.82**</b>	.59**	.05	.40**	.01	-.09
6. Openness	3.55(0.58)	3.39(0.60)	.85	.85	.13	.45**	.45**	.46**	.52**	<b>.75**</b>	-.07	.25**	-.00	.08
Criterion														
7. Actual GPA	6.82(0.70)		-	.05	-.15	.01	.04	-.02	-.04	-	-	-	-	-
8. Self-reported GPA	6.84(0.53)		-	.27**	.22*	.12	.14	.36**	.18	.33**	-	-	-	-
9. Acculturation: Accommodation	5.11(0.88)		.76	.02	.13	.02	.12	.08	.13	.04	.16	-	-	-
10. Acculturation: Maintenance	4.13(1.13)		.83	-.00	.11	.18	.13	-.07	.06	-.21*	.09	.05	-	-

*Note.*  $N = 110$ . For  $M$ ,  $\alpha$ , and  $SD$  the first value relates to the generic MPT-BS, the second value to the contextualized MPT-BS. Correlations between generic and contextualized scales are shown on the diagonal. Correlations below the diagonal relate to the generic MPT-BS scales to each other, above the diagonal to the contextualized MPT-BS. Self-reported personality ranged from 1 to 5. Actual and self-reported GPA ranged from 1 to 10. Acculturation scales ranged from 1 to 7.

\*  $p < .05$ , \*\*  $p < .01$ .

significantly to actual GPA. Incremental validity is only meaningful if both predictors are related to the criterion, therefore computing the incremental validity for these traits would be futile.

For the whole sample, the results showed a clear pattern. The hierarchical regression analysis with Conscientiousness showed that the contextualized questionnaire explained more variance than the generic version in actual GPA ( $\Delta R^2 = .07, p < .01$ ) and self-reported GPA ( $\Delta R^2 = .08, p < .01$ ), and not vice versa (for both GPA measures:  $\Delta R^2 = .01, p > .05$ ). Thus, the incremental criterion validity was small to moderate. Thus, a FoR effect was found and H1 was supported.

To test H2, that the FoR effect would be stronger for minority students than for majority students, hierarchical regression analyses were performed for both groups separately. The FoR effect was found for the majority students. Contextualized Conscientiousness explained more variance than generic Conscientiousness in actual GPA ( $\Delta R^2 = .08, p < .01$ ) and self-reported GPA ( $\Delta R^2 = .10, p < .01$ ), but not vice versa. Contrary to our expectations, we found a complete lack of a FoR effect for the minority students for actual and self-reported GPA. For actual GPA, generic and contextualized Conscientiousness showed no significant correlations (resp.  $r = -.02, p > .05$  and  $r = .05, p > .05$ ; see Table 3). Consequently, neither had incremental criterion validity over the other. The FoR effect was also not found for self-reported GPA of the minority group. In this case generic and contextualized Conscientiousness did show significant correlations (resp.  $r = .36, p < .01$  and  $r = .40, p < .01$ ). However, neither explained significantly more variance over the other (generic:  $\Delta R^2 = .02, p > .05$ ; contextualized:  $\Delta R^2 = .03, p > .05$ ). H2 was thus not supported for Conscientiousness. Furthermore, on the matter of differential validity, generic Conscientiousness ( $r_{\text{majority}} = .29, r_{\text{minority}} = -.02, z = -2.63, p < .01$ ) and contextualized Conscientiousness ( $r_{\text{majority}} = .39, r_{\text{minority}} = .05, z = -2.98, p < .01$ ) were significantly more weakly related to actual GPA among the non-western minority.

To investigate H3, that the FoR effect would be stronger for non-western minority students with lower levels of Accommodation, interaction terms of standardized Conscientiousness (both versions) and standardized Accommodation were computed. Next, both main effects and the interaction effect were computed with a regression analysis in which the interaction effect was entered last. For actual and self-reported GPA, the interaction effect did not add incremental criterion validity for generic Conscientiousness (actual GPA:  $\Delta R^2 = .01, p > .05$ ; self-reported GPA:  $\Delta R^2 = .00, p > .05$ ) and



**Table 4.**

*Hierarchical regression analyses results for Conscientiousness on actual GPA and self-reported GPA.*

		Conscientiousness $R^2(\Delta R^2)$ on actual GPA		
		Whole sample	Majority	Non-western minority
Step 1	Age, gender	.01	.01	.00
Step 2	Generic Context.	.03(.02*)	.10(.09**)	.00(.00)
		.07(.06*)	.17(.16*)	.00(.00)
Step 3	Generic Context.	.08(.07*)	.18(.08*)	.02(.02)
			.18(.01)	.02(.01)
		Conscientiousness $R^2(\Delta R^2)$ on self-reported GPA		
		Whole sample	Majority	Non-western minority
Step 1	Age, gender	.01	.01	.06
Step 2	Generic Context.	.10(.09*)	.11(.10*)	.15(.09*)
		.18(.17*)	.20(.19*)	.18(.12*)
Step 3	Generic Context.	.18(.08*)	.21(.10*)	.18(.02)
		.18(.01)	.21(.01)	.18(.03)

*Note.*  $N = 309$  for the whole sample,  $n = 190$  for the majority sample, and  $n = 110$  for the minority sample. For each sample two hierarchical regression analyses are reported next to each other.

\*  $p < .01$

contextualized Conscientiousness (actual GPA:  $\Delta R^2 = .00$ ,  $p > .05$ ; self-reported GPA:  $\Delta R^2 = .00$ ,  $p > .05$ ). Hence, H3 was not supported. Note that these regression analyses were repeated with an interaction term of Conscientiousness and Maintenance (instead of Accommodation) and that these results were also not significant.

Next, to test H4, differences in participant reactions for the generic and contextualized version were analyzed (see Table 5). There were no significant mean differences in how the majority and minority students viewed the inventories. Minority and majority students equally liked both personality inventories and found them equally face valid and predictive. Hence, H4 was not supported. Both groups found the contextualized version more face valid (majority:  $t = 8.93$ ,  $p < .01$ ; minority:  $t = 6.37$ ,  $p < .01$ ), but did not like the contextualized measure more, and did not find it more predictive.

Last of all, the correlation between actual GPA and self-reported GPA was investigated. For the whole sample together, actual GPA correlated highly to self-reported GPA ( $r = .63$ ,  $p < .01$ ; as shown in Table 1). However, when majority and non-western minority students were analyzed separately, this correlation was significantly stronger ( $z = 6.24$ ,  $p < .01$ ) for the majority students ( $r = .80$ ,  $p < .01$ , as shown in Table 2) than for the non-western minority students ( $r = .33$ ,  $p < .01$ , as shown in Table 3). Note that on average non-western minority students did not over- or underreport their grade ( $t =$

**Table 5.**

*Means, Standard Deviations of participant reactions to the generic and contextualized personality inventory for the whole sample, and separately for the majority and non-western minority students.*

	Whole sample			Majority				Non-western minority				<i>t</i> (majority, minority)
	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$	<i>t</i> (generic, context.)	<i>M</i>	<i>SD</i>	$\alpha$	<i>t</i> (generic, context.)	
Generic personality												
Liking	4.26	0.92	.71	4.26	0.93	.70	-	4.26	0.92	.67	-	-0.04
Face validity	4.37	1.13	.75	4.25	1.10	.78	-	4.45	1.14	.71	-	1.44
Perceived predictive validity	2.83	1.23	.88	2.83	1.27	.87	-	2.83	1.22	.88	-	0.43
Overall participant reactions	3.79	0.73	.77	3.74	0.71	.78	-	3.82	0.75	.72	-	0.81
Contextualized personality												
Liking	4.27	0.96	.74	4.22	1.04	.70	0.92	4.30	0.91	.77	-0.53	0.70
Face validity	5.09	1.06	.75	5.02	1.03	.76	8.93*	5.13	1.08	.74	6.37*	0.84
Perceived predictive validity	2.93	1.21	.86	2.87	1.29	.86	1.85	2.96	1.16	.88	0.45	0.62
Overall participant reactions	4.08	0.71	.76	4.02	0.68	.78	6.73*	4.12	0.72	.72	4.49*	1.10

*Note.*  $N = 309$  for the whole sample,  $n = 190$  for the majority sample, and  $n = 110$  for the minority sample. No significant differences were found between the participant reactions of the majority and minority.

\*  $p < .01$

0.17,  $p > .05$ ) and showed moderate agreement between self-reported GPA and actual GPA ( $r = .33$ ,  $p < .01$ ). On average, the majority students underreported their GPA ( $t = -3.12$ ,  $p < .01$ ), but their self-reported GPA correlated significantly stronger to their actual GPA ( $r = .80$ ,  $p < .01$ ) than the self-reported GPA of the minority students ( $z = 6.24$ ,  $p < .01$ ).

#### 4. Discussion

This study investigated if contextualization is able to reduce the differential validity of personality inventories. First, the well-established Frame-of-Reference effect was found for the whole sample: The predictive validity of the personality inventory increased when the inventory's items were modified to reflect the situation in which the criterion is measured. Next, the suggestion of Church (2010) and De Vries et al. (2012), that contextualization may resolve the differential validity of personality inventories, was investigated. Contrary to their suggestion, neither generic nor contextualized personality was predictive of actual GPA for the non-western minority. Thus, the Frame-of-Reference effect was only present for majority students and not for minority students. Moreover, neither Conscientiousness measure was predictive of the actual GPA in the non-western minority, whereas they were in the majority group. This finding has two implications: 1) personality inventories seem to have differential criterion validity, or even single group validity, in an academic context, and 2) contextualization may not solve this differential validity dilemma. Moreover, the extent to which non-western minority students had included the resident (Dutch) culture in their self-view, their accommodation, did not show any interaction with Conscientiousness in predicting GPA. Hence, the degree to which a student felt part of the majority culture did not explain the differential validity. Despite the differential validity, minority students did not indicate to like the personality inventories differently, or found these less face valid and predictive than majority students. Both groups of students found the contextualized inventory more face valid.

Contrary to the differential validity for long term outcomes, Church, Katigbak, Miramontes, Del Prado, and Cabrera (2007) showed that personality has no differential predictive validity of recent behaviors. Church et al. showed that personality can predict retrospectively reported behavior of the previous month equally well for Americans and Filipinos. Combined with our finding that personality inventories appear to show differential validity for long term outcomes (i.e., GPA), it may be that personality predicts direct behavior equally well across groups, but other contextual or cultural factors may moderate the effect that behavior has on criteria further in time. Indeed, some

research showed that minority students' academic success is affected by factors that are not salient for majority students. For example, minority students often do not have a private social network that is fully able to support them and therefore benefit more strongly from a social network at school (Wolff, 2013) and they may experience more family-study conflict (Meeuwisse, Born, & Severiens, 2014). In short, contrary to some suggestions, contextualization did not resolve the differential validity of personality inventories. A possible explanation for this finding can be that the cause of differential validity of academic success is not rooted inside the criterion's Frame-of-Reference (i.e., the school context) but in the private domain.

In this study we used both actual and self-reported GPA. These two criteria led to two unexpected findings. First, the predictive validity of the personality scales was higher for self-reported GPA than for actual GPA. Second, self-reported GPA of non-western minority students correlated weaker to actual GPA than self-reported GPA of majority students. Whereas for the majority students, most predictor-criterion relations were somewhat amplified for self-reported GPA, for the non-western minority students the results showed drastic changes. For the minority students, the personality scales were uncorrelated to actual GPA, but most of these were correlated to self-reported GPA. More precisely, the average correlation of the generic scales was borderline significantly weaker with actual GPA than with self-reported GPA ( $\bar{r}_{\text{actualGPA}} = -.02$ ,  $\bar{r}_{\text{self-reportedGPA}} = .22$ ,  $z = -1.49$ ,  $p = .07$ ), and the average correlation of the contextualized scales was significantly weaker with actual GPA than with self-reported GPA ( $\bar{r}_{\text{actualGPA}} = -.02$ ,  $\bar{r}_{\text{self-reportedGPA}} = .26$ ,  $z = -1.80$ ,  $p < .05$ ). This higher average correlation between self-reported predictors and self-reported criteria may be attributed to a same-source bias. This same-source variance appears to be stronger for non-western minority students than for majority students. This bias may be caused by a strong overarching response style that contaminates self-reports of minority students and, as such, may distort the measurement of personality and its predictive validity. Indeed, He and Van de Vijver (2013) found that non-western minorities exhibit stronger response styles when filling in personality inventories (i.e., more acquiescence and midpoint responding). These response styles may carry over to self-reported criteria as well, which may explain why personality and self-reported criteria are related stronger for the non-western minority students.

Additionally, note that Kuncel et al. (2005) suggested that minority students may estimate their actual GPA less accurately because they obtain

lower grades on average. These authors found that students with a lower actual GPA reported their GPA less accurately. Therefore, we checked if the average lower actual GPA of the non-western minority students could account for their lower accuracy in self-reported GPA. A mean-split of actual GPA across the whole group put 48% ( $n = 91$ ) of the majority students and 70% ( $n = 76$ ,  $p < .01$ ) of the non-western students under the mean. The majority group under the mean reported their GPA much more accurately ( $r = .45$ ,  $p < .01$ ) than the minority group under the mean ( $r = .17$ ,  $p > .05$ ). Hence, the difference in accuracy of self-reported GPA between majority and non-western minority students cannot be fully explained by differences in actual GPA. Researchers should thus be aware that self-reported GPA may show misleadingly high correlations to actual GPA, which may mask important subgroup differences in predictive validity.

The collection of two criteria, actual and self-reported GPA, is an important strength of the present study. Most studies use one or the other and none has investigated differential predictive validity with two types of criteria. One of the greatest limitations of the present study is the sample of non-western minority students which is still very heterogeneous (i.e., approximately 60% of the participants had a Caribbean, Moroccan, Surinamese, or Turkish background, the remainder ranged from even more diverse backgrounds). We combined three subsamples from different applied universities to collect a sufficient number of non-western minority students. There were no significant differences between these subsamples in the students' self-reported measures. However, it would have been preferable to also distinguish between the ethnic subgroups in the non-western minority group.

Future research may require a multi-method, multi-situation design to more fully understand the differential validity of personality inventories. Differential validity may be caused at three moments: at the moment of the self-report, in the process from personality to criterion, and at the moment of the criterion report. At the moment of the self-report, the ethnic minority may interpret and respond to the items differently (e.g., He and Van de Vijver, 2013). The effect of this response style can possibly be disentangled by using other-ratings of personality. More specifically, a recommendation could be to use majority other-ratings for minority members and vice versa. If the other-rating by the majority judges of the minority targets are predictive of criteria, and the other-rating by the minority judges of the majority targets are not, then the self-report bias is most likely the cause of the lacking predictive validity.

4

Next, the process from personality to behavior to outcomes may be illuminated by alternative measurement methods. To reduce racioethnic differences in selection, Ployhart and Holtz (2008) suggested to not only supplement cognitive measures with measures of KSAO's, they also suggested to use more diverse measurement methods. Heller, Watson, Komar, Min, and Perunovic (2007) described several alternative personality measurement methods, such as diary measures, with which the interaction between personality and cultural cues may be further understood. Diary measurement may be useful to measure personality across a number of situations and also inquire about the participants' perceptions of the situations' affordances. Moreover, this research could investigate which participants feel the need to express certain behaviors in certain situations (also called trait activation; Tett & Burnett, 2003). In this manner the interaction of personality with (perceived) situational affordances could be disentangled and possibly explain why traits are differentially predictive of performance. For example, it may be that the situational affordances of a school setting activate conscientious behaviors for some and different behaviors (e.g., social behaviors) for others.

Last of all, future research may like to consider the possibility that minority students receive lower grades due to majority rater biases (i.e., a criterion bias). More specifically, GPA in our samples consists not only of grades of largely anonymous multiple choice test scores, but also consists of more personal essays, internships, practical assignments and presentations. For work performance, supervisory ratings of white managers have been found to be higher for white employees than for black employees, whereas black managers rated white and black employees equally high (Stauffer & Buckley, 2005). The grades given by majority raters, who constitute the largest part of the educational system and thus produce the largest amount of grades, could possibly be influenced by rater biases such as the Pygmalion/Golem effect (i.e., the finding that teacher/supervisor expectations affect the grade a student receives; e.g., Reynolds, 2007). Future studies investigating differential validity of personality for academic performance could, for instance, try to distinguish between types of grades to determine if a criterion bias exists.

In conclusion, we replicated the FoR effect for majority students. However, the generic and contextualized inventory were not predictively valid for non-western minority students. Therefore, according to the present study contextualization appears not to solve differential validity in personality

inventories. Additionally, we found that non-western minority students estimate their GPA less accurately and that the use of self-reported GPA may mask important subgroup differences in predictive validity.





# Relating the Spherical representation of vocational interests to the HEXACO personality model



5

# **Abstract**

## **Relating the Spherical representation of vocational interests to the HEXACO personality model**

The present study extends previous research on interests-personality relations by comparing recent models of vocational interests (using the Personal Globe Inventory; PGI, Tracey, 2002) and personality (using the HEXACO-PI-R; Ashton, Lee, & De Vries, 2014) with each other. First, the structure of the Spherical representation was adequately replicated in a Dutch sample ( $N = 656$ ). Second, in so far as comparisons were possible, the relations between interests and personality were found to be congruent with previous findings. Third, Prestige interests, the defining feature of the Spherical representation, were related, albeit weakly, to Openness to Experience and to Extraversion. Last of all, Honesty-Humility and Openness to Experience were related to profile elevation in interest scores. All results were obtained for normative and ipsatized scales, revealing several meaningful differences in interests-personality relations depending on the type of interest scoring procedure.

---

This chapter has been published as Holtrop, D., Born, M. Ph., & De Vries, R. E. (2015). Relating the Spherical representation of vocational interests to the HEXACO personality model. *Journal of Vocational Behavior*, 89, 10-20. DOI:10.1016/j.jvb.2015.04.003.

Grateful acknowledgement is provided to Prof. Terence J.G. Tracey for his help with several analyses and to Martine Schut, Rika Mohesi, Loen van Gulick, Angela Bijnsdorp, Elin Hellqvist, Eveline Kreuk, Jim Molenaar, Leontine Hoekemeijer, and Roos Plumiers for their assistance in collecting data.

## 1. Introduction

During the last decade several meta-analytic studies (e.g., Barrick, Mount, & Gupta, 2003; Larson, Rottinghaus, & Borgen, 2002) have advanced our understanding of the associations between interests and personality. These studies established a number of recurring relations between RIASEC vocational interests (Holland, 1959) and Big Five personality (Goldberg, 1990). At approximately the same time revised models emerged both for vocational interests and personality. The Spherical representation (Tracey & Rounds, 1996) was introduced in the field of vocational interests and the HEXACO model (Lee & Ashton, 2004) in the field of personality. This study compares these revised models for the first time.

The main difference between the Spherical representation of vocational interests and the RIASEC model, is the addition of *Prestige interests*. The main difference between the HEXACO model and the Big Five model is the addition of the *Honesty-Humility dimension*. So far, the HEXACO model has been related to vocational interests in only two recent studies (McKay & Tokar, 2012; Pozzebon, Visser, Ashton, Lee, & Goldberg, 2010), but not to the Spherical representation. To our knowledge, the Spherical representation has not yet been related to personality measures.

The purpose of the present study is threefold. First, we investigate the structure of the Spherical representation of vocational interests in a Dutch sample. Second, we investigate the relations between vocational interests and personality and compare these to previously found results. Last, we investigate the relations between Prestige interests and HEXACO personality. This investigation is particularly interesting because the distinguishing features of both models – at face value – seem to revolve around status. That is, Prestige interests seem to pit high-status jobs (i.e., high Prestige) against low-status jobs (i.e., low Prestige) and Honesty-Humility seems to inversely pit high-status motivations (i.e., low Honesty-Humility) against low-status motivations (i.e., high Honesty-Humility).

### 1.1 The structure of the Spherical representation of vocational interests

Holland (1959) proposed that vocational interests may be captured by six themes (called ‘types’ by Holland (1959)). He formulated these themes as the acronym RIASEC: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. The six interests are conceptualized using six evenly distributed vectors in a circumplex (see Figure 1). Despite the fact that the RIASEC model is the most often used model to conceptualize vocational

interests, it is not without its shortcomings. A limitation of the RIASEC model is that some studies outside North-America have not been able to replicate the structure of the RIASEC model (e.g., Einarsdóttir, Rounds, & Su, 2010; Rounds & Tracey, 1996). These divergent findings about the structure of the RIASEC model show that it is important to first approximate the structural (e)qualities of an interest instrument before employing it in research and practice (Tracey & Gupta, 2008).

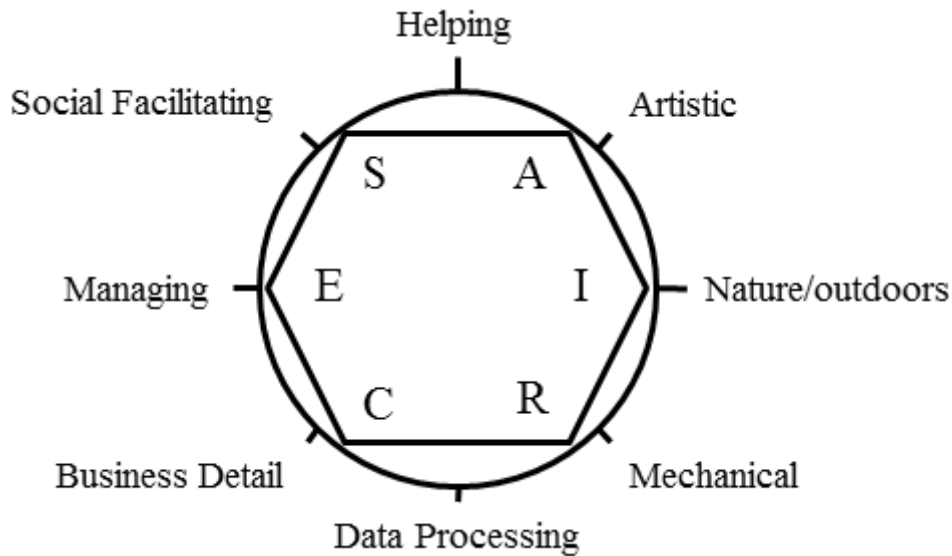
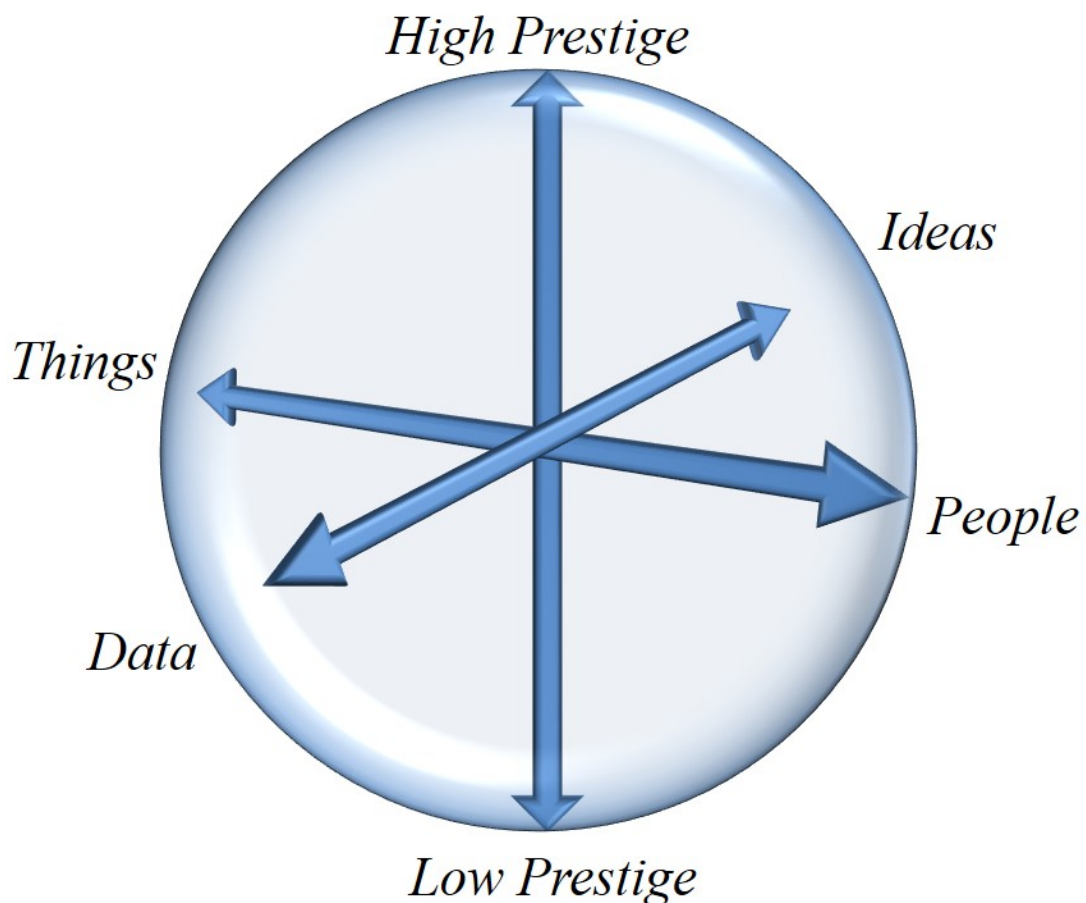


Figure 1. The PGI-octants and RIASEC interest scales.

This study employed the Spherical representation to model vocational interests. So far, the structure of the Spherical representation of vocational interests has been confirmed in several samples outside North-America (e.g., Irish, Chinese, and Croatian; respectively Darcy, 2005; Long, Adams, & Tracey, 2005; Sverko, 2008). However, nothing is yet known about its psychometric and structural properties in a Dutch setting. Consequently, following the suggestions of Tracey and Gupta (2008), the first purpose of our study was to test the psychometric and structural properties of the Spherical representation.

The Spherical representation of vocational interests is a three-dimensional model of interests. It expands Prediger's (1982) model, which structures interests in two right-angled axes, namely Ideas/Data and People/Things. Rounds and Tracey (1993) showed that the two axes proposed by Prediger and the RIASEC are actually two different representations of the same circumplex of vocational interest. Subsequently, Tracey and Rounds (1996) proposed another axis of vocational interest in addition to the two axes proposed by Prediger, which they named the Prestige

interests axis. They proposed that this dimension is orthogonal to the Ideas/Data and People/Things axes identified by Prediger (1982), changing the circular representation into a spherical representation (or globe; Figure 2), on which the original circumplex of vocational interests forms the equator. In the present study the spherical representation is measured with the Personal Globe Inventory (PGI; Tracey, 2002). The PGI measures the circumplex at the equator using an octagon. As illustrated in Figure 1, the octagonal representation covers the same space as the RIASEC hexagonal representation does.



*Figure 2.* A visual representation of the three main dimensions of the Spherical representation of vocational interests.

When measuring vocational interests, the inventories nearly always show profile elevation (i.e., a large general factor) that explains a vast amount of variance in all scales (Tracey, 2012). Some researchers (e.g., Fuller, Holland, & Johnston, 1999; Hirschi & Läge, 2007) have proposed a ‘substantive’ interpretation of this general factor. Accordingly, profile elevation has been related to, for instance, career planning, career exploration (Hirschi & Läge, 2007), and personality traits (Fuller et al., 1999). Fuller et al. showed that profile elevation is related to Openness to Experience and Extraversion.

Other researchers (e.g., Prediger, 1998) have suggested a ‘nuisance’ interpretation of the general factor and have advised interest assessment practitioners to work around the general factor by applying methods that largely ignore its influence. That is, Prediger suggested to use differential values or highest scale scores for interest assessments instead of normative values. Still others (e.g., Tracey, 2012) have proposed an ‘artifact’ interpretation of the general factor, maintaining that the general factor actually constitutes a prominent bias, which contaminates the relations between all interest scales and other variables. To enable a more adequate interpretation of correlations with other variables, Tracey (2012) suggested to partial out the general factor from all interest scales. However, interpreting profile elevation as substance, nuisance, or artifact does not necessarily exclude the other approaches; all three approaches may actually be true. In this study, we tested the general factor using both a substantive and an artifact approach. That is, we tested the structural (e)qualities of the spherical representation and its relations with personality using normative scales, profile elevation (based on the substantive approach), and ipsatized scales (based on the artifact approach).

### *1.2 Known relations between vocational interests and personality traits*

Personality is commonly measured using the Big Five dimensions (Schneider, 2007). Consequently, all meta-analytic studies on the relations between interests and personality have used the RIASEC scales and the Big Five personality dimensions (e.g., Barrick, et al., 2003; Larson et al., 2002). However, as several researchers have indicated, Integrity complements the Big Five personality dimensions (Lee & Ashton, 2004; Schneider, 2007). Based on lexical studies using a similar approach and datasets as prior Big Five studies, six personality dimensions have been uncovered in over nine countries (e.g., Ashton et al., 2004; Ashton, Lee, Marcus, & De Vries, 2007). These six dimensions are known by the acronym HEXACO, which stands for: Honesty-humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O; Lee & Ashton, 2004). Each of these six main dimensions consists of four facets. The HEXACO dimensions together explain more variance than the Big Five dimensions in anti-social organizational behavior (Lee, Ashton, & Shin, 2005), in delinquency (Lee, Ashton, & De Vries, 2005), in counterproductive academic behavior (Marcus, Lee, & Ashton, 2007; De Vries, De Vries, & Born, 2010), and – most importantly for this study – in RIASEC vocational interests (McKay & Tokar, 2012).

Honesty-Humility, the most distinguishing personality dimension in the HEXACO model, closely resembles Integrity (Marcus et al., 2007). The Honesty-Humility scale measures whether an individual has the tendency to be sincere, fair, modest, and to not show greed. Another change compared to the Big Five model is that the personality dimensions Emotionality and Agreeableness have been rotated in the HEXACO model. Content associated with temper and irritability has been moved from Emotionality (in the Big Five known as (low) Emotional Stability) to Agreeableness, and content associated with sentimentality from Agreeableness to Emotionality.

The second purpose of our study was to replicate and contribute to previous findings on the relations between interests and the HEXACO personality model. Currently, only two North American studies (McKay & Tokar, 2012; Pozzebon et al., 2010) have compared the HEXACO personality model to vocational interests and these studies have not used the spherical representation for this comparison. Pozzebon et al. (2010) developed the Oregon Vocational Interest Scales (ORVIS) and studied the relation between the ORVIS and the HEXACO. ORVIS structures vocational interest in eight scales. Although the ORVIS are named differently, the authors posited that seven of the ORVIS together fully account for the six RIASEC scales (two scales make up Realistic interests). Honesty-Humility correlated moderately and negatively with ORVIS Leadership (Enterprising;  $r = -.27$ ). Other personality dimensions and interest scales with correlations higher than .40 included: Extraversion positively with Leadership, Emotionality negatively with Adventure (part of Realistic interests), and Openness to Experience positively with Creativity (Artistic) and Erudition (no RIASEC equal).

McKay and Tokar (2012) related the RIASEC to the HEXACO. Most importantly, they showed that the HEXACO model of personality explained 4-23% (depending on gender and interest scale used) more variance in vocational interests than the Big Five model did. They also found that the RIASEC scale Enterprising correlated negatively with Honesty-Humility and that Realistic correlated negatively with HEXACO Emotionality. The RIASEC scales Investigative and Artistic correlated positively with Openness to Experience. Social interest correlated positively with several personality dimensions: Honesty-Humility, Extraversion, Agreeableness, and Openness to Experience.

To test the overlap between interests and the HEXACO personality model, for our sample, we first computed the relations between PGI-octant

scales to the HEXACO personality dimensions. We then also compared our findings to those of McKay and Tokar (2012) and Pozzebon et al. (2010). Both previous studies used normative, and not ipsatized, scales when comparing interests to personality. We analyzed the relations between interests and personality for our sample, and the congruence between our findings and those of previous studies, with normative and ipsatized interest scales.

### *1.3 Prestige interests and the HEXACO personality model*

So far, no research has related Prestige interests to the HEXACO model and especially to Honesty-Humility. Previous studies have found Prestige interests to be positively related to specific vocational preparation, and level of autonomy (e.g., Gottfredson, 1980). Sodano and Tracey (2008) found that Prestige interests were also related to educational level, required skill and effort for the activity, and competition involved in the activity. In addition, they found that Prestige interests were related to neither men, nor women sex-typing of the activity. Lastly, Sodano (2011) found Prestige interests to be positively related to valuing achievement, and negatively to valuing (physical) power. Our third purpose is to explore the relations between Prestige interests and HEXACO personality.

Based on the literature, it can be argued that Honesty-Humility is either negatively related to Prestige interests or that the two are unrelated. Honesty-Humility includes the facets Modesty and Greed avoidance, which seem opposite to the urge to compete and to strive for higher status and rewards associated with prestigious jobs. Some researchers (e.g., Marcus et al., 2007; Zettler & Hilbig, 2010) have found Honesty-Humility to be negatively related to unethical behavior. Unethical behavior was also found to be related to social class (Piff et al., 2012). In turn, Howard et al. (2011), and Lee and Rojewski (2009) have found (prestigious) socioeconomic class and Prestige interests/career aspirations to be related. Taken together, due to their inverse relation to status/class, the findings above indicate that Prestige interests and Honesty-Humility could be negatively related. However, other evidence seems to indicate that Honesty-Humility and Prestige interests may be unrelated. First, Francis (2012) has disputed the findings by Piff et al. (2012) on the basis of publication bias. Second, Honesty-Humility has been found to be strongly related to counterproductive behaviors, but Roberts, Harms, Caspi, and Moffitt (2007) showed that counterproductive work behaviors and occupational level are unrelated. In turn, occupational level has been found to be related to Prestige interests (Sodano & Tracey, 2008). Taking into account the counter-



arguments of Francis (2012) and the findings by Roberts et al. (2007), Honesty-Humility and Prestige interests may also be unrelated. In this study we explored the existence of a possible relation between the two.

We also expected Prestige interests to be positively related to Openness to Experience. Compared to a lower education, people with a higher education score higher on Openness to Experience (De Vries, Ashton, & Lee, 2009). Educational level is also positively related to Prestige interests (Sodano & Tracey, 2008). Therefore, the required skill and educational level for high Prestige activities is also likely to be related to intellectual curiosity and inquisitiveness which make up Openness to Experience. Consequently, in our study we expected Openness to Experience to have a positive relation with Prestige interests, even after correcting for educational level. That is, we expected people, who are curious and inquisitive (i.e., high Openness to Experience), to be more interested in complex jobs (i.e., have high Prestige interests) independent of their educational level.

#### *1.4 Purpose and research questions*

To summarize, the current study is framed in three overarching topics and for these topics we report results for normative and ipsatized interest scales. First, we investigate if the structure of Spherical representation of vocational interests also applies to our Dutch sample. Second, we compare vocational interests and personality relations in our sample to previous large meta-analytic studies (e.g., Barrick et al., 2003; Larson et al., 2002) and studies using the HEXACO (McKay & Tokar, 2012; Pozzebon et al., 2010). Third, we expand previous research by further investigating the relations between Prestige interests and the HEXACO model of personality.

## **2. Method**

Our study combined five samples. All samples used different additional measures besides the HEXACO inventory and the PGI. These measures are beyond the scope of the present study and are not reported; for more information, please contact the first author.

### *2.1 Participants*

The first sample was a heterogeneous group of people approached through social media by five undergraduate students ( $n = 182$ , ♀ = 58.2%,  $M(\text{age}) = 32.90$  years,  $SD = 14.57$ ). The second a group of professionals working for a higher level agency ( $n = 59$ , ♀ = 25.4%,  $M(\text{age}) = 39.08$  years,  $SD = 8.12$ ). The third sample came from a secondary school for middle

education ( $n = 16$ , ♀ = 93.8%,  $M(\text{age}) = 16.56$  years,  $SD = 1.09$ ). For the fourth sample we approached a large number of students who had previously filled out another vocational interest measure on an open-access website ( $n = 304$ , ♀ = 79.6%,  $M(\text{age}) = 20.56$  years,  $SD = 7.46$ ). The fifth sample was acquired from a heterogeneous group of older white and blue collar workers ( $n = 95$ , ♀ = 56.8%,  $M(\text{age}) = 57.35$  years,  $SD = 7.66$ ) through e-mail invitations and personal requests. Participation was always voluntary, across all samples we had approximately 60% non-response.

The total group ( $N = 656$ ) included more females than males (♀ = 65.9%,  $M(\text{age}) = 30.89$  years,  $SD = 16.18$ , ranging from 14 to 77). Most participants were indigenous Dutch (87.2%). We dummy-coded educational level into lower-middle (13.4%), middle-higher (33.8%), higher (49.2%), and unknown/other (3.5%). Participants were employed (36.0%), studying (36.1%), both employed as well as studying (17.1%), or neither (10.8%, temporary unemployed, retired etc.). The data were collected online via a questionnaire platform. In all samples, participants who completed all measures participated in a raffle, in which they could win various prizes.

## 2.2 Instruments

### 2.2.1 Personality measure

The HEXACO-PI-R (Dutch 100 item version; De Vries et al., 2009) consists of 100 statements measuring six personality dimensions, with 16 items per dimension and four additional items. These dimensions are: Honesty-Humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O). Each dimension consists out of four (four-item) facets that describe parts of the main dimension (e.g., Greed avoidance is a facet of Honesty-Humility). The HEXACO facets are not reported in this study; please contact the first author for these results. Responses are provided on a 5-point Likert scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree*. Reliabilities in earlier studies ranged from .75 for Conscientiousness to .84 for Honesty-Humility (De Vries et al., 2009). The reliabilities in the present study ranged from .78 for Agreeableness to .83 for Extraversion.

### 2.2.2 Vocational interest measure

The PGI (Tracey, 2002) contains three item formats: activity liking, activity competence, and occupation liking, divided over 108 activity statements and 108 occupational titles. The activity statements are rated

twice, once on liking using a 7-point likert scale (1 = *very strongly dislike*, 7 = *very strongly like*) and once on perceived competence using a 7-point likert scale (1 = *unable to do*, 7 = *very competent*). The occupational titles are rated once on liking using a 7-point likert scale (1 = *very strongly dislike*, 7 = *very strongly like*). Example activity items are 'Manage an office' and 'Guard buildings'. Example occupational titles are 'Bank Teller' and 'Sculptor'. The three item formats correlate highly (Tracey, 2002).

The PGI yields 27 scales of which 18 spherical interest scales, six RIASEC scales, and three Spherical axes. The 18 spherical interest scales (e.g., Business systems, Personal service) each exist of 18 items, with six items from each item format. These scales are then weighted and combined to construct the RIASEC scales and the three Spherical axes (Ideas/Data, People/Things, and Prestige). The PGI allows two representations of the equator; it can be either represented by the eight basic interest scales or by the six RIASEC scales. Tracey (2002) found reliabilities ranging from .69 to .92 for the spherical interest scales and .91 to .94 for the three main axes.

For the purpose of the current study, we translated the PGI to Dutch. First the authors, three graduate students, and a professional translator individually translated the English items to Dutch. All translations were then collected and discussed with all contributors, except for the translator. A final list was constructed based on consensus. Disagreements between contributors were resolved through discussion. Lastly, as a final check a second professional translator translated the Dutch items back to English. This final check provided no reasons to adjust the Dutch items. In the present study reliabilities were more than adequate, ranging from .86 to .95 for the basic interest scales, .91 to .96 for the RIASEC scales, and .91 to .96 for the three main axes. The axes reliabilities were calculated using the Feldt and Brennan (1989) formula for composite reliability.

### 2.3 Analyses

Before our main analyses we removed profile elevation from the interest scales as proposed by Tracey (2012). The mean interest score for each participant was computed (i.e., the mean profile elevation for that person) and subtracted from all individual scales, essentially centering/ipsatizing all spherical scales. Due to the computation behind the axes of the Spherical model (Tracey, 2002), the scores for the spherical axes do not differ when the spherical scales are ipsatized. In order to keep our results comparable to

earlier studies we performed most of our main analyses with normative and ipsatized scales.

We analyzed the data in several steps. To analyze our first research question about the structure of the Spherical representation of vocational interests, we chose to perform several circumplex analyses. We tested the fit of the eight basic interest scales (the PGI-octants) with two confirmatory tests to assess the circumplex structure. First we tested the model with a nonparametric randomization test of hypothesized order relations (Hubert & Arabie, 1987) with the program RANDALL (Tracey, 1997). A circumplex structure assumes that scales close to each other have stronger correlations than scales further away on the circumplex. The randomization test of hypothesized order relations tests how many of the correlations correctly decrease in size as the distance between scales increases. This test yields a Correspondence Index (CI) that reflects how many of the order predictions are met, ranging from -1.00 (all predictions violated) to +1.00 (all predictions met). Secondly, to further explore the PGI-octagon structure in our Dutch sample, we performed a covariance circumplex structure modeling analysis (Browne, 1992) with the R package CircE (Grassi, Luccio, & Di Blass, 2010). After testing the PGI-octagon structure, we also tested the fit of our data to the complete spherical model by testing the relations between all eighteen spherical interest scales with the randomization test of hypothesized order relations using the program RANDALL.

To answer our second research question we first looked at the relation between vocational interests and personality traits with simple correlations and multiple linear regression. In order to compare our findings to previous studies we then computed the *z*-values for the correlations in the present study. Thereafter, using Fisher's *r*-to-*z* transformation, we also calculated the weighted *z*-values for the correlations presented in McKay and Tokar (2012) and correlated the *z*-values of McKay and Tokar with our findings. We repeated the same procedure for the correlations found by Pozzebon et al. (2010). To compute the Realistic interest scale we averaged the correlation values on Producing and Adventuring interests.

Lastly, we performed univariate relative weight analyses (RWA's; Johnson, 2000) to specifically look into the relations of the distinguishing features (Prestige interests and Honesty-Humility) in both the interests and personality model to the other model. RWA is a form of regression analysis where the proportion of the shared explained variance and the unique

contribution is calculated for each predictor. The sum of the unique contribution and the proportion of the shared variance of each predictor is labeled the relative weight ( $\epsilon$ ). A RWA accounts for predictor inter-correlation and therefore calculates the contribution of the predictor by itself and in combination with other predictors more accurately than standardized regression coefficients or semipartial correlations do. We used three RWA's to further analyze the relations between Prestige interests and personality, and between Honesty-Humility and interests. Thus, we performed one RWA for the regression of Prestige interests on personality, and two for the regression of Honesty-Humility on the spherical interest scales, one for the normative and one for the ipsatized scales.

### 3. Results

#### 3.1 The structure of the Spherical representation of vocational interests in a Dutch sample

First we investigated the structure of the Spherical representation of vocational interests in our Dutch sample. All the results for the structure analyses are shown in Table 1. The randomization test of hypothesized order relations for the PGI-octant scales in the current sample showed a CI of .882 ( $p < .001$ ). When the PGI-octants structure was tested with the ipsatized scales the CI was .938 ( $p < .001$ ). In both cases the CI indicated an excellent fit for the PGI-octants.

**Table 1.**

*Tests for circumplex structure of the PGI-octants and -sphere.*

	Normative PGI-octants	Ipsatized PGI-octants	PGI-sphere
<i>Randomization test of hypothesized order relations</i>			
Predictions made	288	288	9472
Predictions met	271	279	7341
CI	.882	.938	.552
<i>P</i>	.000	.000	.001
<i>Covariance structure modeling circumplex analysis</i>			
RMSEA	.171	.204	-
CFI	.936	.889	-
SRMR	.070	.046	-

*Note.*  $N = 656$

The structure of the PGI-octants was also tested with covariance structure modeling proposed by Browne (1992) with an unconstrained model.

For the current sample, the circumplex estimation of this analysis showed good (SRMR = .070, CFI = .936) to poor (RSMEA = .171) fit, depending on the fit indices. The circumplex fit of the ipsatized data was also analyzed with an unconstrained model and showed good (SRMR = .046), to moderate (CFI = .889), to poor fit (RSMEA = .204). Overall, we found an acceptable to excellent fit for the PGI-octant scales.

Lastly, the randomization test of hypothesized order relations revealed that the data of the sample also fitted the spherical model of the PGI reasonably ( $CI = .552, p < .001$ ). This fit is comparable to the fit that almost all earlier studies found for the Spherical representation (e.g., Darcy, 2005; Tracey, 2002). Based on all of the fit indices together we concluded that the fit of the PGI-octants and -sphere was adequate and proceeded to analyze the relations between interests and personality.

### 3.2 *The relations between vocational interests and personality traits*

The second topic we analyzed was the congruence between our findings and the findings of previous studies. The means, standard deviations, and reliabilities of all interest scales and personality dimensions are shown in Table 2. Our sample means indicate that our participants showed a higher interest in People over Things and somewhat higher Prestige interests. These higher means are in line with the fact that this sample consists mostly of people working/studying in social domains and with a higher education. Moreover, Prestige interests were significantly higher for participants with a higher education ( $F(2,630) = 48.09, p < .01, \eta^2 = .13$ ). Correlations between interests and personality are shown in Table 3. Gender related strongly to several vocational interest scales and to the personality dimension Emotionality. Vocational interests are known to show pronounced gender differences (e.g., Larson et al., 2002; McKay & Tokar, 2012). For the HEXACO, the personality dimension Emotionality has been known to display the largest gender difference in previous research (De Vries et al., 2009).

The correlations between vocational interests and the personality dimensions in Table 3 show several relations that have also previously been found in other studies (Barrick et al., 2003; Larson et al., 2002). The relations in the present study by-and-large replicate previous meta-analytic findings. However, note that the personality dimension Emotionality showed several significant relations to People and Things related interest scales that do not concur with meta-analytic findings. Additionally, Agreeableness did not have a significant relation with People related interests.

**Table 2.***Means, standard deviations and reliabilities for the PGI and the HEXACO-PI-R.*

Spherical Interest	$M_{normative}$	$M_{ipsatized}$	$SD_{normative}$	$SD_{ipsatized}$	Reliability	Interest axis	$M$	$SD$	Reliability
Social facilitating	3.99	0.85	0.94	0.76	.88	People vs Things	3.21	3.29	.96
Managing	3.62	0.47	1.05	0.78	.91	Ideas vs Data	0.18	2.76	.95
Business detail	3.02	-0.13	1.24	0.96	.95	Prestige	1.39	1.92	.91
Data processing	2.58	-0.56	1.17	0.88	.94	Profile elevation	3.14	0.74	
Mechanical	2.57	-0.57	1.21	0.90	.94	Personality			
Nature/Outdoors	3.27	0.13	1.15	0.83	.91	Honesty-Humility	3.57	0.55	.82
Artistic	3.42	0.28	1.33	1.18	.94	Emotionality	3.11	0.54	.80
Helping	4.33	1.19	1.10	1.01	.91	Extraversion	3.30	0.55	.83
Social Sciences	3.79	0.65	1.31	1.11	.94	Agreeableness	3.11	0.48	.78
Influence	3.45	0.31	1.08	0.64	.89	Conscientiousness	3.50	0.51	.80
Business Systems	2.68	-0.47	1.15	0.76	.93	Openness to experience	3.22	0.57	.79
Quality Control	2.56	-0.59	1.04	0.64	.93				
Manual work	2.57	-0.58	0.98	0.67	.90				
Personal service	3.15	0.00	1.03	0.88	.88				
Financial analysis	2.95	-0.20	1.30	0.94	.95				
Science	3.19	0.04	1.25	0.88	.93				
Construction repair	2.29	-0.86	1.13	0.81	.94				
Basic service	3.18	0.04	0.88	0.68	.86				

*Note.*  $N = 656$ , Gender male = 1 and female = 2. Reliability shown for interest scales is for normative scales. Normative scores for the spherical interest scales ranged from 1 to 7, for personality from 1 to 5.

**Table 3.***Correlations between the PGI and the HEXACO-PI-R.*

	Gender		Age		Honesty-Humility		Emotionality		Extraversion		Agreeableness		Conscientiousness		Openness		$R^2$ (personality)	
Gender	-	-	-	-	.23	.43	-.05	-.04	.16	-.09	.22							
Age	-.29	-	-	-	.08	-.20	.13	.06	.00	-.03	.06							
Social facilitating	.22	<b>.50</b>	-.09	-.06	-.16	.06	.14	<b>.32</b>	<b>.33</b>	<b>.30</b>	.00	-.06	.06	.10	.18	-.07	.21	.25
Managing	-.05	.16	-.08	-.06	-.26	-.10	-.06	.06	.22	.19	-.05	-.13	.06	.12	.06	-.20	.13	.14
Business detail	-.25	-.14	.04	.09	-.29	-.16	-.16	-.10	.02	-.05	-.02	-.08	.06	.10	.03	-.20	.13	.11
Data processing	<b>-.55</b>	<b>-.53</b>	.03	.09	<b>-.30</b>	-.18	<b>-.31</b>	<b>-.30</b>	-.10	-.22	.06	.02	-.11	-.12	.18	-.01	.24	.19
Mechanical	<b>-.53</b>	<b>-.52</b>	.06	.13	-.22	-.08	<b>-.36</b>	<b>-.37</b>	-.04	-.15	.09	.07	-.07	-.06	.16	-.03	.21	.19
Nature/Outdoors	-.21	-.08	-.04	.00	-.14	.04	-.18	-.12	-.01	-.11	.05	.01	-.05	-.04	<b>.38</b>	.26	.20	.12
Artistic	.04	.20	-.09	-.07	-.06	.10	.09	.19	.07	.01	.00	-.04	-.09	-.07	<b>.59</b>	<b>.48</b>	<b>.38</b>	<b>.31</b>
Helping	<b>.32</b>	<b>.53</b>	-.08	-.05	.10	<b>.30</b>	.26	<b>.39</b>	.23	.17	.11	.08	.08	.11	.19	-.01	.20	.29
Social Sciences	.22	.42	-.07	-.05	.07	.26	.18	<b>.30</b>	.17	.13	.11	.09	.02	.05	.17	.00	.12	.19
Influence	-.14	.04	-.11	-.12	-.21	-.05	-.08	.03	.20	.21	.03	-.03	-.01	.02	<b>.38</b>	.28	.21	.12
Business Systems	<b>-.37</b>	<b>-.33</b>	-.06	-.05	<b>-.34</b>	-.26	-.23	-.21	.02	-.08	.01	-.05	-.06	-.05	.11	-.13	.17	.14
Quality Control	<b>-.48</b>	<b>-.50</b>	.09	.21	-.24	-.08	<b>-.30</b>	<b>-.32</b>	-.03	-.17	.09	.07	-.10	-.12	.13	-.13	.17	.18
Manual work	<b>-.44</b>	<b>-.37</b>	.15	.28	-.18	.02	-.25	-.21	.02	-.09	.11	.09	-.10	-.11	.10	-.18	.11	.10
Personal service	<b>.31</b>	<b>.57</b>	-.20	-.19	-.09	.12	.18	.34	.10	.02	-.01	-.07	.01	.05	.15	-.08	.09	.14
Financial analysis	-.28	-.19	.01	.06	<b>-.33</b>	-.25	-.18	-.14	.07	.01	-.02	-.07	.00	.03	.10	-.11	.16	.11
Science	-.14	.01	-.17	-.19	-.14	.02	-.12	-.06	-.01	-.11	.10	.08	-.01	.02	<b>.32</b>	.20	.15	.08
Construction repair	<b>-.53</b>	<b>-.53</b>	.10	.19	-.22	-.07	<b>-.34</b>	<b>-.34</b>	.01	-.09	.08	.05	-.11	-.12	.14	-.08	.17	.15
Basic service	.12	<b>.42</b>	-.12	-.10	-.10	.17	.09	.27	.10	.00	.03	-.03	.00	.04	.10	-.20	.05	.14



**Table 3 continued.**

<i>R</i> <sup>2</sup> (spherical interest scales)	<b>.51</b>	<b>.50</b>	.23	.24	.21	.17	.24	.23	.22	.21	.08	.07	.09	.09	<b>.43</b>	<b>.37</b>	-
People (hi) vs Things (lo)	<b>.59</b>		-.11		.18		<b>.40</b>		.24		-.02		.09		.04		.29
Ideas (hi) vs Data (lo)	.08		-.04		.17		.07		-.06		.07		-.09		<b>.38</b>		.23
Prestige	.14		-.18		-.07		.08		.13		-.03		.08		.22		.08
<i>R</i> <sup>2</sup> (PGI-axes)	<b>.33</b>		.04		.07		.16		.07		.01		.02		.21		-
Profile elevation	-.24		-.05		-.26		-.14		.11		.07		-.04		<b>.30</b>		.18

Note. *N* = 656, Correlations above .07 are significant at a *p* < .05, correlations above .10 are significant at a *p* < .01, correlations above .15 are significant at a *p* < .0001 (*p* value as per Bonferroni correction for the PGI-octants and the spherical axes), correlations above .30 are bolded. For each top variable the first column shows the correlations for normative interest scores, the second column shows the correlations for the ipsatized interest scores. Gender: male = 1 and female = 2. *R*<sup>2</sup> coefficients for gender are Cox and Snell pseudo *R*<sup>2</sup> values.

We compared our findings to previous studies that investigated the relations between RIASEC interests and HEXACO personality and found a very strong similarity between our findings and theirs (McKay & Tokar, 2012:  $r = .80$ ; Pozzebon et al., 2010:  $r = .84$ ). Additionally, we found a weaker correlation to previous (non-ipsatized) studies for our ipsatized scales (McKay & Tokar, 2012:  $r = .69$ ; Pozzebon et al., 2010:  $r = .66$ ). Using a test for differences between dependent correlations (Steiger, 1980), we found that the correlations based on the ipsatized scales were significantly lower than those based on the normative scales (McKay & Tokar, 2012:  $Z = 1.69$ ,  $p < .10$ ; Pozzebon et al., 2010:  $Z = 2.66$ ,  $p < .05$ ).

The relations between Agreeableness and Conscientiousness and interests were small at best, for normative and ipsatized scales alike, the other four personality traits showed several medium-sized relations. Openness to Experience ( $r = .30$ ) and Honesty-Humility ( $r = -.26$ ) were the personality dimensions that were most strongly related to profile elevation. The overall relation to vocational interests for these traits decreased for the ipsatized scales, because some of the shared variance with interests was removed with the removal of the profile elevation from the scale scores. The relations between personality traits Emotionality and vocational interests increased when the scales were ipsatized for Social Facilitating and Helping, two adjacent vocational interest axes that represent the interest in People. The relations between personality traits Extraversion and Data processing interests, Mechanical interests, and Investigative interests (representing interest in Things) became somewhat more pronounced for the ipsatized interest scales.

A RWA for the PGI-octant (Table 4) on Honesty-Humility revealed that the scales on the people and data side of the circumplex were mainly associated with Honesty-Humility. However, when the ipsatized octant scales were entered in the RWA most variance (56%,  $\epsilon = .070$ ) in the relations between vocational interests and Honesty-Humility could be attributed to the Helping scale. That is, the relations of the interest scales with Honesty-Humility changed meaningfully when the variance associated with the general factor of interests was removed.

### *3.3 Prestige interests and the HEXACO personality model*

The correlations in Table 3 revealed significant but weak relations between the Prestige interests and Emotionality, Extraversion, and Openness to Experience ( $r = .08 / .13 / .22$ ). Only the simple correlation with Openness to

Experience remained significant when a Bonferroni correction was applied. Additionally, this correlation between Prestige interests and Openness to Experience remained significant when controlled for Educational level ( $r = .16$ ,  $p < .01$ ). In line with the correlations, a RWA (Table 4) revealed that Openness to Experience explained most variance (56%;  $\epsilon = .045$ ) of all personality dimensions in Prestige interests. These findings were partly in line with our expectations: Prestige interests were significantly and positively related to Openness to Experience, but not significantly and negatively to Honesty-Humility ( $r = -.07$ ,  $p = .09$ ).

**Table 4.**

*Relative weight analyses for interests on Honesty-Humility and personality on interest in Prestige.*

PGI-octant scale	Honesty-Humility				HEXACO dimension	Prestige interests	
	normative		ipsatized			<i>beta</i>	$\epsilon$
	<i>beta</i>	$\epsilon$	<i>beta</i>	$\epsilon$			
Social facilitating	-.301	.025	-.147	.008	Honesty-Humility	-.099	.006
Managing	.001	.023	.011	.004	Emotionality	.121	.010
Business detail	-.095	.028	-.016	.009	Extraversion	.101	.013
Data processing	-.331	.040	-.162	.020	Agreeableness	-.004	.001
Mechanical	.188	.015	.210	.010	Conscientiousness	.072	.005
Nature/Outdoors	-.050	.007	-.005	.001	Openness to Experience	.210	.045
Artistic	-.062	.005	-.012	.003		-	-
Helping	.317	.030	.408	.070		-	-
$R^2$	.173		.125			.081	

*Note.*  $N = 656$ , *beta* denotes the standardized beta weights,  $\epsilon$  denotes the relative weights (Johnson, 2000).

## 4. Discussion

### 4.1 The structure of the Spherical representation of vocational interests in a Dutch sample

The first contribution of the current study is the validation of the Spherical representation of vocational interests in a Dutch sample. This sample consists of both working, studying, and unemployed/retired participants, allowing conclusions not merely limited to a student population. All fit measures taken together, our results show an adequate fit for the structure of the Spherical representation in our sample for both the PGI-octant scales and the complete sphere. However, the PGI-octants structure fit

differed depending on the analysis used: the CI values of the randomization test of hypothesized order relations (a nonparametric test) indicated a good fit, but the RMSEA values of the covariance circumplex structure modeling analysis (a SEM test) indicated a poor to adequate fit. Previous studies found similar differences between these tests (Darcy & Tracey, 2007; Gupta, Tracey, & Gore, 2008). According to Darcy and Tracey (2007), different fit results may be attributed to the different comparisons the analyses make. The nonparametric analysis tests if the structure found in the sample fits better than a random structure. The SEM analysis tests if the structure deviates from a perfect (circular) model. Although the empirical structure of the spherical representation – in our sample – fits better than a random representation, it also deviates from a perfect fit. The PGI-octant scales appear correctly positioned relative to each other, but the scale vectors do not show a perfect octagon structure.

#### *4.2 The relations between vocational interests and personality traits*

The second contribution of the present study is the comparison between the Spherical representation of vocational interests and the HEXACO model. Our results strongly converge with the results from previous studies (McKay & Tokar, 2012, Pozzebon et al., 2010). The two most noticeable differences were as follows. First, compared to the Big Five-Interests relations (Barrick, Mount, & Gupta, 2003; Larson, Rottinghaus, & Borgen, 2002), the HEXACO-interests relations were different for Emotionality and Agreeableness. For the HEXACO model Emotionality – but not Agreeableness – relates to people interests. It is most likely that these differences are caused by the changed rotation of these two personality dimension compared to the Big Five dimensions. These results once again imply that researchers should be careful when comparing results of Big Five and HEXACO Agreeableness and Emotionality/Emotional Stability, because these do not readily translate into each other (cf. Ashton, Lee, & De Vries, 2014). Second, our results also showed significantly less congruence with previous studies when using ipsatized scales instead of normative scale scales.

The different results for Honesty-Humility for normative and ipsatized scales further illustrate the effects of profile elevation on the relations between personality and interests. Our results were initially similar to those of McKay and Tokar (2012) and Pozzebon et al. (2010) for the normative scales. The normative interest scales Helping and Social Facilitating (Social interests) were positively related to Honesty-Humility and Managing and Business detail interests (Enterprising interests) were (weakly) negatively related to Honesty-

Humility. However, when the interest scales were corrected for profile elevation we found that ipsatized Helping interests were most strongly related to Honesty-Humility; 56% of the shared variance between interests and Honesty-Humility could be attributed to just the relation with Helping interests.

In agreement with Tracey (2012) it may be that the purest interests are more adequately reflected using the ipsatized scales, because the general factor of interests adds error variance – a bias – to the measurement of specific interests. On the one hand, we therefore might argue that mainly Helping interests are related to Honesty-Humility and not the scales related to Enterprising interests. On the other hand, the error variance caused by profile elevation also appears to have substance because it shows a moderate positive relation with Openness and a slightly weaker negative relation with Honesty-Humility. When studying relations for specific interests the general factor adds variance that may not be attributed to that specific interest, but this variance is nonetheless substantive. Additionally, the negative relationship between Honesty-Humility and profile elevation advances our knowledge about the substance of profile elevation. So far, profile elevation has only been related to the personality dimensions Openness and Extraversion (Fuller et al., 1999). According to our findings, Openness and Honesty-Humility are the main personality correlates of profile elevation.

#### *4.3 Prestige interests and the HEXACO personality model*

We did not find a relation between Prestige interests and Honesty-Humility. People who like jobs and tasks that require a lot of skill and effort are therefore not more interested in materialistic gains, or less fair, sincere, and modest. On the other hand, Prestige interests did relate to Extraversion and Openness to Experience, even when Openness to Experience was controlled for educational level. This indicates that Prestige interests are related to imagination, curiosity and social energy. Nevertheless both these relations were not very strong.

Because Prestige interests are related to educational level and social economic status the largest limitation of the current study is that most participants were highly educated. Our sample thus mostly represents the upper half of Prestige interests, possibly resulting in a selection bias. A consequence of this selection bias may be that our results are an underestimation of the actual relations between Prestige interests and personality for the general (Dutch) population. Future studies should pay

attention to the educational level of the sample when measuring Prestige interests, in order not to underestimate effects.

#### *4.4 Summary*

In summary, the current study investigated the relations between the Spherical representation of vocational interests and the HEXACO model. We confirmed the structure of the spherical representation in a Dutch sample. Compared to previous studies the relations between interests and personality in our study are very congruent for normative interest scales and a little less congruent for ipsatized interest scales. This difference in congruence with previous studies leads back to the somewhat different interests-personality relations for normative interest scales compared to ipsatized interest scales. The main difference of the HEXACO to existing personality measures is the addition of the Honesty-Humility dimension. We found that Honesty-Humility related significantly and positively to Helping interests. The main difference between the PGI and existing interest measures is the addition of the Prestige interests axis. Results showed that Prestige interests relate significantly to Openness to Experience, but not to Honesty-Humility.

**Perceptions of vocational interests:  
Self-other agreement, assumed similarity,  
similarity, and reciprocity in  
student-parent dyads**



## **Abstract**

# **Perceptions of vocational interests: Self-other agreement, assumed similarity, similarity, and reciprocity in student-parent dyads**

The current study investigated how self- and other-ratings of vocational interests converge among student-parent dyads. Using the Personal Globe Inventory-Short (Tracey, 2010), we obtained data from 271 student-parent dyads across three samples. The students were high school seniors and university students. All participants rated their own vocational interests and those of the other dyad member. First, profile correlations revealed high levels of self-other agreement, moderate levels of assumed similarity, and low levels of similarity and reciprocity in vocational interests. These correlations are highly similar to those found in the field of personality. Second, profile elevation showed a reversed pattern compared to interest perceptions, with high levels of self-other agreement and moderate levels of assumed similarity, indicating that it may mostly be an artifact/rater bias and not a substantive factor. Ipsatization of the vocational interest scales seemed to reduce profile elevation bias. Third, same-gender dyads overestimated their similarity in vocational interests more than different-gender dyads.

---

Grateful acknowledgement is provided to Sandra Kok, Mirte Post, and Natasja Overman for their assistance in collecting data.



## 1. Introduction

Making the right vocational choice straight away is becoming more important. The societal and personal costs of student drop-out are high and government policies actively focus on the reduction of drop-out by imposing laws. For example, the Dutch government has implemented a law that forces institutes of higher education to provide students with a “study choice check”. Even before the implementation of this law more than 54% of all students use a vocational interest questionnaire to help them in choosing a study program (Markteffect, 2011). While vocational interests are predictive of performance (Tracey, Allen, & Robbins, 2012) and have incremental predictive validity over other common individual differences predictors (i.e., intelligence and personality; Van Iddekinge, Putka, & Campbell, 2011), about 55% of the students still drop out because of a “wrong study choice” (Van den Broek, Wartenbergh, Brink, Kurver, Hampsink, & Braam, 2013). Personality research shows that several other-reports have a greater predictive validity for performance than self-rated personality (Connelly & Ones, 2010). Other-rated vocational interests may contribute to predicting a student’s match with their education. Based on personality research it can even be argued that other-rating could offer more predictive validity than self-rating. However, nearly all studies on vocational interests have used self-reports, and only a few (Nauta, 2012; Nelling, Kandler, & Riemann, 2015) have investigated other-reports. Before we use other-ratings of vocational interests it is important to first understand their nature.

In this study we will answer three fundamental questions about self- and other-reports of vocational interests in a dyadic (student-parent) setting. First, how useful are other-reports of vocational interests? Using self- and other-ratings, we investigate the strength of four dyadic vocational interest perceptions, namely self-other agreement (do a person’s self-ratings converge with ratings by others?), assumed similarity (do a person’s self-ratings converge with how s/he rates others?), similarity (do self-ratings of two persons converge?), and reciprocity (do the other-ratings of two persons in a dyad converge?). Second, should vocational interests be ipsatized to adjust for profile elevation? We investigate how different perceptions of regular vocational interests and perceptions profile elevation are. Third, does it matter if the observer is of the same or of a different gender? Investigating the effect of gender similarity between the observer and target on perceptions of vocational interests may indicate whether same- or different-gender dyads are better suited to evaluate each other’s vocational interests.

### 1.1 Self- and other-rated vocational interests

In his Social Relationships Model (SRM), Kenny (1994) described how self- and other-ratings can be combined to form several aspects of interpersonal perception. Within one dyad, self- and other-ratings can yield four different aspects of interpersonal perception, as shown in Table 1. For instance, imagine a dyad consisting of Alexa (A) and Bob (B). Both Alexa and Bob provide self-ratings on vocational interests, which is shown in Table 1 as A(A) and B(B). Alexa provides other-ratings on the interests of Bob and vice versa, which is shown as A(B) and B(A). The self- and other-ratings of Alexa and Bob intersect in six unique ways that can be categorized using four aspects of interpersonal perception (two aspects appear twice).

**Table 1.**

*Types of interpersonal perceptions at a dyadic level with self- and other-ratings.*

Rater	Target	Notation	Person A		Person B	
			Person A	Person B	Person B	Person A
			= A(A)	= A(B)	= B(B)	= B(A)
Person A	Person A	= A(A)	-			
	Person B	= A(B)	Assumed similarity $r_{A(A),A(B)}^\dagger$	-		
Person B	Person B	= B(B)	Similarity $r_{A(A),B(B)}$	Self-other agreement $r_{A(B),B(B)}$	-	
	Person A	= B(A)	Self-other agreement $r_{A(A),B(A)}$	Reciprocity $r_{A(B),B(A)}$	Assumed similarity $r_{B(B),B(A)}$	-

<sup>†</sup> $r_{A(A),A(B)}$  refers to the correlation between person A's self-ratings of vocational interests [A(A)] and person A's other-ratings about the interests of person B [A(B)].

The —probably— most frequently used aspect of interpersonal perception is self-other agreement. Self-other agreement ( $r_{A(A),B(A)}$  or  $r_{B(B),A(B)}$ ) represents the correlation between the self-ratings of one person and the other-ratings by another person, who judges the characteristics (e.g., vocational interests) of the first person. If the target's self-ratings and the observer's other-ratings correlate highly, then the self-other agreement is high. So far, to our knowledge, dyadic research on vocational interests has been limited to self-other agreement (Nauta, 2012; Nelling et al., 2015). However, there are three other aspects of interpersonal perception. First, if a person rates their own characteristics and the characteristics of another person, the correlation between these two ratings is called assumed similarity. Assumed

similarity ( $r_{A(A),A(B)}$  or  $r_{B(B),B(A)}$ ) represents the degree to which a person sees herself as she sees the other. Second, the correlation between self-ratings of one person with the self-ratings of another is called similarity ( $r_{A(A),B(B)}$ ). Sometimes this is referred to as actual similarity to set it apart from assumed similarity. Last, the relation between two people's other-ratings of each other is called reciprocity ( $r_{A(B),B(A)}$ ). Reciprocity is the correlation between two other-ratings of two individuals judging each other. Reciprocity strongly resembles similarity, but uses other-ratings instead of self-ratings to establish the similarity within a dyad.

In sharp contrast to the small number of studies that have looked at other-ratings of vocational interests, a large number of studies have looked at other-ratings of personality. For example, in personality research other-ratings have been used to establish how rater tendencies influence scale scores (Zettler, Lang, Hülshager, & Hilbig, 2015), to establish how stable personality is over time (Costa, McCrae, & Dye, 1991), and in attempts to solve the person-situation debate (Kenny, 2004). Expanding the SRM, Kenny (2004) describes in his PERSON model how an observer may make a more accurate judgment of a target's personality when they make more observations. More observations allow an observer to more accurately rate a target's personality and reduce error (E), the residual of personal stereotypes (R), the influence of stereotypes that are shared by all observers (S), and the influence of norms (N) on judgments. The accurate judgment of a target's personality partly consists of the target's true personality (P; the part that all observers would agree upon), but also partly of the unique opinion (O) that the observer has on the target's personality.

Personality research (e.g., De Vries, 2010; McCann, Lipnevich, Poropat, Wiemers, & Roberts, 2015; Watson & Clark, 1991) has provided empirical evidence for strong self-other agreement relations, moderate assumed similarity relations, and mostly insignificant or small similarity and reciprocity relations. A large body of research (e.g., Barrick, Mount, & Gupta, 2003; Holtrop, Born, & De Vries, 2015; Larson, Rottinghaus, & Borgen, 2002; McKay & Tokar, 2012) has shown significant relations between self-ratings of interests and self-rated personality. Additionally, Nelling et al. (2015) found a high level of self-other agreement in vocational interests ( $\bar{r} = .59$ ), as did Nauta (2012) across two separate samples ( $\bar{r} = .46$  and  $.53$ ). Moreover, when Nauta compared self-other agreement in vocational interests to self-other agreement in personality, she found comparable levels of self-other agreement in vocational interests to personality ( $\bar{r} = .48$  and  $.57$ ). Based on the significant

relation between self-ratings of interests and personality, and the comparable levels of self-other agreement, it can be expected that other aspects of interpersonal perception will show similar magnitudes in vocational interests as found in personality. Thus, we expected to find strong levels of self-other agreement, moderate levels of assumed similarity, and weak levels of similarity and reciprocity.

*Hypothesis 1:* Self-other agreement in vocational interests will be positive; i.e., there is a positive correlation between a target's self-ratings [e.g., A(A)] and other-ratings [e.g., B(A)].

*Hypothesis 2:* Assumed similarity in vocational interests will be positive; i.e., there is a positive correlation between a rater's self-ratings [e.g., A(A)] and other-ratings [e.g., A(B)]. However, assumed similarity will be weaker than self-other agreement in vocational interests.

*Hypothesis 3:* Similarity and reciprocity in vocational interests will be positive; i.e., there is a positive correlation between two targets' self-ratings [similarity; A(A) and B(B)] and two targets' other-ratings [reciprocity; A(B) and B(A)]. However, both similarity and reciprocity will be weaker than self-other agreement and assumed similarity in vocational interests.

The present study not only extends previous research on self- and other-ratings of vocational interests by looking at aspects of interest perceptions, but also by using the spherical model of vocational interests (Tracey & Rounds, 1996). The spherical model of vocational interests is an extension on Holland's (1959) structure of vocational interests. Holland proposed that vocational interests can be captured in six main themes: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. These so-called RIASEC themes represent six, evenly distributed, vectors on a circumplex. Prediger (1982) proposed that two right-angled axes underlie this RIASEC circumplex, namely Ideas versus Data and People versus Things. In addition to the two axes previously proposed by Prediger (1982), Tracey and Rounds (1996) proposed a third axis of vocational interests, namely Prestige interests. Prestige interests describe an interest in complex tasks (versus simple tasks) and have been related to educational level, required skill and effort for the activity, and competition involved in the activity (Sodano & Tracey, 2008). Tracey and Rounds proposed that the Prestige axis is orthogonal to Prediger's Ideas versus Data and People versus Things axes, thus effectively changing the two-dimensional circular representation of vocational interests into a three-dimensional spherical representation, with the original circular representation

on the equator. Tracey and Rounds (1996) called this revised model the spherical representation of vocational interests. The spherical structure has been confirmed in the USA (Tracey, 2002) and in several samples outside the USA (e.g., Irish, Dutch, Chinese, and Caribbean; Darcy, 2005; Holtrop et al., 2015; Long, Adams, & Tracey, 2005; Wilkins, Ramkisson, & Tracey, 2013).

The prestige axis is not the only difference between the spherical representation and the RIASEC model. In the spherical model, the circumplex on the equator is not represented by the six RIASEC themes, but instead by eight 'basic' interest scales: Social facilitating, Managing, Business Detail, Data Processing, Mechanical, Nature/Outdoors, Artistic, and Helping. Tracey and Rounds (1995) showed that the hexagonal and octagonal representations cover the same space on the circumplex of vocational interests and are psychometrically equally valid representations of this circumplex. However, they proposed that an octagonal representation may be slightly preferable, because practically, eight scales allow for a slightly more differentiated formulation of interests that may be more readily understood. The current study measures vocational interests using the Personal Globe Inventory-Short (PGI-Short; Tracey, 2010), which measures the spherical representation including its octagonal equator.

### *1.2 Profile elevation in self- and other-rated vocational interests*

The average score on all scales of an inventory is often called profile elevation. Profile elevation explains large amounts of shared variance in vocational interest scales (Tracey, 2012). Researchers have interpreted profile elevation in two ways: as substance or as artifact. The 'substantive' interpretation of profile elevation (e.g., Fuller, Holland, & Johnston, 1999; Hirschi & Läge, 2007) suggests that profile elevation is a meaningful factor. As a substantive factor, profile elevation has been found to be related to career planning, career exploration (Hirschi & Läge, 2007), and personality traits (Fuller et al., 1999; Holtrop et al., 2015). The 'artifact' interpretation (e.g., Prediger, 1998; Tracey, 2012) suggests that profile elevation is systematic error or rater bias and that its influence should be minimized when measuring vocational interests. Tracey (2012) suggested to minimize the influence of profile elevation by ipsatizing vocational interest scale scores. The thought behind this approach is that profile elevation is part of any normative interest scale and that any relation to another variable is also (possibly) due to profile elevation and not only that particular vocational interest. To ipsatize the interest scale scores a person's mean score on all interest scales is subtracted from each individual scale score. However, profile elevation cannot be

computed for single interest scales (e.g., a scale that only measures interest in STEM areas). Therefore, only broad vocational interest inventories, that measure the whole range of vocational interests, can be ipsatized.

If profile elevation is indeed a substantive factor then it may show the same pattern of interest perceptions as regular interest scales, as described in Hypotheses 1-3. If profile elevation is a rater bias then it may show non-trait-like interest perceptions relations. So far, Nelling et al. (2015) found that profile elevation shows moderate levels of self-other agreement ( $r = .34$ ), in contrast to high levels of self-other agreement for regular interest scales ( $\bar{r} = .59$ ). If profile elevation shows higher levels of assumed similarity than regular interest scales, this would imply that profile elevation is more influenced by the observer's rating tendencies than regular interest scales are, and thus that profile elevation is a bias. Building on the findings of Nelling et al., we expect profile elevation to show moderate levels of self-other agreement, high levels of assumed similarity, and low levels of similarity and reciprocity, indicative of the artifact interpretation.

*Hypothesis 4:* Self-other agreement in profile elevation will be positive and weaker than assumed similarity in profile elevation.

### 1.3 Gender influences on self- and other-rated vocational interests

Of all individual differences, vocational interests arguably show the largest gender differences. For example, Su, Rounds, and Armstrong (2009) found a large gender effect for People versus Things interests ( $d = 0.93$ ), with women showing more interest in People and men in Things. In her circumscription and compromise theory, Gottfredson (1981) describes how people develop vocational aspirations that are in line with their own gender at an early age (from the ages six to eight). Therefore, members of same-gender dyads should—on average—show more strongly overlapping interests than members of different-gender dyads. It may be expected that these pronounced gender mean differences affect the correlations between self- and other-reports of vocational interests and thus influence the perception of vocational interests, specifically for assumed similarity, similarity, and reciprocity, but not for self-other agreement.

In the Realistic Accuracy Model (RAM), Funder, Kolar, and Blackman (1995) described how self-other agreement (or accuracy) varies under four conditions. First, a *good judge* (e.g., trained rater) improves the accuracy of ratings. Second, a *good target* makes it easier for judges to accurately rate his/her characteristics. Third, a *good trait* makes judgments more accurate, as

some traits are more visible than others. For example, some research (e.g., Funder & Colvin, 1988) showed that Extraversion is a very visible trait. Fourth, *good information* should improve the quality of other-ratings, the quality of information generally improves with acquaintanceship. For vocational interests, Nauta (2012) found some evidence that interest in Things showed higher levels of self-other agreement than interest in People and she offered as explanation that interest in Things is more visible. Nauta also found that self-other agreement improved with acquaintanceship. For the current study however, none of these known moderators of self-other agreement describes how self-other agreement may be directly influenced by the gender of the observer, the target, or an interaction thereof. Thus, there is no reason to expect that self-other agreement would be influenced by the gender of a dyad's members.

For some personality traits (e.g., Openness to experience), assumed similarity showed higher levels of actual similarity between well acquainted people, such as friends and (dating) couples (Lee et al., 2009; Watson, Hubbard, & Wiese, 2000); that is, close acquaintances sometimes overestimate their similarity to the other. According to Lee et al., this overestimation might imply that people like to think that, in characteristics that represent their personal values, their close acquaintances are more similar to them than they actually are. As interests may also represent such values, participants in student-parent dyads may be motivated to overestimate their similarity to the other. Specifically, because adolescents may model their behavior more after their same-gender parent (e.g., Dryler, 1998) and because they identify themselves more with the same-gender parent (Starrels, 1994), they may be even more motivated to overestimate their similarity to the same-gender parent. Lastly, due to the aforementioned gender effects in vocational interests, the interests of two people with the same gender are on the average more similar than interests of two people with a different gender. Therefore same-gender dyads are expected to share more interests than different-gender dyads, which should result in higher similarity and reciprocity in same-gender dyads.

*Hypothesis 5a.* Assumed similarity in vocational interests is stronger than (actual) similarity.

*Hypothesis 5b.* Assumed similarity in vocational interests is stronger for same-gender dyads than for different-gender dyads, even when controlled for (actual) similarity.

*Hypothesis 6.* Similarity and reciprocity in vocational interests are stronger for same-gender dyads than for different-gender dyads

## **2. Method**

### *2.1 Participants and procedure*

The present study combines three Dutch samples of self- and other-ratings of vocational interests: 1) a high school student sample, 2) a small heterogeneous snowball sample, and 3) a university student sample. Note that the phrase “students” in this manuscript refers to sons/daughters who are (high school) students, all of who were close to making their first major vocational decision (i.e., choosing a study programme) or recently made that decision.

In all samples, the students were first invited to participate and after their participation their parent was invited. All samples used different additional measures besides the PGI-Short (Tracey, 2010). Because these measures are beyond the scope of the present study, they are not reported. For more information, please contact the first author. The procedure to collect the student-ratings was slightly different between the samples, as described below. The subsequent procedure to collect the parent-ratings was similar across all samples.

For the first sample, Dutch high school students were approached in the last years of their high school. Since Dutch high schools have different durations depending on their educational level, fourth through sixth year students were approached. Beforehand, parents were informed that the research was taking place. Then, approximately 1200 students across six high schools, in four cities, were approached for voluntary participation via email and classroom presentations. The students were first asked to fill out a self-rating of their vocational interests and then to choose one of their parents and judge this parent’s interests. Across the six high schools, 341 students completely filled out the self- and other-ratings. At the first high school, we initially asked the students to judge the interests of both parents, resulting in a total of 29 self-ratings with double parent ratings. However, students found it very tedious to fill out the same questions three times. Therefore the design was changed to include only one parent. For the 29 students with two other-ratings only the other-rating by the first parent was used. For the second sample, a group of undergraduate students approached people within their social network. Forty-seven people completely filled out the self- and other-ratings. For the third sample, university students participated in exchange for



credits. In the span of two years, 132 undergraduate psychology students completely filled out the self- and other-ratings. Thus, self- and other-ratings of vocational interests were available from 520 people in total. As a reward all participants received a short report of their personal interest scores and some vouchers were raffled.

All participants filled out the e-mail address of one of their parents. The parents were approached via e-mail within two weeks of their students' participation. For the first sample of 341 high school students, 59.53% of the parents completely filled out the self- and other-ratings, resulting in 203 "high school student-parent" dyads (students: ♀ = 62.6%,  $M(\text{age}) = 16.40$  years,  $SD = 0.99$ ; parents: ♀ = 56.7%,  $M(\text{age}) = 49.12$  years,  $SD = 5.29$ ). For the second sample of 47 students, 44.68% of the parents completely filled out the self- and other-ratings, resulting in 21 "student-parent" dyads (students: ♀ = 57.1%,  $M(\text{age}) = 22.40$  years,  $SD = 5.24$ ; parents: ♀ = 76.2%,  $M(\text{age}) = 52.80$  years,  $SD = 9.23$ ). For the third sample of 132 undergraduate students, 35.61% of the parents completely filled out the self- and other-ratings, resulting in 47 "undergraduate student-parent" dyads (students: ♀ = 91.5%,  $M(\text{age}) = 19.62$  years,  $SD = 1.47$ ; parents: ♀ = 68.1%,  $M(\text{age}) = 49.28$  years,  $SD = 5.23$ ). Further analyses in the present study are only conducted with participants in a complete dyad.

Before the samples were merged we investigated if the three groups differed in self-other agreement, assumed similarity, similarity and reciprocity. There were no such differences ( $F(2, 268) = 0.30$  to  $1.48$ ,  $p > .20$ ), also if the second and third sample were merged ( $F(1, 269) = 0.15$  to  $2.56$ ,  $p > .10$ ), because these two samples consisted of slightly older participants compared to the high school students. Because the groups did not show any significant differences on interpersonal perception of vocational interest it was decided that the groups could be combined for all subsequent analyses. In total, self- and other-ratings of vocational interests were available for 271 dyads (students: ♀ = 67.2%,  $M(\text{age}) = 17.42$  years,  $SD = 2.56$ ; parents: ♀ = 60.1%,  $M(\text{age}) = 49.47$  years,  $SD = 5.79$ ). Of these dyads, 63.4% was of the same gender (49 son-father and 123 daughter-mother) and 36.5% was of different genders (40 son-mother parent and 59 daughter-father).

## 2.2 Vocational interest measure

*The PGI-Short* (Tracey, 2010) measures vocational interests and contains two item-formats: activity liking and activity competence, divided over 40 activity statements. Each statement is rated twice on a 7-point likert scale,

once on liking (1 = *very strongly dislike*, 7 = *very strongly like*) and once on perceived competence (1 = *unable to do*, 7 = *very competent*). Example items are ‘Paint a portrait’ and ‘Oversee sales’. For the purpose of the current study, we used a Dutch translation of the PGI-Short. The items were previously translated to Dutch in a study (Holtrop et al., 2015) that used the complete PGI (Tracey, 2002). For the self-ratings we used the original instructions. For the other-ratings the participants were instructed to “indicate how much you think your [mother/father/daughter/son] likes the activities and how competent you think your [mother/father/daughter/son] is at the activities”. The survey tool automatically filled in the appropriate other where mother/father/daughter/son is shown.

The PGI-short yields several scales. We used the PGI’s ten spherical interest scales to compute the aspects of interpersonal perception, because together these ten scales fully represent the spherical representation of vocational interests. These ten scales consist of the eight basic interest scales, and the High and Low prestige interests scales. Each scale includes eight items with four items from each format. Alpha reliabilities for the ten scales ranged from .80 to .92 for the self-ratings and from .83 to .93 for the other-ratings. We also computed the three main axes of the spherical model: Ideas versus Data, People versus Things, and Prestige interests. These axes are computed based on the ten spherical interest scales (Tracey, 2010). The axes also showed adequate composite reliabilities (Feldt & Brennan, 1989), ranging from .78 to .93 for self-ratings and .80 to .95 for other-ratings.

Because the Dutch PGI-Short was used for the first time its circumplex structure (the equator with the basic interest scales) was tested with a nonparametric randomization test of hypothesized order relations (Hubert & Arabie, 1987) utilizing the program RANDALL (Tracey, 1997). This investigation is particularly important as the structure of the other-ratings was also tested. If the self- and other-ratings would not show an equivalent structure of vocational interests then comparing the two would be compromised. In a circumplex structure scales close to each other are expected to show higher inter-correlations than scales further away from each other. The randomization test of hypothesized order relations is used to test whether the correlations correctly decrease in magnitude as the distance between scales increases. These hypothesized order relations are called order predictions. The Correspondence Index (CI) reflects how many of the order predictions are met, ranging from -1.00 (all predictions violated) to +1.00 (all predictions met). In previous research (Tracey, 2010), the PGI-Short showed a

high average CI of .88 for the basic interest scales. For our data, the self- and other-ratings of vocational interests showed a significant circumplex fit with CI's ranging from .80 to .87 (Table 2). It was decided that this fit was adequate enough to proceed with further analyses.

**Table 2.**

*Randomization test of hypothesized order relations for a circumplex structure in self- and other ratings on the PGI-Short Octant.*

Rater	Student		Parent	
	Self	Other: Parent	Self	Other: Student
Predictions met	269	259	259	263
Predictions tied	0	1	0	0
<i>CI</i>	.868	.802	.799	.826
<i>p</i>	.000	.000	.000	.000

*Note.*  $N = 271$ ; *CI* = Correspondence Index (Hubert & Arabie, 1987); for every hypothesized order test 288 predictions were made.

### 2.3 Analyses

To test hypotheses 1-3 zero-order correlations were computed for all separate spherical interest scales for each aspect of interpersonal perception within a dyad (see Table 1). Next, to be able to compute an average correlation per aspect, an  $r$  to  $Z$  transformation was applied, then the  $Z$ -values were averaged, and then the average  $Z$  was transformed back to  $\bar{r}$ . This average correlation statistic summarizes the correlations between overlapping interest scales for each aspect of interpersonal perception. However, psychological profiles often show some degree of normativeness, which means that -on the average- two random profiles are somewhat similar to each other. An averaged correlation between two profiles, such as the one computed from the simple correlations, may therefore not just reflect the knowledge about one particular person, but also of people in general. This normativeness may cause an overestimation of the unique overlap between two interest ratings. Therefore, an  $r$  to  $Z$  to  $\bar{r}$  transformation was also applied to all interest scales to compute a baseline that reflects normativeness. The baseline thus consists of the average correlation between matching interest scales (e.g., correlation between Artistic interest ratings for A(A) and for B(A)) and non-matching scales (e.g., correlation between Artistic interest ratings for A(A) and Business detail ratings for B(A)).

Additionally, to further test hypotheses 1-3 we also computed profile correlations (Furr, 2008). These correlations describe the overlap between two individual multivariate profiles. The profile correlations were estimated with the R package Multicon (Sherman & Serfass, 2015). Using Multicon we computed

two types of profile correlations for every dyad for each aspect of interpersonal perception: Normative profile correlations and distinctive profile correlations. Normative profile correlations describe the overlap between two multivariate profiles, but still include normativeness. Distinctive profile correlations describe the overlap between two multivariate profiles with normativeness removed. To remove normativeness, the overlap between all non-paired profiles is computed and subsequently removed from the normative profile correlations. Therefore, distinctive profile correlations can be viewed as a summary of the unique overlap between all interest scales of two profiles into one correlation statistic. For the interpretation of the results we focused on the distinctive profile correlations. Please note that both the average normative and distinctive correlations were estimated with the *describe.r* function in Multicon. This function computes the average correlation applying an  $r$  to  $Z$  to  $\bar{r}$  transformation. The average normative and distinctive profile correlation presented is therefore somewhat different from the straightforward average of the normative and distinctive profile correlations.

For hypothesis 4, profile elevation (i.e., the mean score across all spherical interest scales) was computed for all self- and other-ratings of vocational interests. These profile elevation values were then used to compute a zero-order correlation for each aspect of interpersonal perception. Additionally, profile elevation was subtracted from all normative scale scores, creating a second -ipsatized/centered- interest scale score for every participant (Tracey, 2012). Due to the nature of the computation of profile correlations, the profile correlation scores of the ipsatized scales do not differ from the normative scales (in the computation of profile correlations the scores are already ipsatized). Next, the average correlation of the matching interests and non-matching interests were compared between the normative and ipsatized scales with a  $Z$ -test. If ipsatization increases self-other agreement of matching interest scales and decreases self-other agreement of non-matching interest scales, then ipsatization improves the measurement of self-other agreement (i.e., accuracy of other-ratings). Also, if ipsatization reduces the baseline/average correlation between all interest scales (i.e., the normativeness) of assumed similarity, then ipsatization reduces rater tendencies.

To test hypotheses 5 and 6, vocational interest perceptions (see Table 1) of same-gender and different-gender dyads were compared using several AN(C)OVA's. For these tests, the distinctive profile correlations were compared between groups. However, because correlations are generally not

**Table 3.***Normative interest scores: Means, Standard deviations and Reliabilities.*

Rater Target	Student						Parent					
	Self			Other: Parent			Self			Other: Child/student		
	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$
Spherical interest scales												
Social Facilitating	4.44	1.08	.84	4.09	1.16	.86	4.37	1.12	.80	4.86	1.15	.88
Managing	3.80	1.20	.88	4.10	1.27	.88	3.96	1.12	.80	3.97	1.17	.87
Business Detail	3.17	1.37	.92	3.76	1.45	.92	3.48	1.31	.87	3.47	1.33	.91
Data Processing	2.58	1.06	.84	2.88	1.28	.88	2.75	1.21	.85	2.89	1.12	.83
Mechanical	2.43	1.11	.87	2.87	1.44	.92	2.97	1.51	.91	2.69	1.14	.87
Nature/ Outdoors	3.30	1.26	.85	3.23	1.29	.87	3.33	1.24	.82	3.87	1.26	.84
Artistic	3.23	1.39	.90	3.15	1.52	.93	3.26	1.36	.87	3.81	1.47	.92
Helping	3.97	1.35	.86	3.89	1.34	.86	4.12	1.24	.83	4.34	1.31	.86
High prestige	3.63	1.21	.84	3.65	1.18	.84	3.51	1.33	.85	3.96	1.25	.85
Low prestige	2.65	1.10	.84	3.06	1.32	.87	3.53	1.29	.83	3.01	1.22	.84
Interest axes												
People (hi) vs Things (lo)	3.35	3.28	.93	2.16	3.95	.95	2.72	3.55	.92	3.51	3.21	.92
Ideas (hi) vs Data (lo)	-0.65	3.19	.93	-1.45	3.27	.93	-0.80	3.03	.89	-0.05	2.92	.91
Prestige	0.98	1.46	.78	0.59	1.67	.84	-0.02	1.77	.83	0.95	1.55	.80
Profile elevation	3.32	0.72	-	3.47	0.77	-	3.53	0.73	-	3.69	0.77	-

*Note.*  $N = 271$ ; scores on the spherical interest scales ranged from 1 = *very strongly dislike/ unable to do* to 7 = *very strongly like/very competent*.

normally distributed, first the  $Z$ -values were computed using Fisher's  $r$  to  $Z$  formula. These values were used as input for the AN(C)OVA's. After the analyses the  $Z$  values were transformed back to  $r$ 's.

### 3. Results

Table 3 shows the means, standard deviations, and reliabilities of the self- and other-reports of vocational interests. Overall, participants seemed to be mostly interested in activities with people over things, somewhat more interested in activities with data over ideas, and somewhat more interested in Prestige.

#### 3.1 Self- and other-rated vocational interests

Table 4 shows all scale and profile correlations for the normative and ipsatized scores. Cohen's (1992) guidelines were used to interpret the strength of the correlations. In these guidelines he describes correlations from .10 to .30 as weak, from .30 to .50 as moderate, and correlations higher than .50 as strong. Altogether, the normative results in Table 4 support the first three hypotheses. With respect H1, the distinctive profile correlations showed significantly positive and strong levels of self-other agreement,  $r = .55$  for agreement about the students' interests and  $r = .60$  for agreement about the parents' interests. For H2, the distinctive profile correlations showed significantly positive and moderate levels of assumed similarity,  $r = .46$  for assumed similarity by the students (i.e., how similar the students think they are to the parents) and  $r = .49$  for assumed similarity by the parents (i.e., how similar the parents think they are to the students). Additionally, the distinctive profile correlations for assumed similarity were significantly weaker than those for self-other agreement ( $Z = -3.18$  and  $-7.75$ , both  $p < .01$ ). Among the zero-order correlations of assumed similarity, the Business detail scale showed the weakest relations for both students' and parents' assumed similarity. For H3, the distinctive profile correlations showed significantly positive and weak similarity ( $r = .18$ ) and reciprocity ( $r = .22$ ). The distinctive profile correlations for similarity and reciprocity were significantly weaker than those for self-other agreement ( $Z = -4.53$  to  $-5.95$ , all  $p < .01$ ). However, the distinctive profile correlations for similarity and reciprocity were mostly not significantly weaker than those for assumed similarity, as only (actual) similarity was weaker than assumed similarity by the parent ( $Z = -2.31$ ,  $p < .05$ ).

**Table 4.**

Correlations between self- and other ratings within student(S)-parent(P) dyads.

	Self-other agreement				Assumed similarity				Similarity		Reciprocity	
	$R_{S(S),P(S)}^\dagger$		$r_{P(P),S(P)}$		$r_{S(S),S(P)}$		$r_{P(P),P(S)}$		$r_{S(S),P(P)}$		$r_{S(P),P(S)}$	
	Norm	Ips	Norm	Ips	Norm	Ips	Norm	Ips	Norm	Ips	Norm	Ips
Spherical interest scales												
Social facilitating	.502	.502	.471	.593	.542	.410	.594	.528	.243	.291	.395	.356
Managing	.444	.539	.461	.527	.414	.307	.440	.472	.185	.289	.180	.238
Business detail	.473	.530	.615	.669	.280	.187	.252	.252	.090	.108	.184	.233
Data processing	.486	.505	.616	.683	.495	.295	.510	.482	.307	.252	.356	.336
Mechanical	.514	.617	.649	.740	.487	.314	.513	.392	.275	.266	.344	.329
Nature/Outdoors	.542	.593	.432	.585	.536	.423	.564	.475	.294	.317	.316	.348
Artistic	.536	.625	.529	.638	.422	.354	.480	.372	.165	.206	.330	.333
Helping	.546	.657	.622	.773	.359	.354	.470	.354	.157	.253	.300	.335
High prestige	.491	.507	.418	.522	.520	.382	.478	.372	.161	.182	.282	.293
Low prestige	.332	.403	.557	.638	.477	.294	.555	.458	.217	.201	.295	.254
Interest axes												
Ideas (hi) vs Data (lo)	.598	.570	.667	.671	.367	.397	.452	.432	.271	.251	.344	.335
People (hi) vs Things (lo)	.714	.725	.803	.799	.395	.339	.509	.469	.372	.309	.435	.384
Prestige interests	.510	.504	.618	.609	.348	.382	.429	.435	.199	.200	.289	.286
Correlation aggregates												
Fisher's Z to $\bar{r}$ (all interests)	.149	.007	.120	.011	.243	.001	.225	.002	.055	.000	.105	.001
Fisher's Z to $\bar{r}$ (matching interests)	.489	.522	.542	.644	.457	.334	.490	.419	.210	.237	.300	.306
Normative profile correlation	.728	-	.730	-	.520	-	.554	-	.397	-	.448	-
Distinctive profile correlation	.547	-	.598	-	.327	-	.361	-	.177	-	.223	-
Profile elevation	.378	-	.330	-	.666	-	.619	-	.153*	-	.277	-

Note.  $N = 271$ ; Norm = correlation coefficients for normative scores; Ips = correlation coefficients for ipsatized scores.

$^\dagger r_{S(S),P(S)}$  indicates the self-other agreement and refers to the correlation between the children's self-ratings of vocational interests [S(S)] and the parents' other-ratings about the interests of their children [P(S)].

Correlations in *italic* are n.s., \* $p < .05$ , all other correlations are  $p < .01$ . Fisher aggregated correlation coefficients do not have a significance level.

### 3.2 Profile elevation and self- and other-rated vocational interests

Table 4 also shows the interest perceptions for profile elevation (i.e., the average score over all interest scales). In contrast to the findings on vocational interests (H1-3), self-other agreement and assumed similarity correlations in profile elevation seemed to be reversed in strength, as predicted in H4. There was a significantly positive and moderate self-other agreement in profile elevation ( $r = .38$  and  $.33$ , with respectively students or parents as target) and an significantly positive and strong assumed similarity ( $r = .67$  in students and  $.62$  in parents). Furthermore, profile elevation scores of self-other agreement were significantly lower than those of assumed similarity ( $Z = -4.70$  and  $-4.41$ ,  $p < .01$ ). Profile elevation scores of similarity and reciprocity showed comparable correlations (respectively  $r = .15$  and  $.28$ ) to the vocational interest scales. Next, self-other agreement and assumed similarity in profile elevation were compared to self-other agreement and assumed similarity in regular interests. For self-other agreement, the correlations of profile elevation were significantly lower than the distinctive profile correlations of the regular interest scales for students ( $Z = -2.50$ ,  $p < .05$ ) and parents ( $Z = -4.02$ ,  $p < .01$ ). For assumed similarity, the correlations of profile elevation were significantly higher than the distinctive profile correlations of the regular interest scales for students ( $Z = 5.37$ ,  $p < .01$ ) and parents ( $Z = 4.00$ ,  $p < .01$ ).

In addition to the normative interest scores, all regular interest scores were adjusted for profile elevation by ipsatizing the scale scores, as described in the analyses plan. For self-other agreement, the average correlation of matching ipsatized scales was not significantly higher than the correlation of the normative scores ( $Z = 0.51$ ,  $p = .61$  and  $Z = 1.83$ ,  $p = .07$ ), and the average correlation between non-matching interests scales (i.e., the normativeness) showed a non-significant decrease ( $Z = -1.66$ ,  $p = .10$  and  $Z = -1.27$ ,  $p = .20$ ). For assumed similarity, the average correlation of matching ipsatized scales was not significantly lower than the correlation of the normative scores ( $Z = -1.69$ ,  $p = .09$  and  $Z = -1.04$ ,  $p = .30$ ), but the average correlation between non-matching interests scales did show a significant decrease for assumed similarity ( $Z = -2.86$ ,  $p < .01$  and  $Z = -2.63$ ,  $p < .01$ ). Thus, ipsatization did not significantly affect self-other agreement or assumed similarity, except for a significant reduction in normativeness of assumed similarity. Note that the effect of the ipsatization on self-other agreement and assumed similarity seemed to be very small (if it would exist after all). This study provided low power to detect such small differences ( $1 - \beta$  ranged from  $.13$  to  $.57$  for the non-significant results).



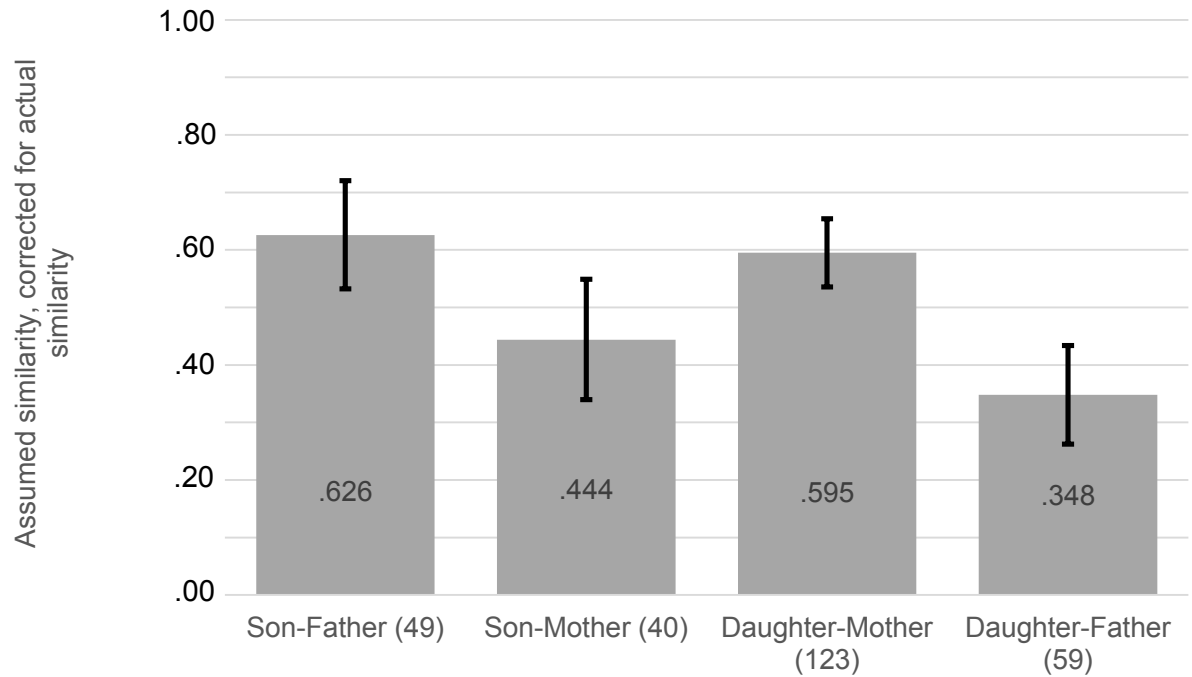


Figure 1. Means and 95% confidence intervals for assumed similarity of vocational interests of students with their parents ( $r_{S(S),S(P)}$ ), corrected for actual similarity ( $r_{S(S)=P(P)}$ ).  $R^2(\text{model}) = .197$ ,  $p < .01$ , observed power ( $1-\beta$ ) = .998.

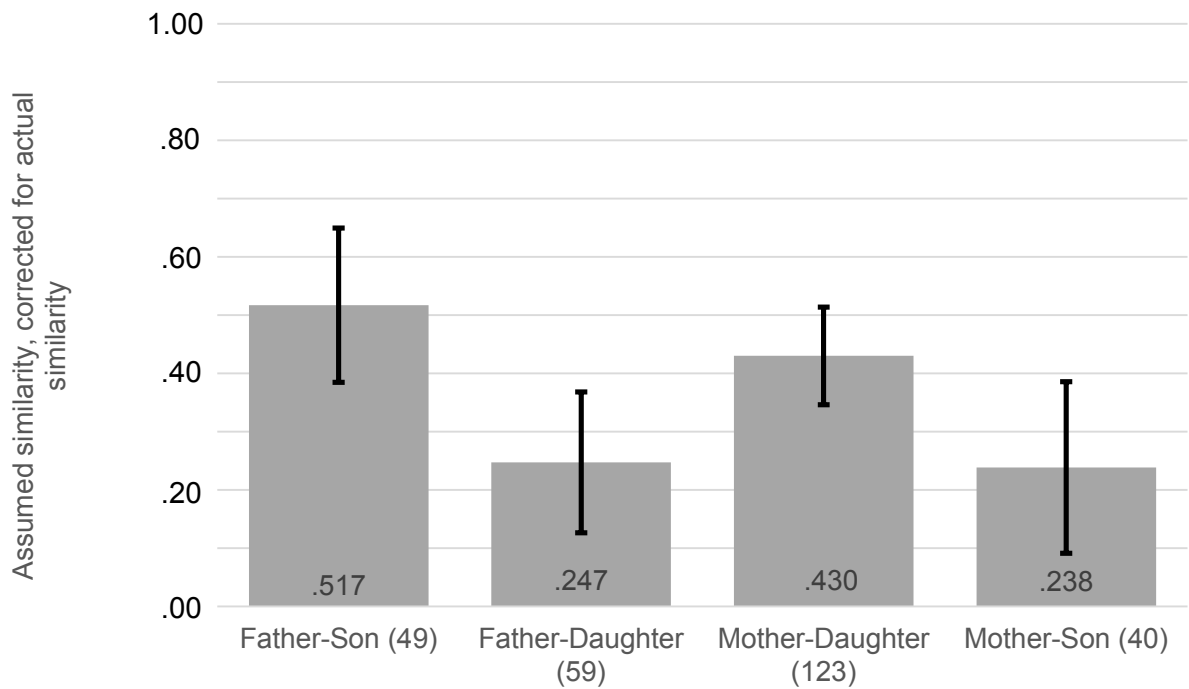


Figure 2. Means and 95% confidence intervals for assumed similarity of vocational interests of parents with their children ( $r_{P(P)=P(S)}$ ), corrected for actual similarity ( $r_{P(P)=S(S)}$ ).  $R^2(\text{model}) = .106$ ,  $p < .01$ , observed power ( $1-\beta$ ) = .956.

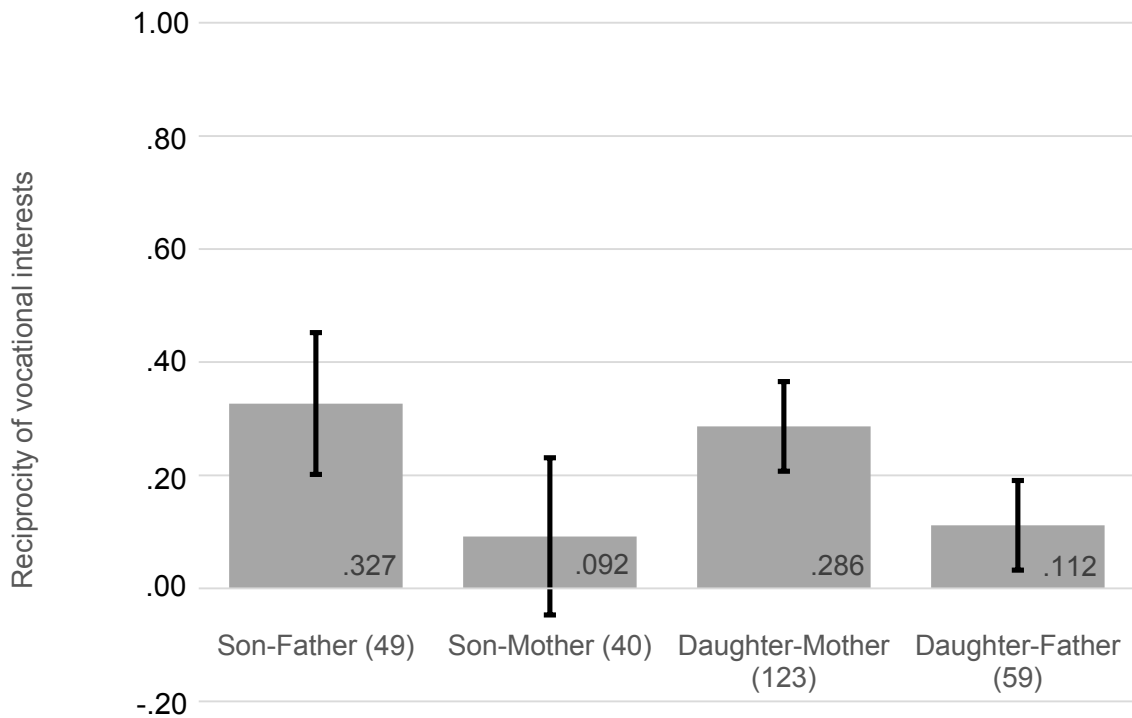


Figure 3 Means and 95% confidence intervals for reciprocity of vocational interests of students and their parents ( $r_{S(P),P(S)}$ ).  $R^2(\text{model}) = .043$ ,  $p < .01$ , observed power ( $1-\beta$ ) = .931.

### 3.3 Gender influences on self- and other-rated vocational interests

Lastly, we tested whether same-gender dyads differed in interest perceptions compared to different-gender dyads. For same-gender dyads, no differences were expected in self-other agreement and greater magnitudes were expected of assumed similarity (H5), and similarity and reciprocity (H6). The distinctive profile correlation was used to calculate the gender interaction for all four aspects of interest perceptions (see Table 1). Assumed similarity was also corrected for actual similarity.

For self-other agreement, no interaction effect was found – between the gender of the rater and the gender of target – for students ( $\eta^2 = .00$ ,  $p = .94$ ) and for parents ( $\eta^2 = .00$ ,  $p = .30$ ). For assumed similarity (H5a), a significant interaction effect was found – between the gender of the rater and the gender of the target – for students ( $\eta^2 = .09$ ,  $p < .01$ ) and parents ( $\eta^2 = .06$ ,  $p < .01$ ). Thus, higher assumed similarity was indicated when a person of the same gender was judged. When the main effect of actual similarity on assumed similarity was included in this model (H5b), the interaction effect for gender diminished slightly but remained significant for students ( $\eta^2 = .08$ ,  $p < .01$ ) and parents ( $\eta^2 = .05$ ,  $p < .01$ ). The interaction effects for assumed similarity are visually shown in Figures 1 and 2. Unexpectedly, for similarity (H6) no significant interaction effect was found between the gender of the targets ( $\eta^2 =$

.01,  $p = .08$ ). For reciprocity (H6) a significant interaction effect was found between the gender of the targets ( $\eta^2 = .04$ ,  $p < .01$ ). This interaction effect for reciprocity ( $\eta^2 = .04$ ,  $p < .01$ ). This interaction effect for reciprocity is visually shown in Figure 3. Thus, H6 was only partially confirmed. Overall, neither self-other agreement, assumed similarity, similarity, nor reciprocity showed significant main-effects for gender.

## 4. Discussion

### 4.1 Self- and other-rated vocational interests

The results showed that interpersonal perceptions of vocational interests are quite similar to interpersonal perceptions of personality. Using self- and other-reports of vocational interests we replicated Nauta's research (2012). She found high levels of self-other agreement in vocational interests and personality. Our results showed a similar high level of self-other agreement in vocational interests. Additionally we extended her research, by showing that vocational interests have relations similar to personality for three other interpersonal perceptions, namely for assumed similarity, similarity, and reciprocity. Assumed similarity was moderately strong for vocational interests. This means that, when judging others' vocational interests, students and parents generally indicated that those interests were moderately similar to their own interests. We confirmed only a weak presence of similarity and reciprocity, meaning that the correlation between two self-ratings or two other-ratings of vocational interests seems to be low in students-parent dyads. Note that vocational interests are partially heritable (e.g., Nelling et al., 2015) and that the dyads in our study were formed by parents and their children. Therefore, we expect that the moderate assumed similarity, weak similarity, and weak reciprocity found in the present study may be an overestimation when using dyads comprised of non-related persons.

Based on the high similarity between the interpersonal perceptions of vocational interests and personality, it could be argued that other research findings regarding self- and other-reports of personality, are likely to generalize to vocational interests. For example, Connelly and Ones (2010) showed that several other-reports of personality have a higher predictive validity for work performance than self-reported personality. This could mean that multiple other-reports of vocational interests may be a more valid measure of someone's interests (than a self-report) and possibly more predictive for a person's career choices. Consequently, future studies might like to compare

the predictive validity of self- and other-reports of vocational interests for study choice or performance.

#### *4.2 Profile elevation and self- and other-rated vocational interests*

Profile elevation showed a non-trait-like pattern of interpersonal perception in the present study. First, self-other agreement in vocational interests was high, implying that observers can judge the vocational interests of another person well. However, self-other agreement in profile elevation was moderate and significantly lower than in interests, indicating that observers cannot judge profile elevation as accurately as they can judge vocational interests. Second, assumed similarity in vocational interests was moderate, implying that observers judge others somewhat similar to themselves. However, assumed similarity in profile elevation was strong and significantly higher than in interests. This reversed pattern of self-other agreement and assumed similarity – compared to regular interest scales – indicates that profile elevation is probably less like a trait than substantive interest scales. Moreover, the finding that profile elevation shows strong assumed similarity indicates that profile elevation mainly resides within the ratings of one person and means that it is likely a rater tendency/bias. These findings thus support the ‘artifact’ interpretation (Prediger, 1998; Tracey, 2012) of profile elevation and oppose the ‘substantive’ interpretation (e.g., Fuller et al., 1999; Hirschi & Läge, 2007).

Ipsatization seemed to alleviate a large part of the bias created by profile elevation. Specifically, it completely removed the random correlation between self- and other-ratings from the same observer. Matching vocational interest scales showed a non-significant increase of self-other agreement and a non-significant decrease of assumed similarity. These non-significant results could be due to a lack of power. Overall, ipsatization seems to improve interpersonal perception measures of vocational interests somewhat. Practitioners and researchers should thus consider removing the bias of profile elevation by ipsatizing their vocational interest measures.

#### *4.3 Gender influences on self- and other-rated vocational interests*

The third major finding of the present study was that same-gender dyads showed higher assumed similarity and reciprocity than different-gender dyads. This interaction effect did not occur for self-other agreement and—unexpectedly—neither for similarity. Earlier studies have shown that observers have a tendency to rate targets more similar to themselves than that they actually are (e.g., Lee et al., 2009; Watson et al., 2000). In the present study,

assumed similarity, as rated by parents, was significantly higher than their actual similarity to the students, but not vice versa. Our findings are thus partially in line with previous findings. The tendency to overestimate the similarity to a target may also occur for vocational interests. Moreover, assumed similarity was significantly higher in same-gender dyads than in different-gender dyads (among both students and parents), even if corrected for actual similarity. Thus, observers with the same gender as the target overestimate the overlap in vocational interests even more than observers in different-gender dyads. When studying interpersonal relations with self- and other-rated vocational interests researchers should keep in mind that different-gender raters may provide more unique information.

The absence of a gender interaction effect for similarity is unexpected as vocational interests show strong gender-effects. Moreover, similarity and reciprocity indicate the same underlying interpersonal comparison (both are a between-person comparison with either self- or other-ratings). Because a gender interaction effect did occur for reciprocity, the non-significant result for similarity seems puzzling. However, the obvious difference between similarity and reciprocity may lie at the root of this difference. If an observer strongly identifies with a target then they may tend to project their own values on the target in their other-ratings (Lee et al., 2009; Watson, Hubbard, & Wiese, 2000). Since children identify themselves more strongly with same-gender parents (Starrels, 1994), other-ratings in a same-gender child-parent dyad correlate stronger than in different-gender dyads. However, taking into account that the interaction effect of similarity was nearly significant and that of reciprocity only showed a moderate interaction effect, future research needs to clarify the stability of the gender interaction effect of both these aspects of interests perception.

In summary, this study shows that dyadic perceptions of vocational interests have strong self-other agreement, moderate assumed similarity, and weak similarity and reciprocity. These findings are strongly aligned with findings in personality research. Furthermore, some differences are observed between perceptions of profile elevation and perceptions of vocational interests. Specifically, profile elevation shows moderate self-other agreement and strong assumed similarity, indicating that profile elevation in vocational interests is probably mostly a rater bias. Furthermore, compared to observers in different-gender dyads, observers in same-gender dyads overestimate their similarity in vocational interests to the target. All in all, this study shows that

other-ratings are a highly useful complement to self-ratings in the assessment of vocational interests.

# Improving personality and interest measurement for purposes of selection and assessment: Summary and general discussion



7





Personality and vocational interest measures have been found to be predictive of behavior and academic/job outcomes, consequently, they are often used in practice. Vocational interests are usually measured when people need to decide which vocations they wish to pursue. More than 50% of all Dutch aspiring students fills in a vocational interest measure before they choose the vocational path that they wish to pursue (Markteffect, 2011). Personality is usually measured when someone has already chosen a vocation and is applying for a job. Based on annual revenue estimates of personality tests (e.g., *The Economist*, 2013), a very conservative estimate would be that over 16 million people—globally—fill out a personality measure each year. Personality and vocational interests are thus important for major life decisions (i.e., finding a vocation and being selected for a job) and are used by a vast number of people. Therefore, the quality of personality and vocational interest measurement is of the utmost importance.

The present dissertation starts with an investigation of an improvement to personality measurement that is usually referred to as contextualization (Schmit, Ryan, Stierwalt, & Powell, 1995). Contextualization is a process in which a meaningful situation is added to personality items. This modification is known to increase the predictive validity of personality measures for behaviors relevant to the added situation. Separate studies of the dissertation investigated which types of contextualization are more preferable in terms of predictive validity and participant reactions. Next, the effect of contextualization on differential predictive validity, which has been shown to occur in personality measures (e.g., De Meijer, Born, Terlouw, & Van der Molen, 2008; De Vries, Born, & De Vries, 2012), was investigated between ethnic groups.

Subsequently, a recent model of personality, the HEXACO model (Lee & Ashton, 2004), was related to a recent model of vocational interests, the Spherical representation of vocational interests (Tracey & Rounds, 1996). For this purpose we translated a US-developed measure of the Spherical representation to the Dutch language. Last, we investigated the measurement of vocational interests using self- and other-ratings.

This discussion chapter concludes the dissertation by answering the six research questions posed in the introduction in light of the findings of the five empirical studies described in chapters two through six. Additionally, strengths, limitations, and practical implications of the present dissertation are discussed and suggestions for future studies are proposed. At the end of this chapter, a general conclusion is provided.

## 1. Summary and discussion of main findings

### **Research Question 1: Is a completely contextualized personality measure more predictive of academic and work performance than a tagged personality measure?**

Although personality has been found to be relatively stable (e.g., McCrae & Costa, 1994), personality-related behavior also tends to vary across situations. For instance, some people who are orderly at home, may be less so at school or at work and vice versa. The result of this personality variation may be that the predictive validity of personality traits are situation-dependent. Consequently, according to the Frame-of-Reference effect (FoR effect; Schmit et al., 1995), situation-specific personality measurement may explain more variance in (work and academic) behaviors relevant to that situation than generic personality measurement. To achieve the FoR effect, generic personality items need to be contextualized, i.e., modified to include a situation relevant to the behaviors that are to be predicted. Often, items are modified with a tag which directly refers to the context, such as “...at school” or “...at work”. However, Lievens, De Corte, and Schollaert (2008) suggested that contextualizing items further than merely tagging them may increase incremental criterion validity further too. Therefore, in the first two empirical chapters, a tagged and a completely contextualized inventory were compared. **Chapter two** focused on the comparison of these two Frame-of-Reference modifications in terms of their criterion validity in a setting of higher vocational education (a so-called university of applied sciences). In **chapter three** this study was repeated in a work setting with pharmacy assistants. In both studies the participants filled out a generic, a tagged, and a completely contextualized personality measure. Whereas the same method was used in both studies, the results were surprisingly different.

A total of 531 students participated in the first study of the Frame-of-Reference effect. The students filled out three versions (i.e., generic, tagged, and completely contextualized) of two personality measures. A total of 316 students completed the versions of the Multicultural Personality Test Big Six (MPT-BS; NOA, 2009; De Vries, De Vries, & Born, 2011) and 215 completed the versions of the HEXACO Personality Inventory Revised (HEXACO-PI-R; De Vries, Ashton & Lee, 2009; Lee & Ashton, 2004). Additionally, the students filled out a modified version of the Inventory of Counterproductive Behavior (ICB; Marcus, Lee, & Ashton, 2007). Finally, we collected the students' actual Grade Point Average (GPA) from the applied university's records. The results of this study showed that generic Conscientiousness was predictive of

academic performance and that Honesty-Humility/Integrity was predictive of Counterproductive Academic Behavior. Moreover, when compared to generic scales, the tagged scales showed an increase in criterion validity and the completely contextualized measures showed an even further increase in criterion validity. Thus, this study was able to replicate earlier findings of the FoR effect (e.g., Shaffer & Postlethwaite, 2012) and showed that complete contextualization evokes the largest FoR effect. The latter finding was the study's added value to the already existing literature.

In the second study of the FoR effect, data were collected among 139 pharmacy assistants from 33 different pharmacies. Similar to the students, the pharmacy assistants filled out three versions of the HEXACO-PI-R. In the case of the pharmacy assistants, the contextualization implied that the items were reformulated so that they focused on issues and tasks in a pharmaceutical context. Additionally, they filled out the Abridged Job In General scale (AJIG; 8 items; Bowling Green State University, 2009), which is a job satisfaction inventory. Lastly, individual job performance ratings were provided by the pharmacists (i.e., their supervisors). The results showed that generic Conscientiousness was predictive of job performance and that Honesty-Humility was predictive of job satisfaction. However, neither tagged, nor completely contextualized Conscientiousness was significantly related to job performance. Consequently, the contextualized Conscientiousness measures did not show any incremental criterion validity over the generic measure. Tagged and completely contextualized Honesty-Humility were, however, significantly related to job satisfaction. Still, neither contextualized measure was able to show incremental criterion validity over generic Honesty-Humility. Thus, this second study could not replicate the FoR effect and neither method of contextualization outperformed the other. In contrast, the results showed a reversed FoR effect: The generic measure of Conscientiousness was related to job performance and the contextualized Conscientiousness measures were not.

Whereas the study in chapter two positively answered the first research question, the study in chapter three did not. The purpose of chapter two and three was to investigate which contextualization method was preferable. It had not been expected that both contextualized measures would be unrelated to supervisory rated performance and that the generic measure would be related to it. However, the FoR effect is a robust phenomenon that is even meta-analytically supported (Shaffer & Postlethwaite, 2012), one null finding may not be a reason to dismiss it. Nonetheless, it may also be too early to dismiss the

findings of chapter three as just ‘error,’ associated with the design process or the particular sample used. The non-significant relation between the completely contextualized personality measure and job performance to our view is not likely to be caused by the design process. In both studies the contextualized measures were designed in a—mostly similar—structured collaboration with practitioners and subject matter experts. Moreover, both the tagged and completely contextualized measure showed the same non-significant relation to job performance in chapter three. This makes it unlikely that the non-effect was only due to the complete contextualization process and makes it more likely that it was due to contextualization as such. Furthermore, there is no reason to assume that the sample introduced specific error. Although the pharmacy assistants in the sample rated themselves only somewhat higher on (generic) Conscientiousness and Honesty-Humility than the HEXACO norm group, the scores on these dimensions were normally distributed. That is, there did not appear to be a ceiling effect. Overall, these considerations suggest that contextualization had an unexpected null-effect in this study and that the non-significant relations were not due to a specific contextualization method or to this particular sample. Moreover, a recent study (Robie & Risavy, 2016) found a similar null-effect for students when predicting GPA. However, neither our design, nor the design of Robie and Risavy allow for one clear explanation of these null-effects.

We propose two possible explanations for the non-significant relation between the contextualized measures and job performance in the pharmacist study, based on the fact that the pharmacy assistants rated their Conscientiousness significantly higher on the contextualized measures and also the standard deviation appeared lower for these measures.

First, considering that contextualization clarifies the situation a participant needs to imagine, the purpose of the personality measure becomes clearer to the participants, making them understand that their behavior in a work setting is being measured. Even though it was emphasized that the results would be treated confidentially, some participants indicated that they had their reservations when filling in the questionnaire because they did not feel secure that the data would not be shared with their supervisor. We did not receive comparable comments from the students in the first study. Based on this observation and the higher scores on the contextualized measures (compared to the generic measure), the contextualized measures may have elicited a self-enhancement bias from the pharmacy assistants. Subsequently,

this response style may have introduced error to the measurement of contextualized personality and may have reduced its criterion validity.

Second, the lower predictive validity of contextualized Conscientiousness may have been caused by the fact that conscientious behaviors at a pharmacy are restricted to a certain range. A pharmacy is a highly regulated environment, the assistants have to follow procedures that do not allow any sloppiness or lack of discipline. For example, it is not allowed for a pharmacy assistant to not maintain clear records nor is it optional to double check a medication's proportions before handing it to a client, such actions are always executed according to a rigid protocol. Therefore, it may be possible that, for pharmacy assistants, conscientious behaviors at the job may be restricted to a too limited range. The measurement of generic Conscientiousness is not limited to these specific behaviors in a job setting. Some variance of behaviors outside the job may also be important for job performance. For example, being conscientious in general (having an organized life) may also help an assistant to be on time at work, which in turn results in higher supervisory ratings.

**Research Question 2: Is a completely contextualized personality measure more positively perceived by participants than a tagged personality measure?**

In **chapters two, three, and four** participant reactions to contextualized personality measures were investigated. Participants in chapters two and three were asked to rate the generic personality measure and both contextualized personality measures in terms of liking, face validity, and perceived predictive validity. In chapter four, 309 applied university students from three different universities of applied sciences completed the generic and completely contextualized MPT-BS measure used in chapter two.

On the matter of liking, in all studies the participants on average indicated liking all personality measures somewhat higher than the scale average of '4' (on a scale ranging from (1) *completely disagree* to (7) *completely agree*). In chapters two and three, the tagged measure was liked the least. The completely contextualized version was liked less than the generic measure in chapter two and liked similarly (to the generic measure) in chapters three and four. It is not very surprising that the tagged measure is liked the least, because it feels very repetitive if the same tag is used over and over again in each item of a scale.

On the matter of face validity, which refers to how relevant the participants find the measure for their tasks/role, in all three studies the completely contextualized version was rated most face valid. Additionally the tagged measure was rated more face valid than the generic version by the students in chapter two and equally face valid to the generic version by the pharmacy assistants in chapter three. Overall, contextualization seems to have a positive effect on the face validity of a personality measure. Moreover, this effect appears to be larger for completely contextualized measures than for tagged measures.

On the matter of perceived predictive validity, the results are not as clear as on the previous two participant reactions. In chapter two, the students only perceived the completely contextualized inventory significantly more predictive. In chapter three, the pharmacy assistants found only the tagged measure more predictive, and, in chapter four, the students found the generic and completely contextualized measure equally predictive. Altogether, contextualization may improve the perceived predictive validity of a personality measure somewhat, but this effect seems to not occur systematically and is not more present for either method of contextualization.

Overall, contextualization seems to positively improve participant reactions, with the exception that tagged contextualization (the easiest form of contextualization) is liked less than generic measures by participants. Moreover, complete contextualization improves participant reactions more than tagged contextualization does. Therefore, if the goal is to improve participant reactions, then completely contextualized measures seem to be preferable over tagged measures.

**Research Question 3: Does contextualization reduce the differential validity of personality measures across ethnic groups?**

Some studies have shown that personality measures may have differential predictive validity (De Meijer et al., 2008; De Vries et al., 2012). **Chapter four** reports on a study that investigated whether contextualization reduces differential validity. The argument was that, when filling in a generic personality measure, Dutch majority students may automatically apply a school FoR more often to the (context-less) items than non-western minorities. The FoR effect works through the removal of non-relevant variance from generic items by ensuring that all items are filled out with the relevant context in mind. Non-western minorities include the resident culture at school in their self-view to different extents. Generic personality measures, filled out by

members of non-western minorities, may thus include more non-relevant variance, because the school context is less frequently used for generic items. Thus, we expected the FoR effect to be stronger for non-western minority students, which could result in a reduction of personality measures' differential validity.

A generic and a completely contextualized personality measure were filled out by 326 students. Of these students, 190 were part of the Dutch majority group and 110 were part of a non-western minority group. All students were also asked to estimate their GPA. Additionally, the non-western minority students were asked to fill out an acculturation measure that measured the Maintenance of the maternal/paternal culture and their Accommodation to the Dutch culture. The students' actual GPA was available from the applied universities' records.

The analyses showed that for the entire group the FoR effect was replicated: The completely contextualized personality measure predicted more variance in actual and self-reported GPA than the generic measure did. However, when the analyses were performed separately for the majority and non-western minority group, the predictive validity of the generic and the contextualized personality measure was only significant for the Dutch majority group. These results showed two things. First, the Frame-of-Reference effect was only present for the majority group and did not exist for the non-western minority. Second, both personality measures showed differential validity for the prediction of GPA. More precisely, these measures showed single group validity, implying that they were only predictive of the majority group's GPA. Therefore, the results of chapter four contradict the suggestion that contextualization may reduce the differential validity of personality measures (Church, 2010; De Vries, Born, & De Vries, 2012).

There are cultural differences in how non-western minorities and western majorities fill out personality inventories (e.g., He & Van de Vijver, 2013). However, it is important to note that the differential validity of personality inventories may not (only) be due to how people with different (cultural) backgrounds fill out personality inventories. This differential validity could perhaps also be attributed to the construction of the criterion. In chapter four's study, the students were also asked to estimate their GPA. When the predictive validity for actual GPA (as obtained from the institutional records) was compared to self-reported GPA, a mostly similar predictive validity was found for the majority students, but different validities were found for the non-western minority group. For the non-western minority students, a number of

personality traits that were not predictive of their actual GPA, unexpectedly showed predictive validity for self-reported GPA. Conscientiousness (generic and contextualized) was the most predictive trait of self-reported GPA. Additionally, generic Integrity and Emotional Stability were predictive and, for the contextualized measure, only Agreeableness did not show a significant correlation to self-reported GPA. Moreover, across these specific traits, the predictive validity of the contextualized inventory seemed slightly, but non-significantly, higher than the generic inventory.

Clearly, personality traits related stronger to self-reported GPA than to actual GPA for the non-western minority. This leads to the suggestion that the differential validity of personality measures may not only be attributable to cultural differences in the measurement of personality alone, but perhaps to a broader factor that also affects the measurement of criteria. The most salient possibility is that the enhanced correlation between personality and self-rated GPA (compared to actual GPA) is caused by an overarching rater response style that affects both predictor and criterion. If the same response style affects personality measurement and self-rated academic performance then they innately share some (error) variance and the strength of the relation between predictor and criterion may thus be an overestimation.

Another explanation could be that the differential validity of personality measures may also be caused by the measurement of the criterion. Even though we labeled GPA obtained from the institutional records as 'actual GPA', this label does not mean that it is a fully objective measure. A large part of this GPA is not made up of test scores, but exists of assignment ratings, such as essays, internships, and practical assignments. These assignments are rated by—mostly—majority teachers, whose ratings could possibly be affected by rater biases. In the US-based organizational literature, supervisory performance ratings of white managers have been found to be higher for white employees than for black employees, whereas black managers rated white and black employees equally high (Staufer & Buckley, 2005). Possibly actual GPA is partly affected by rater biases that subsequently cause differential validity in personality measurement. However, we hasten to note that there is no evidence to support such rater biases in an academic context so far.



**Research Question 4: Is profile elevation in interest measurement a substantive factor or a nuisance/artifact?**

For a long time scholars have investigated profile elevation of vocational interests (e.g., Prediger, 1998). Profile elevation is a large overarching general factor dominating interest measurement. Profile elevation refers to the phenomenon that some people—on average—score higher across all scales of a vocational interest measure. Some scholars have argued that profile elevation is a factor that has substantial meaning and that it can be used for counseling purposes (e.g., Fuller, Holland, & Johnston, 1999). For example, Fuller et al. suggested that low profile elevation may be an indicator that someone has maladjustment issues. Others have argued that profile elevation is a statistical artifact that needs to be ignored or statistically corrected for (e.g., Tracey, 2012). So far, the evidence for either interpretation is inconclusive.

In chapter five profile elevation was related to personality traits. The results showed that profile elevation was moderately correlated to Openness to Experience, which is congruent with previous research findings (e.g., Fuller et al., 1999). Additionally, profile elevation showed a small negative relation to Honesty-Humility. Individuals who are curious, creative, less honest, and less modest appear to have a higher interest profile elevation. These findings support the notion that profile elevation is at least partially a substantial factor, because it relates to substantial personality traits.

Consecutively, the raw scores on the vocational interest scales (i.e., the normative scores) were ipsatized with the purpose to partial out profile elevation. This means that the average interest score of each participant was computed (i.e., the profile elevation) and subtracted from all individual scales, essentially removing profile elevation from the separate interest scales. Vocational interests of which the normative scales were related to profile elevation, correlated differently with personality traits when the measures were ipsatized. For example, the normative 'interest in Helping scale' correlated non-significantly (when Bonferroni-corrected for the number of analyses) to Honesty-Humility, whereas the ipsatized interest in Helping scale remained significantly and moderately related to Honesty-Humility. This finding shows that profile elevation indeed changes the relation of vocational interest scales to other variables and that profile elevation may be a suppressor variable of the relation of (some) interest scales to other variables.

In chapter six, self- and other-ratings of vocational interests were used to compute self- and other-rated profile elevation. Next, the self- and other-ratings of vocational interests and profile elevation were compared. The results showed distinctly different correlational patterns of self- and other-ratings for vocational interests than for profile elevation. Profile elevation showed higher so-called assumed similarity than vocational interests did, and a lower self-other agreement.

For the ease of interpretation, imagine a dyad consisting of Anouschka (A) and Badr (B). First, imagine that Anouschka rates her own vocational interests (A(A)) and Badr's interests (A(B)), which are supposedly not related. The correlation between Anouska's self-rating and her other-rating (about Badr's interests) is referred to as assumed similarity (i.e.,  $r_{A(A),A(B)}$ ). The assumed similarity of profile elevation between *her* self-ratings and *her* other-ratings is strong, whereas the mean assumed similarity of vocational interests is moderate. Thus, profile elevation shows a higher correlation than vocational interests between two supposedly unrelated ratings given by the same person and may thus be influenced more by individual biases than ratings of vocational interests. Second, imagine that Anouschka rates Badr's interests (A(B)) and Badr rates his own interests (B(B)). The correlation between Anouska's other-rating (about Badr's interests) and Badr's self-rating is referred to as self-other agreement (i.e.,  $r_{A(B),B(B)}$ ). The self-other agreement of the profile elevation between *her* other-ratings and *his* self-ratings is moderate, whereas the mean self-other agreement of vocational interests is strong. This latter finding indicates that profile elevation is not as easily observable by another person as vocational interests are.

This finding indicates that profile elevation exists mostly within one rater and may therefore be attributable to a substantial rater response style, such as acquiescence. Moreover, profile elevation should not be viewed as a 'strength of vocational interests' factor because it shows distinctly different correlational patterns of self- and other-ratings.

Altogether, the findings in chapters five and six support the notion that profile elevation may be an artifact in the measurement of vocational interests, because it mostly exists within one rater. However, this artifact appears to also have some form of substance because it is related to substantial traits. Previous research in personality (e.g., He & Van de Vijver, 2013; Vigil-Colet, Morales-Vives, & Lorenzo-Seva, 2013; Zettler, Lang, Hülshager, & Hlbig, 2015) found response styles to be related to personality traits. Response styles include rater tendencies such as acquiescence (where a participant

agrees with most items regardless of their content), midpoint responding (where the participant tends to use the scale midpoint), and extremity (where the participant tends to use the far ends of the scale). He and Van de Vijver (2013) found that these response styles are strongly related to the general factor of personality—a putative and mostly controversial (e.g., Ashton, Lee, De Vries, & Goldberg, 2009; Van der Linden, Te Nijenhuis, & Bakker, 2010) higher-order factor of personality—, much like response styles in vocational interests may be related to profile elevation<sup>4</sup>.

As shown in chapter five, vocational interests and personality are two related fields of individual differences. Moreover, both fields usually measure their respective constructs with self-reports. Future studies may show to what degree profile elevation of vocational interests and the general factor of personality are similar. Because the interest and personality general factor both seem to relate to rater biases, it may be true that they are also caused by similar response styles.

**Research Question 5: Prestige vocational interests: to which personality dimensions are these interests related?**

In **chapter five** the Spherical representation of vocational interests (Tracey & Rounds, 1996) was introduced. This model includes so-called Prestige interests as a dimension of vocational interests. The interest scale highest on the Prestige dimension is labelled ‘Influence’ and the scale lowest on this dimension is labelled ‘Manual Work’. Prestige interests represent the difficulty, training, knowledge, education, and effort required of the activities or jobs a person is interested in (e.g., Roe, 1956; Sodano & Tracey, 2008). The name of the Prestige interests dimension might also imply that people who are interested in status are attracted to vocations high in Prestige interests.

In this dissertation, the Spherical representation was empirically related to the HEXACO model of personality (Lee & Ashton, 2004), to further our understanding of how vocational interests and personality are related. For the purposes of this study we translated the U.S.-developed Personal Globe Inventory (PGI; Tracey, 2002), a questionnaire to assess the Spherical representation, to Dutch.

Subsequently, 656 participants filled out the Dutch PGI and the Dutch HEXACO-PI-R (100 item version; De Vries et al., 2009). The translated PGI

---

<sup>4</sup> Note that this general factor is normally not found in the HEXACO (De Vries, 2011). A facet level Principal Component Analysis with Oblique rotation was performed on the data of chapter five ( $n = 656$ ; HEXACO 100 item version) and indicated that a six-factor solution best fitted the data. Results can be obtained from the first author.

showed good psychometrical properties. Although Prestige interests and Honesty-Humility might conceptually appear to be negatively related, Prestige interests did not show a significant negative relation to Honesty-Humility. Therefore, it was *not* supported that people who like vocations and activities that are in positions of influence and that require skill and effort, are more interested in materialistic gains, or less fair, sincere, and modest. On the other hand, Prestige interests were positively related to Openness to Experience and somewhat to Extraversion.

Therefore, the results showed that Prestige interests are related to imagination, curiosity, and social energy. People who are open to new experiences and who are more extraverted appear to be drawn more towards influential jobs that require more effort and training. Openness to Experience is the one personality dimension that is somewhat related to educational level and therefore it is not unexpected that people high on Openness are also more interested in vocations that require more training. Jobs high in Prestige interests can be considered complex in nature, demanding and dynamic. The relation between Extraversion and Prestige interests may be explained by the fact that Extraverts are attracted by the higher effort and energy required of vocations higher in Prestige interests.

**Research Question 6: Do other-ratings of vocational interests show self-other agreement, assumed similarity, similarity, and reciprocity?**

Usually, if personality is measured in a dyadic context, research finds strong self-other agreement (do a person's self-ratings converge with ratings by others?), moderate assumed similarity (do a person's self-ratings converge with how s/he rates others?), and weak similarity (do self-ratings of two persons converge?) and reciprocity (do the other-ratings of two persons in a dyad converge?) (e.g., De Vries, 2010; McCann, Lipnevich, Poropat, Wiemers, & Roberts, 2015; Watson & Clark, 1991).

The main purpose of **chapter six** was to investigate if the properties of self- and other-ratings of vocational interests are similar to those of personality. In this study, 271 adolescents (age ranging from 15 to 35) and one of their parents rated their own and each other's vocational interests. The short version of the PGI (Tracey, 2012) was used to measure self- and other-rated vocational interests. Indeed, the profile correlation (a summary correlation coefficient that was used to summarize the correlations between all interests) showed that other-rated vocational interests do have similar properties to other-rated personality: The results showed strong self-other agreement,

indicating that others can accurately judge a person's vocational interests. Additionally, the results showed a moderate assumed similarity, and a weak similarity and reciprocity. Moreover, the results showed that dyads usually overestimated their similarity in vocational interests (assumed similarity remained significant even if actual similarity was partialled out), and that same-gender dyads overestimated their similarity in interests (assumed minus actual) even more.

The strong self-other agreement found, is an indication that others can indeed be relied upon to judge the interests of a person. Therefore, other-rated interests can be considered a useful measure for vocational counseling. Moreover, some observers may be more informative than others of a person's interests. Overall, participants were found to overestimate the similarity (assumed minus actual similarity) with the other somewhat. This finding is not unusual, as personality research has also found that people tend to overestimate their similarity to people they are close to (Lee et al., 2009). However, the fact that same-gender dyads tend to overestimate their similarity in interests more is an indication that people of a similar gender may be less informative judges because they may distort other-ratings by projecting personal interests on the similar gender others. Additionally, because the properties of other-rated vocational interests appear highly similar to those of other-rated personality, some of the findings about other-ratings of personality could possibly be expected to be similar for other-ratings of vocational interests. For example, Connely and Ones (2010) found that several other-ratings have greater and incremental validity to self-ratings of personality. The same may be true for other-ratings of vocational interests. Although our results do not allow any conclusions about this, it may indeed be true that several other-ratings of interests may be more predictive of career choice or performance than one self-rating.

## **2. Strengths and limitations**

One of the major contributions of the present dissertation is the development, validation, and application of personality and vocational interest measurement methods. First, chapter two presented a procedure for the contextualization of generic personality inventories. This procedure was subsequently applied to two personality measures, the HEXACO (De Vries et al., 2009; Lee & Ashton, 2004) and the MPT-BS (NOA, 2009; De Vries et al., 2011), in a school context, and to the HEXACO in a work (pharmacy) context. In all instances the modified inventories maintained their original factor structure, the alpha coefficients remained approximately similar, and the

contextualized scales correlated strongly to the generic original. Therefore, it may be concluded that the modification procedure is useful for future researchers or practitioners who wish to construct their own contextualized inventory. However, we recommend to not make the contextualized items overly specific to behaviors that a person will rarely need to display, as this may negatively impact the predictive validity. For example, an over-contextualized item to measure Emotional Stability could be “I stay calm during a robbery”, because most people will never have encountered this situation.

Second, the present dissertation also contributed to the measurement of vocational interests. In chapter five, the only existing measure of the Spherical representation of vocational interests, the U.S.-developed PGI (Tracey, 2002), was translated to Dutch. The PGI's (and thus the Spherical representation's) structural properties were investigated and confirmed in a large and diverse Dutch sample. Subsequently, in chapter six, we used the abbreviated version of the PGI to investigate other-ratings of vocational interests. Based on the results we argue that vocational interest measures, similar to the PGI, may also be applied to collect other-ratings. This opens up new possibilities for vocational counselors, because they can now involve others in the counseling process in a structured manner.

Third, the first three empirical chapters contributed to a replication of the FoR effect. Makel, Plucker, and Hegarty (2012) investigated how many published psychology studies are replications of previous work. At the beginning of their paper they quoted John Tukey (1969, p. 84): “Confirmation comes from repetition. Any attempt to avoid this statement leads to failure and probably to destruction”. Subsequently they showed only 2.39% of the published psychology studies (after the year 2000) had indicated to be a replication. The FoR effect had been investigated by a number of previous studies (Schmit et al., 1995; Shaffer & Postlethwaite, 2012) and is now a well-established phenomenon. Overall, the FoR effect was replicated in the present dissertation. In two studies we replicated the FoR effect for ethnic majority students, but failed to replicate it for the pharmacy assistants and non-western minority students. The latter finding is most likely due to differential validity and not necessarily evidence against the FoR effect. However, the findings of the pharmacy assistants and the recent findings of Robie and Risavy (2016) indicate that the FoR effect may not always occur. In addition to the replications, the studies in the present dissertation show that complete contextualization should be preferred over tagged contextualization.

Another strength of the present dissertation is the use of objective and supervisory rated performance criteria in all studies using performance as a criterion. Our findings in chapter four clearly show how predictive validity can be affected by the use of self-reported performance. All Grade Point Average (GPA) ratings were retrieved from institutional records and performance ratings in chapter three were provided by supervisors. Supervisory ratings and GPA may be affected by certain biases, but these are still most representative for the criteria that a person is evaluated on in practice. Additionally, we extended the criteria on which contextualization is evaluated by including Counterproductive Academic Behavior and participant reactions. So far, no previous research had investigated the FoR effect for (self-reported) counterproductive behaviors and only one study (Holtz, Ployhart, & Dominguez, 2005) had investigated the differences in affective reactions to generic and contextualized inventories. Moreover, Holtz et al. mostly focused on attitude differences towards the administering organization, such as recommending the organization to others, whereas we looked at attitude differences towards the tests.

Some limitations are also worth mentioning. The first and foremost limitation is that the studies in the present dissertation are cross-sectional in nature. This prevents any causal inferences to be drawn. This may be especially relevant to the chapters on the FoR effect. A recent longitudinal study (Liu & Huang, 2015) asked expat students to rate their contextualized extraversion (using tagged contextualization) at three points in time. The study's results showed that the expat students' initial contextualized Extraversion ratings (measured directly after arriving in the U.S.) changed over the course of four months. Moreover, the initial Extraversion and the change (i.e., increase) in Extraversion were separately predictive of adjustment outcomes. The authors concluded that stable personality traits (i.e., generic traits) may change within one context (i.e., contextualized traits) over the course of several years, and that this malleability of personality may be especially relevant for cross-cultural (adjustment) outcomes. However, note that an alternative interpretation of their findings could be that—due to an increase in experience with the context—the (subjectively felt) context 'changes' (and maybe not so much personality).

A second limitation of the present dissertation is that all samples were mostly highly educated. Therefore, our samples may over-represent certain key characteristics that allow them to function at higher levels, such as higher general mental abilities, socio-economic status, and higher educated social

network. This limits the generalizability of our findings to approximately 43% of the Dutch population (CBS, 2012). This sampling limitation may have especially affected the findings regarding the Spherical representation, because Prestige interests are related to educational level and social economic status. A restriction of range in a sample may lead to an underestimation of correlations. As a consequence of this sampling limitation the correlations between Prestige interests and personality may thus be an underestimation.

### **3. Practical implications**

First, our studies showed that completely contextualized measures appear to receive more positive participant reactions and appear to have a higher criterion validity than generic and tagged measures. Note that there are some previously mentioned caveats to this incremental criterion validity. For practice we would recommend against the use of tagged inventories as participants appear to like these less than non-tagged generic or completely contextualized inventories. Based on our results, it is advisable to use completely contextualized personality measures over generic measures. However, designing a completely contextualized personality measure involves a lengthy procedure with several stages. In our case, designing and pre-testing took approximately 65 hours (excluding obtaining a norm group). At first glance, the relatively small incremental criterion validity of completely contextualized scales may not seem encouraging enough to engage in this process. In some cases, the small increase in validity may still improve the utility of a selection procedure, for example in a situation with a low selection rate and an average base rate. Moreover, participants like completely contextualized personality inventories more and consider these to be more face valid. Therefore, the chance of objections to assessments with completely contextualized personality inventory may be lower than the chance of objections to generic inventories. Thus, completely contextualized measures may have a reduced chance of complaints—or even lawsuits—by improving (positive) participant reactions. The improved predictive validity and participant reactions are strong arguments in favor of designing completely contextualized questionnaires, especially when a context is widely applicable (e.g., school or jobs in the service sector).

A second practical implication relates to the finding that generic and contextualized personality inventories showed differential criterion validity for ethnic groups. Some previous studies already indicated that personality inventories may have differential validity. The results in chapter four even



showed single group validity for personality tests in predicting academic performance. Contrary to the suggestions of some scholars (Church, 2010; De Vries et al., 2012), contextualization did not reduce this differential validity in our study. But then why do personality tests have differential criterion validity? We would argue that there may be third variables that need to be taken into account. Non-western participants seem to rely on different answering strategies (He & Van de Vijver, 2013). These strategies may affect the predictive validity of personality inventories, especially if the strategies of the non-western participants alter personality measurement more than the strategies of the majority. It may therefore be advisable, when assessing a diverse group, to not overly rely on personality measures when there is indeed evidence of differential validity, but instead include measures that have shown less differential validity and adverse impact, such as Situational Judgment Tests, work samples, or the use of open-ended responses (Ployhart & Holtz, 2008).

Finally, the present dissertation shows that other-ratings of vocational interests may be used to measure a person's interests. To our knowledge, interests are rarely (possibly never) measured via other-ratings. Others are usually involved in vocational counseling through interviews or discussions. The fact that other-ratings accurately reflect a person's interests, but are not completely similar to self-ratings, opens up new possibilities for vocational counseling. Using other-rated vocational interests will allow counselors to involve acquaintances in a structured and reliable manner. If the participant has little vocational experience, then counselors should aim to collect other-ratings from well acquainted others with more vocational experience. These acquaintances may even be able to better estimate which jobs and activities the participant likes than the participant him/herself, because they have more knowledge of the person as well as of vocations. For example, an aspiring student may not understand what an accountant does and thus may find it hard to estimate how much he or she would like working as one. Another person, who knows the student well and knows more precisely what an accountant does, may find it easier to estimate how much the student would like working as an accountant. Subsequently, the student could use the results of the other-reports as an additional aid for conversations about his or her vocational choice with the observer and counselor.

We would recommend to only use the highest interest scores of other-assessments and warn against using the lowest scores, because the purpose of vocational interest measurement is to explore what someone would like to

do. It may be detrimental to this process if a well-acquainted other points out what the person would not like. Moreover, counselors should be aware that that other may have ulterior motives by ranking vocations higher or lower, especially if they are invested in the person's future. For example, a parent may be against an Arts specialization because "there is no money to be made in those professions".

#### **4. Suggestions for future research**

The studies in this dissertation and also previous studies (e.g., Shaffer & Postlethwaite, 2012) have shown that behavior can be more specifically predicted if individual differences are measured when taking a context into consideration. However, individual differences may not have simple relations to these situational affordances: Personality may interact with situations via complex processes, and situations may affect another situation's interaction with personality. For instance, one of the major drawbacks of chapter four is that only the school situation is taken into consideration. Some studies (Meeuwisse, Born, & Severiens, 2014; Wolff, 2013) indicated that the home situation is crucial for the academic performance non-western minority. Therefore, our first suggestion for future research would be to investigate the relation between situational affordances and individual differences comprehensively (i.e., include a multitude of situations and individual differences), to fully understand how situations and individual differences produce behaviors and subsequently affect performance. This line of research may be especially useful to understand the processes that underlie cultural differences in person-situation interaction.

Rauthman et al. (2014) proposed a taxonomy of situations to define, describe, and measure the characteristics of situations. In a series of studies, they showed that situations can be described with eight dimensions: Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, and Sociality (DIAMONDS). Subsequently, they showed that these situational characteristics are highly predictive of behavior. They suggested that future research may use the DIAMONDS taxonomy to more fully understand the relation between traits and situations. Future research could measure personality (e.g., with the HEXACO model) and the situations (with the DIAMONDS model) in a diary study. The results could then be interpreted in light of the recently proposed Situation-Trait-Outcome Activation (STOA; De Vries et al., 2015) model that describes how individual differences and situations produce behaviors. The STOA model argues that traits and situations lead to behavior via three pathways: 1) Situation activation (Buss,

1987), which means that people differ in the way they select, manipulate, or evoke situations that allow for *exploitation* (allowing for the expression of traits)

- 2) Trait activation (Tett & Burnett, 2003) which means that situations may activate traits for some people and not for others
- 3) Outcome activation (De Vries et al., 2015), which means that the consequences of behaviors in situations that allowed for *exploitation* stimulate or discourage future behaviors.

We believe that the processes described in the STOA model may help explain differential validity of personality inventories. Because values and personality are related (Roccas, Savig, Schwartz, & Knafo, 2002), the value of the expression of some traits may be partially dependent on cultural differences. Therefore, cultural background may influence the selection and manipulation of situation that allow for an expression of traits. Moreover, the same situation may be perceived differently depending on cultural background and therefore elicit different behaviors.

Another line of research that future studies could pursue is the further investigation of other-rated measurement of vocational interests. We conducted one of the first studies that charted the fundamental attributes of these ratings. Based on the results of this study and the study by Nauta (2012), future research can build on the finding that other-ratings of vocational interests behave mostly similar to other-ratings of personality. However, the practical value of other-ratings of vocational interests has not yet been established. First and foremost, future research would be advised to look into the predictive validity of other-rated interests for career decisions. Ideally, interests would be rated in a round robin design as Kenny (1994) described in the Social Relationships Model (SRM). This design requires a group of participants to all rate each other. Multiple other-ratings could then also be averaged and compared in terms of predictive validity to self-rated interests. Note that it is also possible that such a design would not provide more informative results, because the added value has been found to be relatively small when compared to a design that only employees self- and (multiple) other-ratings in a standard (non round robin) design (De Vries, 2010). Second, it remains to be seen if other-rated interests yield new insights to a participant and if different views on the participant's interests will be accepted by the participant. Future research could investigate how much people appreciate the opinion of others about their interests and under which conditions they use other-ratings to make career decisions.

## 5. Conclusion

A vast number of people use personality and vocational interest measures when making major life decisions. In psychological science, these measures are continually improved. The goal of the present dissertation was to further improve these measures.

First, the effect of contextualization on personality measures was investigated. For ethnic majority students, completely contextualized measures are more predictive of performance and counterproductive behaviors. Additionally, completely contextualized measures—overall—generate more positive participant reactions. However, we could not replicate the FoR effect for pharmacy assistants and for ethnic minority students. Therefore, the FoR effect is only supported for a large majority group and not for more specific groups. Also, contextualization of personality inventories does not seem to solve the differential predictive validity dilemma shown by personality measures of academic performance. Contextualization thus appears to be a useful method, but it does not solve some important limitations of personality inventories that still need to be addressed.

Second, we investigated the relations between the HEXACO personality model and the Spherical representation of vocational interests. The results showed that people high on Honesty-Humility are more likely to be drawn to Helping vocations. Additionally, people high on Openness to Experience and Extraversion are more likely to be drawn to activities and jobs high on Prestige interests.

Third, in the last empirical chapter, the usefulness of other-ratings of vocational interests was explored. Self- and other-rated interests showed mostly comparable relations to those of personality measures. On the whole, parents and children were able to rate each other's interests fairly accurately. It appears that others can be used as a valuable source of information in the measurement of someone's vocational interests.

Altogether, personality and vocational interests are two related individual differences that are measured in vast amounts of people for major life decisions. Therefore, improving the quality of personality and vocational interest measurement is of great importance. Based on this dissertation, and with the caveats noted above, we would recommend scientists and practitioners to further explore complete contextualization as a useful method to improve the quality of personality measurement and to further explore the usefulness of other-ratings of vocational interests to improve the quality of vocational interest measurement.

**Nederlandse samenvatting**

**(Dutch summary)**



8



Persoonlijkheids- en interessevragenlijsten voorspellen gedrag en prestaties op school en op het werk. Persoonlijkheidsvragenlijsten worden vaak gebruikt om te bepalen of iemands karakter goed aansluit bij de eisen van een studie of baan. Werkt iemand bijvoorbeeld wel nauwkeurig genoeg om een administratieve baan aan te kunnen? Interessevragenlijsten worden vaak gebruikt als iemand belangrijke keuzes voor zijn/haar verdere leven moet maken, bijvoorbeeld bij het zoeken naar een nieuwe beroepsrichting of bij het kiezen van een opleiding. Persoonlijkheids- en interessevragenlijsten worden dus gebruikt voor belangrijke beslissingen over personen, en ze worden daarbij in grote aantallen gebruikt. Gebaseerd op schattingen van de wereldwijde omzet, vullen minimaal 16 miljoen mensen per jaar een persoonlijkheidsvragenlijst in en hoogstwaarschijnlijk zelfs veel meer (The Economist, 2013). Meer dan 50% van alle Nederlandse studenten vult een interessevragenlijst in voordat ze hun studie kiezen (Markteffect, 2011). De kwaliteit van persoonlijkheids- en interessevragenlijsten is dan ook van groot belang. Deze metingen worden immers niet alleen door veel mensen ingevuld, maar het gaat ook om belangrijke beslissingen. In dit proefschrift is daarom onderzocht hoe persoonlijkheidsvragenlijsten beter prestaties kunnen voorspellen en hoe een driedimensionaal model van interesses met behulp van zelf- en anderbeoordelingen gemeten kan worden.

Bij het invullen van een persoonlijkheids- of interessevragenlijst beantwoordt iemand een lijst met stellingen (meestal items genoemd). Een persoonlijkheidsvragenlijst onderzoekt wat voor karakter iemand heeft en beantwoordt de vraag: *Wie ben ik?* Hiervoor gebruikt een persoonlijkheidsvragenlijst items zoals *“Ik houd alles netjes”*. Kandidaten kunnen bij dit item aangeven in hoeverre de stelling op hen van toepassing is. Van de in totaal vijf onderzoeken in dit proefschrift, betreffen de eerste drie het verbeteren van persoonlijkheidsvragenlijsten. Deze verbetering houdt in dat in deze onderzoeken de items van bestaande persoonlijkheidsvragenlijsten zijn aangepast, met de bedoeling een betere voorspelling te bewerkstelligen van school- en werkprestaties.

Een interessevragenlijst onderzoekt wat voor werk of studie iemand wil doen en beantwoordt de vraag: *Hoe leuk vind ik...?* Hiervoor gebruikt een interessevragenlijst items zoals *“Computerprogramma’s schrijven”* of *“Haarstylist”*. Kandidaten kunnen bij deze items aangeven hoe leuk ze de activiteit of het beroep vinden. De laatste twee onderzoeken in dit proefschrift betreffen het gebruik van interessevragenlijsten. Het eerste onderzoek daarover bestudeert de samenhang tussen persoonlijkheids- en

interessevragenlijsten en toetst een driedimensionaal interessemodel in een Nederlandse steekproef. Het tweede onderzoek bekijkt of het mogelijk is om met een interessevragenlijst de interesses van een andere persoon te beoordelen. Voor sommige deelnemers zijn interessevragenlijsten die ze over zichzelf invullen niet heel informatief, andere personen zouden dan een waardevolle bron van informatie kunnen zijn.

Deze samenvatting begint met de bevindingen van de eerste drie onderzoeken naar de voorspellende waarde van persoonlijkheidsvragenlijsten. Vervolgens worden de onderzoeken naar interesses beschreven. Hierna worden de sterke punten, beperkingen en praktische implicaties van de studies beschreven. Deze samenvatting sluit af met een conclusie.

### **1. Contextualisatie van persoonlijkheidsvragenlijsten**

Dit proefschrift begint met drie onderzoeken naar het verbeteren van persoonlijkheidsvragenlijsten door middel van contextualisatie. Persoonlijkheid is in deze onderzoeken benaderd volgens het HEXACO-model van persoonlijkheid (Lee & Ashton, 2004), dat ook weleens het zesfactor-model van persoonlijkheid wordt genoemd. In dit model wordt er vanuit gegaan dat mensen in meer of mindere mate zes stabiele en onafhankelijke persoonlijkheidstrekken bezitten: Integriteit (*Honesty-Humility*), Emotionaliteit (*Emotionality*), Extraversie (*eXtraversion*), Verdraagzaamheid (*Agreeableness*), Consciëntieusheid (*Conscientiousness*) en Openheid voor ervaringen (*Openness to experience*). Deze zes persoonlijkheidskenmerken liggen volgens het HEXACO-model ten grondslag aan gedrag en de daaruit voortvloeiende prestaties. Echter, gedrag wordt niet alleen bepaald door persoonlijkheidskenmerken, maar het hangt ook af van de situatie waarin iemand zich bevindt (Mischel & Shoda, 1995). Contextualisatie is een methode waarbij een situatie wordt toegevoegd aan een situatieloze (algemene) persoonlijkheidsvragenlijst (e.g., Mount, Barrick, & Strauss, 1994; Schmit, Ryan, Stierwalt, & Powell, 1995). Op deze manier wordt de wisselwerking tussen persoonlijkheid en situatie verwerkt in een persoonlijkheidsvragenlijst. Een item van een algemene persoonlijkheidsvragenlijst is bijvoorbeeld:

Voorbeelditem 1. Ik stel hoge eisen aan mijzelf

Als het bovenstaande item wordt gecontextualiseerd door een schoolsituatie toe te voegen, dan verandert het item bijvoorbeeld in:

Voorbeelditem 2. Ik stel hoge eisen aan mijzelf op school

of in:



### Voorbeelditem 3. Ik ben pas tevreden als ik hoge cijfers haal

Bij contextualisatie wordt de toegevoegde situatie een *Frame-of-Reference* (FoR) genoemd. In het voorbeeld is dat dus een schoolsituatie, dit zou ook een andere situatie kunnen zijn zoals werk of thuis. Het toevoegen van een FoR verhoogt de voorspellende waarde van een persoonlijkheidsvragenlijst voor prestaties in het FoR (bv. Schmit et al., 1995). Bijvoorbeeld, als een schoolsituatie is toegevoegd aan een persoonlijkheidsvragenlijst, voorspelt deze iemands schoolprestaties beter (en voorspelt deze overigens prestaties buiten het FoR slechter; Lievens, De Corte, & Schollaert, 2008). Deze toegevoegde voorspellende waarde wordt ook weleens het *Frame-of-Reference effect* genoemd (FoR-effect; Schmit et al., 1995).

Voorbeelden 2 en 3 laten zien dat een persoonlijkheidsvragenlijst op verschillende manieren gecontextualiseerd kan worden. Eerdere onderzoeken hebben contextualisatie op drie verschillende manieren uitgevoerd. De eerste en minst sterke vorm van contextualisatie gebeurt door middel van de instructie aan de kandidaat vooraf aan het invullen van de vragenlijst. Een kandidaat wordt bij deze methode gevraagd om te denken aan een bepaald FoR bij het beantwoorden van algemene items (bv. Hunthausen, Truxillo, Bauer, & Hammer, 2003). De tweede en de meest gebruikte contextualisatiemethode is het toevoegen van een zogenaamde *tag* aan de bestaande items, zoals "... *op school*", of "... *op het werk*" (zie voorbeelditem 2; bv. Bing, Whanger, Davison, & VanHook, 2004; Lievens et al., 2008; Robie, Schmit, Ryan, & Zickar, 2000; Schmit et al., 1995). Echter, Lievens et al. (2008) suggereerden dat het mogelijk is dat de voorspellende waarde van persoonlijkheidsvragenlijsten nog verder verhoogd kan worden als er meer context wordt toegevoegd dan alleen een *tag*. De derde en minst onderzochte contextualisatiemethode is volledige/complete contextualisatie (Butter & Born, 2012; Murtha, Kanfer, & Ackerman, 1996; Pace & Brannick, 2010). Bij deze methode worden de items volledig aan de situatie aangepast (zie voorbeeld 3).

#### **1.1 Voorspelt een volledig gecontextualiseerde persoonlijkheidsvragenlijst school- en werkprestaties beter dan een getagde persoonlijkheidsvragenlijst?**

In hoofdstukken twee en drie is de suggestie van Lievens et al. (2008) onderzocht dat een volledig gecontextualiseerde persoonlijkheidsvragenlijst prestaties beter voorspelt dan een getagde persoonlijkheidsvragenlijst. In

**hoofdstuk twee** hebben 531 hogeschoolstudenten een algemene, getagde en volledig gecontextualiseerde persoonlijkheidsvragenlijst ingevuld. In dit onderzoek zijn hiervoor twee verschillende algemene persoonlijkheidsvragenlijsten gebruikt, namelijk de Multiculturele Persoonlijkheidstest Big Six (MPT-BS; NOA, 2009; De Vries, De Vries, & Born, 2011) en de Nederlandstalige HEXACO Personality Inventory-Revised (HEXACO-PI-R; De Vries, Ashton, & Lee, 2009; De Vries, De Vries, & Born, 2011). Van beide vragenlijsten werden twee FoR-versies ontwikkeld voor dit onderzoek, te weten een getagde en een volledig gecontextualiseerde versie. Bij de getagde vragenlijst werd achter elk item “op school” toegevoegd. Bij de volledig gecontextualiseerde vragenlijst werd elk item volledig veranderd zodat het zo goed mogelijk aansloot bij de situatie van een student. De studenten werden willekeurig toegewezen aan één van de twee vragenlijsten (de MPT-BS of de HEXACO-PI-R) en vulden daarna de algemene vragenlijst en beide FoR-vragenlijsten in. Vervolgens vulden de studenten een vragenlijst in over de mate waarin zij Contraproductief Academisch Gedrag (CAG; Marcus, Lee, & Ashton, 2007) vertonen. Ten slotte werden hun gemiddelde cijfers verzameld uit de database van de hogeschool als maat voor hun studieprestaties.

De resultaten in hoofdstuk twee laten zien dat Consciëntieusheid studieprestaties voorspelt en dat Integriteit CAG voorspelt. De resultaten laten verder zien dat de getagde schalen, vergeleken met de algemene (situatieloze) persoonlijkheidsschalen, studieprestaties en CAG beter voorspellen (bv. getagde Consciëntieusheid voorspelt studieprestaties beter dan algemene Consciëntieusheid). Bovendien voorspellen de volledig gecontextualiseerde schalen studieprestaties en CAG nog beter dan de getagde schalen en de algemene schalen (bv. volledig gecontextualiseerde Consciëntieusheid voorspelt studieprestaties beter dan algemene Consciëntieusheid en getagde Consciëntieusheid). Deze studie bevestigt met deze bevindingen het eerder vastgestelde FoR-effect (bv. Shaffer & Postlethwaite, 2012) en laat zien dat volledige contextualisatie het sterkste FoR-effect teweeg brengt.

**Hoofdstuk drie** beschrijft de tweede studie naar het FoR-effect. In deze studie zijn gegevens verzameld van 139 apothekersassistentes van 33 verschillende apotheken in de Randstad. De assistentes werd gevraagd om drie verschillende versies van de HEXACO-PI-R in te vullen. Net als de studenten, vulden de assistentes een algemene, een getagde en een volledig gecontextualiseerde persoonlijkheidsvragenlijst in. Vervolgens werd de

assistentes ook gevraagd om een vragenlijst in te vullen over hun werktevredenheid (de Abridged Job In General scale; Bowling Green State University, 2008). Ten slotte werden over alle assistentes prestatiebeoordelingen ingevuld door hun leidinggevende apotheker. De resultaten laten zien dat hun algemene Consciëntieusheid hun werkprestaties enigszins voorspelt en dat hun Integriteit (bestaande uit bijvoorbeeld bescheidenheid en behoefte aan status en bezit) werktevredenheid voorspelt. Echter, getagde en volledig gecontextualiseerde Consciëntieusheid voorspellen werkprestaties niet, zo bleek uit de resultaten. Getagde en volledig gecontextualiseerde Integriteit voorspellen wel werktevredenheid, maar geen van beide voorspelt werktevredenheid beter dan algemene Integriteit. In deze studie is dus zowel voor Consciëntieusheid als voor Integriteit geen FoR-effect gevonden en geen van beide contextualisatiemethoden voorspelt beter dan de andere. Sterker nog, de resultaten tonen een omgekeerd FoR-effect aan, waarin de FoR-vragenlijsten slechter voorspellen dan de algemene persoonlijkheidsvragenlijst.

Hoofdstukken twee en drie schetsen elk een ander beeld van het FoR-effect. De studie in hoofdstuk twee bevestigde het FoR-effect en laat zien dat een volledig gecontextualiseerde persoonlijkheidsvragenlijst gedrag en prestaties beter voorspelt dan een getagde vragenlijst. De studie in hoofdstuk drie toont aan dat alleen de algemene persoonlijkheidsvragenlijst (met de persoonlijkheidsdimensie Consciëntieusheid) een relatie heeft met werkprestaties en de FoR-vragenlijsten juist niet. Het FoR-effect is een robuust fenomeen dat niet door de bevindingen van hoofdstuk drie als volledig onjuist beschouwd dient te worden. Echter, de bevindingen van hoofdstuk drie moeten ook niet terzijde worden gelegd. Zeker omdat een recente studie (Robie & Risavy, 2016) een vergelijkbaar nul-effect heeft gevonden.

Wij stellen twee mogelijke verklaringen voor onze nulbevindingen in hoofdstuk drie voor. Deze mogelijke verklaringen worden ingegeven door de bevinding dat de apothekersassistentes zich gemiddeld hoger op Consciëntieusheid beoordeelden met de FoR-vragenlijsten dan op de algemene vragenlijsten, en een kleinere standaarddeviatie vertoonden op de FoR-scores. Ten eerste is het mogelijk dat, ondanks het feit dat vertrouwelijkheid werd benadrukt, de assistentes het gevoel hadden dat ze beoordeeld werden. Sommige assistentes gaven tijdens de studie bijvoorbeeld aan dat ze zich afvroegen of de resultaten zichtbaar zouden zijn voor hun leidinggevenden. Door dit gevoel hebben de assistentes mogelijk de vragenlijsten sociaal wenselijk ingevuld waardoor de voorspellende waarde

kan zijn afgenomen. Omdat de FoR-vragenlijsten duidelijk meer aan werk zijn gerelateerd dan de algemene vragenlijst, bestaat de kans dat assistentes bij de FoR vragenlijst hebben ingevuld “hoe ze zich *zouden moeten* gedragen” en niet “hoe ze zich gedragen”. Hierdoor kan de voorspellende waarde van de FoR-vragenlijsten minder geworden zijn dan van die van de algemene vragenlijst. Ten tweede is het mogelijk dat de voorspellende waarde van FoR-Consciëntieusheid lager is dan algemene Consciëntieusheid omdat de assistentes in een zeer gereguleerde omgeving werken. In een apotheek vinden de meeste handelingen plaats volgens strikte procedures; de assistentes kunnen dus niet veel variëren in gedrag. Hierdoor is het mogelijk dat Consciëntieusheid in een werksituatie heel weinig verschillen laat zien tussen assistentes, waardoor FoR-Consciëntieusheid werkprestaties slechter voorspelt dan algemene Consciëntieusheid. Algemene Consciëntieusheid kan ook belangrijk zijn voor werkprestaties, maar valt niet binnen deze strikte protocollen die de variatie tussen assistentes beperken.

## **1.2 Denken deelnemers positiever over een algemene, een *getagde*, of een volledig gecontextualiseerde persoonlijkheidsvragenlijst?**

Bij het inzetten van vragenlijsten is het belangrijk om te weten wat gebruikers ervan vinden. Als de gebruikerservaring tegenvalt, kan een deelnemer het gevoel krijgen dat de test niet nuttig is en daardoor weerstand opbouwen tegen de resultaten of zelfs de resultaten aanvechten. In de studies in hoofdstukken twee, drie en vier werden daarom verschillende deelnemersreacties op algemene persoonlijkheidsvragenlijsten en FoR-vragenlijsten gemeten. De deelnemers van de studies in hoofdstukken twee en drie werd gevraagd de algemene en de twee FoR-vragenlijsten te beoordelen op *leuk vinden*, *indruksvaliditeit* en *geschatte voorspellende waarde*. In hoofdstuk vier werden 309 hogeschoolstudenten gevraagd om de algemene en volledig gecontextualiseerde MPT-BS te beoordelen in termen van dezelfde drie deelnemersreacties.

In alle genoemde studies beoordeelden de deelnemers hoe leuk ze de vragenlijsten vonden aan de hand van vier items, bijvoorbeeld ‘Ik vind het leuk om dit soort vragenlijsten in te vullen’. Ze beoordeelden alle persoonlijkheidsvragenlijsten net iets hoger dan het schaalgemiddelde, op een schaal die van 1 (helemaal niet mee eens) tot 7 (helemaal mee eens) liep. De *getagde* vragenlijst werd als *minst leuk* beoordeeld. Dit is niet een verrassende bevinding, omdat deze vragenlijst heel repetitief is doordat steeds dezelfde tag wordt gebruikt. Door de deelnemers aan de studie in hoofdstuk twee werd de volledig gecontextualiseerde vragenlijst minder leuk gevonden dan de

algemene vragenlijst en even leuk door de deelnemers aan de studies in hoofdstukken drie en vier.

De indrukvaliditeit gaat over de vraag hoe toepasbaar deelnemers de vragenlijst vinden voor hun rol/taken. Deze validiteit werd gemeten met vijf items, bijvoorbeeld 'De inhoud was duidelijk gerelateerd aan mijn werk/school'. In alle studies werd de volledig gecontextualiseerde test als meest indrukvalide beoordeeld. De getagde vragenlijst werd als meer indrukvalide beoordeeld dan de algemene vragenlijst door de deelnemers aan de studie in hoofdstuk twee, en even indrukvalide door de deelnemers aan de studie in hoofdstuk drie. Contextualisatie lijkt dus de indrukvaliditeit te verhogen en dit effect lijkt het sterkst voor volledig gecontextualiseerde persoonlijkheidsvragenlijsten.

De ingeschatte voorspellende waarde gaat over de vraag hoe voorspellend deelnemers de vragenlijst vinden voor hun prestaties. Dit werd gemeten met vijf items, bijvoorbeeld 'Mijn prestatie op deze vragenlijst geeft aan hoe goed ik ben in mijn werk'. De studenten van de studie in hoofdstuk twee vonden de volledig gecontextualiseerde vragenlijst het meest voorspellend. De apothekersassistenten vonden de getagde vragenlijst het meest voorspellend, en de studenten van de studie in hoofdstuk vier vonden de algemene versie en de gecontextualiseerde versie even voorspellend. Als we deze bevindingen samennemen, blijkt contextualisatie de ingeschatte voorspellende waarde te verhogen, maar het effect wordt niet systematisch gevonden en is niet van toepassing op één bepaalde contextualisatiemethode.

Samengevat lijkt contextualisatie een positief effect te hebben op gebruikersreacties. De uitzondering hierop is dat deelnemers getagde vragenlijsten (de makkelijkste manier van contextualisatie) consequent minder leuk vinden. Bovendien worden volledig gecontextualiseerde vragenlijsten positiever beoordeeld dan getagde vragenlijsten. Als het dus gaat om het verbeteren van gebruikersreacties, dan is volledige contextualisatie nuttiger dan het toevoegen van een tag aan een vragenlijst.

### **1.3 Vermindert een volledig gecontextualiseerde persoonlijkheidsvragenlijst de differentiële voorspellende waarde van persoonlijkheidsvragenlijsten tussen etnische groepen?**

De voorspellende waarde van persoonlijkheidsvragenlijsten kan verschillen tussen groepen. Dit fenomeen wordt differentiële validiteit genoemd: Differentiële validiteit houdt in dat de vragenlijst beter voorspelt voor de ene groep dan voor een andere groep. Sommige studies rapporteerden dat

persoonlijkheidsvragenlijsten differentiële validiteit vertonen voor autochtone ten opzichte van niet-westerse allochtone groepen (De Meijer, Born, Terlouw, & Van der Molen, 2008; De Vries, Born, & De Vries, 2012). Verschillende auteurs hebben gesuggereerd dat de differentiële validiteit tussen autochtone en allochtone groepen verminderd kan worden door contextualisatie van persoonlijkheidsvragenlijsten (Church, 2010; De Vries et al., 2012).

Contextualisatie zorgt ervoor dat alle items met hetzelfde FoR worden beantwoord. Het zou kunnen dat groepen verschillende contexten gebruiken bij het beantwoorden van items in een persoonlijkheidsvragenlijst (bv. de ene groep denkt vooral aan school en de andere groep vooral aan thuis). Voor een groep die vooral een criterium-relevant FoR gebruikt bij het invullen van een algemene persoonlijkheidsvragenlijst (bv. school FoR als het doel is om schoolprestatie te voorspellen), kan deze algemene vragenlijst mogelijk beter de uitkomst voorspellen dan voor een groep die vooral een niet-relevant FoR gebruikt. Door contextualisatie gebruikt iedereen hetzelfde FoR bij het beantwoorden van alle items. Contextualisatie zou op deze manier differentiële validiteit kunnen verminderen.

In **hoofdstuk vier** is onderzocht of contextualisatie inderdaad de differentiële validiteit beïnvloedt. In deze studie vulden 326 studenten (190 autochtonen en 110 niet-westerse allochtonen van drie verschillende hogescholen) een algemene en een volledig gecontextualiseerde persoonlijkheidsvragenlijst in. Alle studenten werd vervolgens gevraagd om zelf hun gemiddelde cijfer in te schatten. Ten slotte werd de studenten gevraagd om een acculturatievragenlijst in te vullen. Deze vragenlijst meet 'het behouden van de moedercultuur' en 'het overnemen van de Nederlandse cultuur'. Verder werd het ongewogen gemiddelde cijfer van alle studenten uit de database van de hogescholen gehaald.

De resultaten laten het FoR-effect zien voor de hele groep studenten: volledig gecontextualiseerde Consciëntieusheid voorspelt het gemiddelde studiecijfer beter dan algemene Consciëntieusheid. Echter, na verdere analyses blijkt dat de algemene en gecontextualiseerde persoonlijkheidsvragenlijst alleen het objectief geregistreerde gemiddelde studiecijfer van de autochtone groep voorspellen en niet dat van de allochtone groep. Deze resultaten impliceren dat 1) het FoR-effect niet voor de allochtone groep wordt gerepliceerd en 2) zowel de algemene als de gecontextualiseerde persoonlijkheidsvragenlijst differentiële voorspellende waarde vertonen. Omdat de persoonlijkheidsvragenlijsten geen relatie vertonen met het gemiddeld cijfer van de allochtone groep is er zelfs sprake van de meest

extreme vorm van differentiële validiteit: *single group validity*. Dit betekent dat een vragenlijst voor de ene (hier: autochtone) groep wel voorspelt, maar voor een andere (hier: allochtone) groep niet. De resultaten geven dus aan dat differentiële validiteit niet opgelost kan worden door contextualisatie van persoonlijkheidsvragenlijsten.

Echter, bij de studie in hoofdstuk vier was niet alleen het gemiddelde cijfer uit de studentendatabase gehaald, maar de studenten was ook gevraagd om hun gemiddelde cijfer zelf te rapporteren. Voor de autochtone groep bleek Consciëntieusheid het daadwerkelijke gemiddelde cijfer even goed te voorspellen als het zelf-gerapporteerde gemiddelde cijfer. Voor de allochtone groep voorspelde geen enkele persoonlijkheidstrek het daadwerkelijke gemiddelde cijfer, maar bleken er enkele persoonlijkheidstreken (voornamelijk Consciëntieusheid) het zelf-gerapporteerde gemiddelde cijfer te voorspellen. Verder bleek er bij het voorspellen van het zelf-gerapporteerde cijfer alleen een FoR-effect op te treden voor de autochtone groep en niet voor de allochtone groep.

Voor de niet-westerse allochtone studenten heeft persoonlijkheid dus een sterkere relatie met het zelf-gerapporteerde cijfer dan met het echte gemiddelde cijfer. Deze sterkere samenhang kan zijn veroorzaakt doordat de persoonlijkheidsmeting en het zelf-gerapporteerde cijfer beide door de student zelf zijn gerapporteerd. Mogelijk is er een responsstijl die invloed uitoefent op de persoonlijkheidsmetingen en het zelf-gerapporteerde cijfer van de niet-westerse allochtone studenten waardoor deze metingen aan elkaar zijn gerelateerd.

## **2. Interesses meten**

In de laatste twee studies die dit proefschrift rapporteert zijn interessevragenlijsten onderzocht. Interesses zijn in deze onderzoeken gemeten met de *Personal Globe Inventory* (PGI; Tracey, 2002), die in het Nederlands is vertaald. De PGI meet interesses volgens het Sferische interessemodel (Tracey & Rounds, 1996). Eerdere modellen van interesses onderscheiden twee dimensies waarop mensen verschillen in hun interesses (Prediger, 1982). De eerste dimensie is Data versus Ideeën (*Data versus Ideas*). Personen die meer geïnteresseerd zijn in Data werken graag met feiten en getallen; personen die meer geïnteresseerd zijn in Ideeën werken graag onderzoekend en creatief. De tweede dimensie is Mensen versus Dingen (*People versus Things*). Personen die meer geïnteresseerd zijn in Mensen doen graag werk waarbij ze anderen verzorgen of beïnvloeden;

personen die meer geïnteresseerd zijn in Dingen werken graag met machines, gereedschappen en computers. Het Sferische model heeft de dimensie Prestige-interesse (*Prestige interests*) toegevoegd aan de twee eerder aangetoonde dimensies. Personen die hoog scoren op Prestige-interesse werken graag in complexe banen waarvoor veel opleiding nodig is; personen die laag scoren op Prestige-interesse werken liever in eenvoudiger banen waarvoor minder geleerd hoeft te worden.

## 2.1 Prestigieuze beroepsinteresse: Aan welke persoonlijkheidsdimensies is Prestige-interesse gerelateerd?

**Hoofdstuk vijf** bespreekt een studie waarin 656 mensen de Nederlandse PGI en Nederlandse HEXACO-PI-R (100 item versie; De Vries et al., 2009) invulden. Veel onderzoek is de relatie al nagegaan tussen interesses en persoonlijkheid, maar er zijn nog geen studies bekend die de meest recente modellen van interesses (het Sferische model) en persoonlijkheid (het HEXACO-model) aan elkaar gerelateerd hebben. Deze vergelijking is extra interessant omdat de nieuwe dimensies van het Sferische model (Prestige-interesse) en het HEXACO-model (Integriteit) conceptueel negatief aan elkaar gerelateerd lijken te zijn: Het is namelijk mogelijk dat hoog prestigieuze banen meer aanzien genieten dan laag prestigieuze banen, waardoor mensen met veel behoefte aan status geïnteresseerd zijn in prestigieuze banen. Behoeft aan status en bezit is een negatief facet van Integriteit: mensen met een hoge behoefte aan status scoren lager op Integriteit. Om deze redenen is het mogelijk dat Prestige-interesse negatief gerelateerd is aan Integriteit.

Allereerst is de kwaliteit van de—voor dit onderzoek—vertaalde PGI bestudeerd. Deze vertoonde goede psychometrische eigenschappen: de structuur van de vragenlijst was niet veranderd ten opzichte van het Amerikaanse origineel en de betrouwbaarheden van de schalen waren hoog. De vertaling bleek dus geslaagd te zijn. Vervolgens is de relatie tussen Prestige-interesse en persoonlijkheid onderzocht.

Hoewel het de verwachting was dat Prestige-interesse en Integriteit negatief met elkaar zouden samenhangen, tonen de resultaten geen relatie aan. Mensen die complexe en veeleisende beroepen leuk vinden, zijn dus *niet* meer geïnteresseerd in status en bezit of minder oprecht, eerlijk en bescheiden. Prestige-interesse is wel positief gerelateerd aan Openheid voor ervaringen en, in mindere mate, aan Extraversie. Dit betekent dat Prestige-interesse voornamelijk gerelateerd is aan verbeelding en nieuwsgierigheid en



enigszins aan sociale energie. Mensen die openstaan voor nieuwe ervaringen, worden aangetrokken tot ingewikkelde banen die veel moeite vergen en veel training vereisen. Openheid voor ervaringen is de persoonlijkheidsdimensie die gerelateerd is aan opleidingsniveau en daarom is het niet verrassend dat mensen die hoog scoren op Openheid banen interessant vinden waarvoor veel training nodig is. Banen hoog op Prestige-interesse zijn complex, veeleisend en dynamisch. De vereiste moeite en energie voor deze banen zou kunnen verklaren waarom extraverte personen meer tot deze banen worden aangetrokken.

## **2.2 Tonen anderoordelen van interesses *zelf-ander overeenstemming, gelijkens en verwachte gelijkens?***

Over het algemeen worden interesses gemeten door een vragenlijst die iemand over zichzelf invult. **Hoofdstuk zes** bespreekt een studie over 271 adolescenten (van 15-35 jaar oud) en hun ouders die hun eigen en elkaars interesses hebben beoordeeld. Elke deelnemer vulde eerst de PGI over zichzelf in en vervolgens over de ander (de eigen ouder of het eigen kind). Er zijn in deze studie dus zelfbeoordelingen en anderbeoordelingen van interesses verzameld. Voor deze studie is de korte vorm van de PGI gebruikt die bestaat uit 40 items. Vervolgens is de overeenstemming tussen de verzamelde zelf- en anderbeoordelingen berekend. De hoofdvraag bij dit onderzoek was: "Kunnen de interesses beoordeeld worden door een ander?"

Om de hoofdvraag te beantwoorden zijn drie maten van overeenstemming berekend tussen zelf- en anderbeoordelingen van interesses. Ter illustratie, stel dat dochter Alexa en moeder Bea als duo hebben deelgenomen aan dit onderzoek. Als eerste is de *zelf-ander overeenstemming* berekend tussen de zelfbeoordeling van dochter Alexa en de anderbeoordeling van moeder Bea (over de interesses van haar dochter). Bij een hoge zelf-ander overeenstemming schatten Alexa en Bea de interesses van Alexa hetzelfde in: ze hebben dan hetzelfde beeld van de interesses van Alexa. De zelf-ander overeenstemming is vervolgens ook berekend voor de interesses van Bea. Deze is berekend aan de hand van de zelfbeoordeling van Bea en de anderbeoordeling van Alexa (over de interesses van haar moeder, Bea). Ten tweede is de *gelikens* in interesses berekend tussen Alexa en Bea door de overeenstemming te berekenen tussen de zelfbeoordeling van Alexa en de zelfbeoordeling van Bea. Bij een hoge gelijkens hebben Alexa en haar moeder Bea voornamelijk dezelfde interesses. Ten derde is de *verwachte gelijkens* in interesses berekend door de overeenstemming te berekenen tussen de zelf- en anderbeoordeling van

Alexa of Bea. In het geval van Alexa: bij een hoge verwachte gelijkenis beoordeelt Alexa haar eigen interesses hetzelfde als dat zij de interesses van haar moeder beoordeelt. Ze verwacht in dat geval dat haar interesses vergelijkbaar zijn aan die van haar moeder. De verwachte gelijkenis is ook uitgerekend voor Bea over haar dochter. Tot nu toe hebben slechts twee studies onderzocht of interesses door anderen beoordeeld kunnen worden (Nauta, 2012; Nelling, Kandler, & Riemann, 2015). Deze studies gebruikten echter alleen zelf-ander overeenstemming.

De overeenstemming tussen de verschillende interessebeoordelingen werd uitgerekend met zogeheten profielcorrelaties, die een opsomming vormen van alle losse correlaties tussen de verschillende interessedimensies. De profielcorrelaties laten zien dat bij het beoordelen van interesses (bij ouder-kind paren) de *zelf-ander overeenstemming* hoog is, de *verwachte gelijkenis* gemiddeld is, en de daadwerkelijke *gelijkenis* laag is. Dit betekent dat 1) ouders en hun kinderen de interesses van elkaar accuraat kunnen inschatten, omdat de zelf-ander overeenstemming hoog is, en dat 2) zij de gelijkenis met elkaar overschatten, omdat de verwachte gelijkenis hoger is dan de daadwerkelijke gelijkenis. Vervolganalyses laten zien dat duo's met hetzelfde geslacht (zoon-vader en dochter-moeder) de overeenstemming meer overschatten dan duo's met een gemengd geslacht. Beoordelaars van hetzelfde geslacht verwerken mogelijk hun eigen zelfbeeld meer in de beoordeling van de interesses van een ander. Hierdoor zijn ze wellicht minder informatief dan een beoordelaar van het andere geslacht.

De studie in hoofdstuk zes toont aan dat anderbeoordelingen gebruikt kunnen worden om interesses te meten. De zelf-ander overeenstemming, gelijkheid en verwachte gelijkheid van interessebeoordelingen, zoals gevonden in de studie uit hoofdstuk zes, zijn zeer vergelijkbaar met de resultaten van persoonlijkheidsonderzoek naar anderbeoordelingen (bv. De Vries, 2010; McCann, Lipnevich, Poropat, Wiemers, & Roberts, 2015; Watson & Clark, 1991). Er is echter veel meer onderzoek gedaan naar anderbeoordelingen van persoonlijkheid dan naar anderbeoordelingen van interesses. Op basis van de vergelijkbare resultaten kan worden verwacht dat overige bevindingen over anderbeoordelingen van persoonlijkheid mogelijk ook gelden voor anderbeoordelingen van interesses. Zo hebben Connely en Ones (2010) gevonden dat meerdere anderbeoordelingen van persoonlijkheid prestaties beter voorspellen dan één zelfbeoordeling. Op basis van onze resultaten kunnen we niets concluderen over de voorspellende waarde van anderbeoordelingen van interesses, maar het zou kunnen dat meerdere

anderbeoordelingen van interesses een match met een baan of opleiding beter voorspellen dan één zelfbeoordeling. Voor deze anderbeoordelingen van interesses kunnen ook vrienden (Nauta, 2012) gevraagd worden.

### **2.3 Is *profielverhoging* in interessevragenlijsten een inhoudelijke factor of een meetfout/artefact?**

Als de structuur van interessevragenlijsten wordt onderzocht, komt er altijd een dominerende hoofdfactor naar voren naast de interessedimensies. Dit betekent dat analyses naar de structuur van interesse(vragenlijsten) altijd een superdimensie vinden. Alle items van de interessevragenlijst laden dan op deze superdimensie, daarnaast laadt elk item ook altijd op een tweede interessedimensie (bv. Data versus Ideeën). Deze superdimensie wordt vaak profielverhoging (*profile elevation*) genoemd. Als iemand op alle items in een interessevragenlijst hoog scoort, heeft iemand veel profielverhoging. Door de tijd heen is er veel onderzoek gedaan naar profielverhoging bij interessevragenlijsten, maar er is nog geen consensus bereikt over wat deze profielverhoging precies betekent (Tracey, 2012).

Sommige onderzoekers hebben voorgesteld dat profielverhoging aangeeft of iemand over het algemeen geïnteresseerd is in allerlei beroepen en activiteiten, en dus een grote algemene interesse heeft (bv. Fuller, Holland, & Johnston, 1999). Omdat alle items van een interessevragenlijst op profielverhoging laden, lijkt het dat iemand met een verhoogd profiel in van alles geïnteresseerd is. Andere onderzoekers hebben juist voorgesteld dat profielverhoging een statistisch artefact is dat genegeerd moet worden of waar analyses voor moeten worden gecorrigeerd (bv. Tracey, 2012). Deze laatste groep onderzoekers vinden over het algemeen dat profielverhoging niet onderdeel is van interesses en dat het daarom geen rol mag spelen bij het bepalen van waarin iemand geïnteresseerd is.

In hoofdstuk vijf is de profielverhoging van interesses berekend en gerelateerd aan persoonlijkheid. De profielverhoging blijkt positief gecorreleerd te zijn aan Openheid voor ervaringen en negatief aan Integriteit. Personen die nieuwsgierig, creatief, minder eerlijk en minder bescheiden zijn, vertonen dus meer profielverhoging op interessevragenlijsten. Als vervolgens de profielverhoging statistisch werd verwijderd uit de dimensies van interesses dan vertonen de interesseschalen een andere relatie met persoonlijkheid. Deze bevindingen laten zien dat profielverhoging de relatie beïnvloedt die beroepsinteresses met andere variabelen hebben, en deze relatie soms zelf onderdrukt. Het lijkt dus verstandig om individuele interesseschalen te

corrigeren voor profielverhoging bij het bepalen van de relatie met andere variabelen, zoals persoonlijkheid.

In hoofdstuk zes is ook de mate van zelf-ander overeenstemming over profielverhoging berekend. Vervolgens is de zelf-ander overeenstemming over profielverhoging vergeleken met de zelf-ander overeenstemming over interesses. Als profielverhoging een vorm van interesse is, zou de zelf-ander overeenstemming van profielverhoging—net als de (andere) dimensies van beroepsinteresses—hoog moeten zijn. Echter, de resultaten laten zien dat profielverhoging een gemiddelde zelf-ander overeenstemming vertoont, terwijl de reguliere interesses een hoge mate van zelf-ander overeenstemming vertonen. Profielverhoging lijkt dus moeilijker waar te nemen dan interesses. Hiernaast vertoont profielverhoging een hoge verwachte gelijkensis terwijl reguliere interesses daarentegen een gemiddelde verwachte gelijkensis vertonen. Profielverhoging lijkt dus vooral te worden veroorzaakt door de manier waarop iemand een interessevragenlijst invult. Dit kan betekenen dat profielverhoging niet zo zeer een vorm van interesse (in van alles) is, maar eerder een manier is waarop mensen een vragenlijst invullen, ook weleens een responsstijl genoemd.

De resultaten in hoofdstuk zes ondersteunen de gedachte dat profielverhoging een artefact is bij het meten van beroepsinteresses. Echter, de resultaten in hoofdstuk vijf geven wel aan dat profielverhoging niet een willekeurige meetfout is omdat deze verhoging gerelateerd is aan Openheid voor ervaringen en (negatief) aan Integriteit. Eerder onderzoek naar persoonlijkheid heeft aangetoond dat responsstijlen gerelateerd zijn aan persoonlijkheidseigenschappen (bv. He & Van de Vijver, 2013; Vigil-Colet, Morales-Vives, & Lorenzo-Seva, 2013; Zettler, Lang, Hülshager, & Hilbig, 2015). Deze responsstijlen zijn bijvoorbeeld bevestigend antwoorden (ook wel *acquiescence* genoemd; de neiging om een vraag of stelling positief te beantwoorden), neutraal antwoorden (de neiging om het midden van een antwoordschaal te gebruiken) en extreem antwoorden (de neiging om de uitersten van een antwoordschaal te gebruiken). Zettler et al. (2015) vonden bijvoorbeeld dat Integriteit een negatieve relatie heeft met extreem antwoorden. Het zou kunnen zijn dat Integriteit negatief gerelateerd is aan profielverhoging omdat integere personen niet de uiterste punten van de interesseschaal gebruiken. Hiernaast vonden He en Van de Vijver (2013) dat responsstijlen sterk gerelateerd zijn aan de algemene factor van persoonlijkheid, een controversiële superdimensie (bv. Ashton, Lee, De Vries, & Goldberg, 2009; Van der Linden, Te Nijenhuis, & Bakker, 2010) die in een

aantal persoonlijkheidsvragenlijsten wordt gevonden. Profielverhoging zou op dezelfde manier grotendeels verklaard kunnen worden door individuele verschillen in responsstijlen. Toekomstig onderzoek zou kunnen nagaan of dezelfde responsstijlen ten grondslag liggen aan profielverhoging van interesses en de algemene factor van persoonlijkheid die in sommige vragenlijsten gevonden wordt.

### **3. Sterke punten en beperkingen van de studies**

Dit proefschrift heeft een belangrijke bijdrage geleverd aan de ontwikkeling, validatie en toepassing van persoonlijkheids- en interessemeetmethoden. Er is een procedure ontwikkeld voor de volledige contextualisatie van algemene persoonlijkheidsvragenlijsten (hoofdstuk 2). Deze methode werd vervolgens toegepast op twee bestaande persoonlijkheidsvragenlijsten voor een school- en werkcontext. De ontwikkelde volledig gecontextualiseerde vragenlijsten behielden de originele factorstructuur en betrouwbaarheid en correleerden sterk met het origineel. We kunnen dus concluderen dat de contextualisatieprocedure uit hoofdstuk twee bruikbaar is voor onderzoekers en testontwikkelaars die hun eigen gecontextualiseerde vragenlijst willen ontwerpen. We raden wel aan om erop te letten dat items niet te ver gecontextualiseerd worden naar te specifieke situaties: dit zou een negatief effect kunnen hebben op de voorspellende waarde van de vragenlijst. Bijvoorbeeld, het item “*Ik blijf kalm tijdens een beroving*” kan te ver zijn gecontextualiseerd, omdat de meeste mensen deze situatie nooit zijn tegengekomen.

Dit proefschrift heeft ook bijgedragen aan het meten van beroepsinteresses. De PGI is vertaald naar het Nederlands (zie hoofdstuk vijf). Vervolgens is met een grote en diverse Nederlandse steekproef bevestigd dat de structuur van de originele Amerikaanse vragenlijst behouden blijft. Daarna is (hoofdstuk zes) de verkorte versie van de PGI gebruikt om anderbeoordelingen van interesses te onderzoeken. Gebaseerd op de resultaten van deze laatste studie menen wij dat beroepsinteressevragenlijsten ook gebruikt kunnen worden om anderbeoordelingen te verzamelen. Met anderbeoordelingen van interesses kunnen studie- en beroepskeuzeadviseurs mensen uit de omgeving van een coachee op een gestructureerde manier betrekken bij het coachingsproces.

Een verdere bijdrage van dit proefschrift is dat de eerste drie studies (hoofdstukken een, twee en drie) hebben geprobeerd het FoR-effect te repliceren. Makel, Plucker en Hegarty (2012) hebben onderzocht hoeveel

gepubliceerde psychologiestudies eerdere studies repliceren. Aan het begin van hun publicatie halen ze John Tukey (1969, p. 84) aan: “*Confirmation comes from repetition. Any attempt to avoid this statement leads to failure and probably to destruction.*” Vervolgens tonen Makel et al. aan dat slechts 2,39% van de gepubliceerde psychologiestudies heeft aangegeven een replicatie te zijn van eerder werk. Het FoR-effect is inmiddels bevestigd door een groot aantal studies. Toch konden twee van de studies in dit proefschrift dit effect niet (volledig) repliceren met verschillende steekproeven. Het FoR-effect werd voor autochtone studenten wel gevonden, maar niet bij apothekersassistenten en allochtone studenten. Wat betreft de nul-bevinding bij allochtone studenten is dit waarschijnlijk te wijten aan differentiële validiteit en niet zozeer aan de contextualisatie. Echter, de bevindingen bij de apothekersassistenten en recente bevindingen van Robie en Risavy (2016) geven aan dat het FoR-effect niet altijd optreedt.

Ten slotte is ook het gebruik van uitsluitend objectieve studieprestaties en leidinggevendendebeoordelingen een sterk punt van dit proefschrift. Voor alle onderzoeken in dit proefschrift zijn de maten van studieprestaties uit de database van de deelnemende organisaties gehaald. Werkprestaties zijn beoordeeld door de directe leidinggevenden van de deelnemers. De bevindingen in hoofdstuk vier laten duidelijk zien dat zelfbeoordelingen van (studie)prestaties tot andere resultaten kunnen leiden dan objectieve maten van prestaties.

Een beperking van dit proefschrift is dat alle studies cross-sectioneel van aard zijn. Hierdoor kunnen geen causale verbanden worden bevestigd. Deze beperking lijkt vooral relevant voor de studies naar het FoR-effect. Bij een recente longitudinale studie (Liu & Huang, 2015) werden buitenlandse studenten, die net hun studie de V.S. begonnen, gevraagd hun school (FoR-) Extraversie te beoordelen. Binnen vier maanden na aankomst in de V.S. werden de studenten nog twee keren gevraagd om hun FoR-Extraversie te rapporteren. De resultaten lieten zien dat gecontextualiseerde Extraversie van de buitenlandse studenten door de tijd heen veranderde. Mensen kunnen zich dus door de tijd heen binnen een bepaalde context anders (bv. extravert) gaan gedragen. Bovendien vonden Liu en Huang dat FoR-Extraversie (bij aanvang van de studie) én de verandering van FoR-Extraversie (tijdens de vier maanden na aankomst), elk een afzonderlijke voorspellende waarde hadden voor acclimatisering en uitvalintenties van de studenten. Ze concludeerden dat stabiele gecontextualiseerde persoonlijkheidstrekken door de tijd heen kunnen veranderen en dat deze verandering ook nuttige

informatie bevat. Een alternatieve verklaring zou echter kunnen zijn dat mensen een nieuwe context na enige tijd beter leren kennen en zo een ander beeld van zichzelf krijgen in deze context. Mensen vullen dan de vragenlijst over zichzelf anders in omdat hun zelfbeeld is veranderd. In dit geval verandert persoonlijkheid niet, maar verandert iemands beeld over de context en daardoor hoe iemand zich tot de context denkt te verhouden.

Een tweede beperking van dit proefschrift is dat de steekproeven voornamelijk bestaan uit hoogopgeleiden. Dit is een beperking die regelmatig voorkomt in arbeids- en organisatiepsychologisch onderzoek (Bergman & Jean, 2015). Hierdoor is het mogelijk dat een aantal belangrijke kenmerken—waardoor deze groepen op hoger niveau functioneren—oververtegenwoordigd zijn, zoals sociaaleconomische status, cognitieve verschillen en een hoogopgeleid sociaal netwerk. Dit beperkt de generaliseerbaarheid van de bevindingen in dit proefschrift tot ongeveer 43% van de Nederlandse bevolking. Het is verder mogelijk dat deze beperking vooral invloed heeft gehad op de resultaten van het Sferische interessemodel. Prestige-interesse is sterk gerelateerd aan opleidingsniveau, daarom zal vooral bij deze interessedimensie een variantiebeperking zijn opgetreden. Door deze beperkte variantie zijn de relaties van Prestige-interesse met persoonlijkheid (Openheid voor ervaringen en Extraversie) mogelijk anderszins onderschat.

#### **4. Praktische implicaties**

Een belangrijke praktische implicatie van dit proefschrift is dat volledig gecontextualiseerde persoonlijkheidsvragenlijsten positievere deelnemersreacties oproepen. Daarnaast lijken volledig gecontextualiseerde vragenlijsten ook beter te voorspellen. Let op dat de betere voorspellende waarde niet altijd werd teruggevonden, zoals besproken. In de praktijk raden wij het *niet* aan om getagde vragenlijsten te gebruiken, omdat deelnemers deze minder leuk lijken te vinden dan algemene of volledig gecontextualiseerde persoonlijkheidsvragenlijsten. Hoewel het aantrekkelijker lijkt om volledig gecontextualiseerde persoonlijkheidsvragenlijsten te gebruiken in plaats van algemene vragenlijsten, is het wel belangrijk om in overweging te nemen dat het eenmalig ongeveer 65 uur kost om een dergelijke vragenlijst te ontwerpen (exclusief het verzamelen van een normgroep). Ten opzichte van bestaande algemene persoonlijkheidsvragenlijsten lijkt de kleine toegevoegde voorspellende waarde van volledig gecontextualiseerde persoonlijkheidsvragenlijsten deze moeite niet waard, tenzij een zeer strikte selectie deze toegevoegde waarde kan verantwoorden. Volledig gecontextualiseerde vragenlijsten worden echter vaak leuker en relevanter

(indruksvalider) gevonden door deelnemers. Hierdoor neemt de kans op weerstand tegen een assessment/selectie waarschijnlijk af. Op basis van deze overwegingen lijkt het dus nuttig om volledig gecontextualiseerde persoonlijkheidsvragenlijsten te gebruiken. Als de toegevoegde situatie (bv. school) van toepassing is op een grote groep kandidaten dan is de ontwerptijd per deelnemer bovendien relatief gering.

Een tweede praktische implicatie van de bevindingen in dit proefschrift is dat algemene en volledig gecontextualiseerde persoonlijkheidsvragenlijsten beide differentiële validiteit lijken te vertonen voor verschillende etnische groepen. Deze bevindingen sluiten aan bij enkele eerdere studies die vonden dat persoonlijkheidsvragenlijsten mogelijk differentiële validiteit vertonen voor etnische minderheden in Nederland (De Meijer et al., 2008; De Vries et al., 2012). In tegenstelling tot de suggesties van sommige academici (Church, 2010; De Vries et al., 2012) verminderde contextualisatie de differentiële validiteit niet. Het is op dit moment echter niet duidelijk waarom persoonlijkheidsvragenlijsten wel voor de ene groep prestaties voorspellen en niet voor de andere groep. Het is mogelijk dat een derde variabele een rol speelt, die tot nu toe nog niet in beschouwing is genomen. Niet-westerse etnische minderheden lijken persoonlijkheidsvragenlijsten op andere manieren te beantwoorden dan de autochtone meerderheid (He & Van de Vijver, 2013). Niet-westerse etnische minderheden gebruikten bijvoorbeeld vaker het midden van een antwoordschaal. Deze antwoordstrategieën beïnvloeden mogelijk de voorspellende waarde van persoonlijkheidsvragenlijsten, vooral als de invloed van deze antwoordstrategieën sterker is bij niet-westerse minderheden. Vanwege de gevonden differentiële validiteit van persoonlijkheidsvragenlijsten lijkt het voor assessments van een cultureel diverse groep, verstandig om vooral instrumenten te gebruiken die minder differentiële validiteit vertonen, zoals *Situational Judgement Tests*, werkproeven en open vragen (Ployhart & Holtz, 2008). De bevindingen in dit proefschrift zijn overigens van toepassing op assessments voor coaching en begeleidingsdoeleinden (*low stakes*) en niet op assessments voor selectie (*high stakes*).

In hoofdstuk zes hebben we laten zien dat anderbeoordelingen van interesses gebruikt kunnen worden om iemands interesses te meten. Voor zover wij weten worden anderbeoordelingen van interesses bijna nooit in de praktijk gebruikt. Dit staat in sterk contrast tot het feit dat anderen vaak door middel van interviews en discussies worden betrokken bij beroeps- en studiekeuzes. De bevinding dat anderbeoordelingen van interesses een goede inschatting geven van iemands interesses—maar niet precies hetzelfde zijn—



creëert nieuwe mogelijkheden voor beroeps- en studiekeuzecoaching. Het gebruik van anderbeoordelingen stelt carriërecoaches in staat om bekenden op een gestructureerde en betrouwbare manier te betrekken bij het coachingsproces. Vooral als de deelnemer zelf weinig kennis heeft van beroepen en activiteiten kunnen bekenden met meer ervaring wellicht een goede bijdrage leveren aan de studie- of beroepskeuze. De bekenden kunnen mogelijk zelfs beter inschatten welke banen en activiteiten de deelnemer leuk vindt, omdat zij de deelnemer kennen én goed begrijpen wat de beroepen en activiteiten inhouden. Bijvoorbeeld, een beginnend student begrijpt misschien niet goed wat een accountant doet en zal het daardoor moeilijk vinden om in te schatten hoe leuk het is om deze baan te beoefenen. Een goede bekende, die wel begrijpt wat een accountant doet en de beginnend student goed kent, kan misschien beter inschatten hoe leuk de student het zou vinden om accountant te zijn. De beginnend student kan de anderbeoordelingen van de goede bekende vervolgens gebruiken als ondersteuning bij gesprekken met zijn studiekeuzecoach.

Ten slotte willen we graag aanbevelen om waakzaam te zijn voor tegenstrijdige motieven van de bekenden. Het is mogelijk dat bekenden de deelnemer niet beoordelen op wat ze denken dat de deelnemer leuk vindt, maar op wat ze willen dat de deelnemer gaat doen. Dit kan een rol spelen als de bekenden belang hebben in de toekomst van de deelnemer. Bijvoorbeeld, een ouder kan tegen een kunstzinnige studie zijn omdat “er geen geld te verdienen is met zo’n studie”.

## **5. Conclusie**

Een groot aantal mensen gebruikt persoonlijkheids- en interessevragenlijsten bij belangrijke levensbeslissingen. In de psychologie worden deze meetmethoden constant verbeterd. Ook dit proefschrift had als doel om verdere verbeteringen aan dergelijke vragenlijsten te onderzoeken.

Ten eerste is het effect van contextualisatie op persoonlijkheidsvragenlijsten onderzocht. Voor autochtone studenten waren compleet gecontextualiseerde vragenlijsten beter voorspellend voor studieprestaties dan algemene persoonlijkheidsvragenlijsten. Daarnaast werden compleet gecontextualiseerde vragenlijsten ook positiever ervaren door de deelnemers. Echter, het FoR-effect kon niet worden gerepliceerd voor allochtone studenten en apothekersassistentes. Het FoR werd dus alleen teruggevonden bij een grote meerderheidsgroep en niet voor kleinere specifieke groepen. Hiernaast verminderde contextualisatie niet de

differentiële validiteit van persoonlijkheidsvragenlijsten. Contextualisatie lijkt dus een nuttige methode, maar een paar belangrijke beperkingen van persoonlijkheidsvragenlijsten worden hiermee niet opgelost.

Ten tweede zijn de relaties tussen het HEXACO-persoonlijkheidsmodel en het Sferische model van interesses onderzocht. De resultaten lieten zien dat integere mensen *niet* meer aangetrokken zijn tot Prestigieuze banen. Extraverte en nieuwsgierige mensen lijken juist wel aangetrokken te worden door Prestigieuze banen en activiteiten.

Ten derde is de bruikbaarheid van anderbeoordelingen van interesses onderzocht. Ouders en hun kinderen konden elkaars interesses redelijk nauwkeurig beoordelen met een vragenlijst. Zelf- en anderbeoordelingen van interesses laten dus een hoge mate van overeenstemming zien die vergelijkbaar is met beoordelingen van persoonlijkheid. Het lijkt erop dat anderen een waardevolle bron van informatie kunnen zijn bij het optimaliseren van iemands studie- of beroepskeuze.

# References

- Andriessen, I., & Phalet, K. (2002). Acculturation and school success: A study among minority youth in the Netherlands. *Intercultural Education, 13*, 21-36.
- Arends-Tóth, J., & Van de Vijver, F. J. R. (2007). Acculturation attitudes: A Comparison of measurement methods. *Journal of Applied Social Psychology, 37*, 1462-1488.
- Arends-Tóth, J., & Van de Vijver, F. J. R. (2003). Multiculturalism and acculturation: Views of Dutch and Turkish-Dutch. *European Journal of Social Psychology, 33*, 249-266.
- Armstrong, I. A., & Anthony, S. F. (2009). Personality facets and RIASEC interests: An integrated model. *Journal of Vocational behavior, 75*, 346-359.
- Ashton, M. C., & Lee, K. (2008). The prediction of Honesty-Humility-related criteria by the HEXACO and five-factor models of personality. *Journal of Research in Personality, 42*, 707-721.
- Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18*, 139-152.
- 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.
- Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology, 87*, 707-721.
- Ashton, M. C., Lee, K., Marcus, B., & De Vries, R. E. (2007). German lexical personality factors: Relations with the HEXACO model. *European Journal of Personality, 21*, 23- 43.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Blas, L. D., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psychological studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356-366.
- Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology, 56*, 45-74.
- Bergman, M. E., & Jean, V. A. (2015). Where Have All the "Workers" Gone? A Critical Analysis of the Unrepresentativeness of Our Samples Relative to the Labor Market in the Industrial–Organizational Psychology Literature. *Industrial and Organizational Psychology, 1*-30.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied psychology: An International Review, 46*, 5-34.
- Bing, M. N., Whanger, J. C., Davidson, 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.
- Boehm, V. R. (1977). Differential prediction: A methodological artifact? *Journal of Applied Psychology, 62*, 146-154.
- Bowling Green State University, (2009). Abridged Job Descriptive Index including Abridged Job in General Scale. BGSU.
- Browne, M. W. (1992). Circumplex models for correlation matrices. *Psychometrika, 57*, 469-497.
- Buss, D. M. (1987). Selection, evocation, and manipulation. *Journal of Personality and Social Psychology, 53*, 1214-1221.

- Butter, R., & Born, M. Ph. (2012). Enhancing criterion-related validity through bottom-up contextualization of personality inventories: The construction of an ecological conscientiousness scale for Ph.D. candidates. *Human Performance*, 25, 303-317.
- CBS (2012). Obtained from: <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71822NED&D1=0-1&D2=0&D3=0&D4=0-4&D5=a&D6=0&D7=0,4-l&HDR=T,G2,G1,G5,G6&STB=G4,G3&VW=T>
- CBS (2015a). Obtained from: <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71055NED&D1=0&D2=a&D3=0&D4=1,5-7&D5=0&D6=15-19&HDR=T,G4,G1,G2&STB=G3,G5&P=T&VW=T>
- CBS (2015b). Obtained from: <http://www.cbs.nl/en-GB/menu/themas/dossiers/allochtonen/methoden/begrippen/default.htm?Languageswitch=on&ConceptID=1013>
- Celenk, O., & Van de Vijver, F. J. R. (2011). Assessment of acculturation: Issues and overview of measures. *Online Readings in Psychology and Culture*, 8.
- Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality*, 17, 237-250.
- Church, A. T. (2010). Current perspectives in the study of personality across cultures. *Perspectives on Psychological Science*, 5, 441-449.
- Church, A. T., Katigbak, M. S., Miramontes, L. G., Del Prado, A. M., & Cabrera, H. F. (2007). Culture and the behavioural manifestations of traits: An application of the act frequency approach. *European Journal of Personality*, 21, 389-417.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- 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.
- Cook, M. (2004). Minorities, fairness and the law. In Cook, M. (eds.) *Personnel selection: Adding value through people fourth edition*. West-Sussex, UK: John Wiley & Sons.
- Costa, P. T., & McCrea, R. R. (1989). *The NEOPI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences*, 12, 887-898.
- Costa, P. T., & McCrae, R. R. (1992). *NEO Personality Inventory - Revised (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Cousins, S. D. (1989). Culture and self-perception in Japan and the United States. *Journal of Personality and Social Psychology*, 56, 124-131.
- Darcy, M. U. A. (2005). Examination of the structure of Irish students' vocational interests and competence perceptions. *Journal of Vocational Behavior*, 67, 321-333.
- Darcy, M., & Tracey, T. J. (2003). Integrating abilities and interests in career choice: Maximal versus typical assessment. *Journal of Career Assessment*, 11, 219-237.
- Darcy, M. U. A., & Tracey, T. J. (2007). Circumplex structure of Holland's RIASEC interests across gender and time. *Journal of Counseling Psychology*, 54, 17-31.
- De Meijer, L. A., Born, M. P., Terlouw, G., & Van Der Molen, H. T. (2008). Criterion-related validity of Dutch police-selection measures and differences between ethnic groups. *International Journal of Selection and Assessment*, 16, 321-332.

- De Vries, A. De Vries, R. E., & Born, M. Ph. (2011). Broad versus narrow traits: Conscientiousness and Honesty-Humility as predictors of academic criteria. *European Journal of Personality, 25*, 336-348.
- De Vries, A., Born, M. Ph., & De Vries, R. E. (2012). Conscientiousness and Integrity among ethnic majority and minority students: Score differences and predictive validity. In De Vries, A. (Unpublished doctoral dissertation), *Specificity in Personality Measurement* (pp. 45-58).
- De Vries, R. E. (2010). Lots of target variance: An update of SRM using the HEXACO personality inventory. *European Journal of Personality, 24*, 169-188.
- 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., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO Persoonlijkheidsvragenlijst. [The six most important personality dimensions and the HEXACO Personality Inventory.] *Gedrag & Organisatie, 22*, 232-274.
- De Vries, R. E., Tybur, J. M., Pollet, T. V., & Van Vugt, M. (2015). Evolution, Situational Affordances, and the HEXACO Model of Personality. *Manuscript submitted for publication*.
- Dryler, H. (1998). Parental role models, gender and educational choice. *The British Journal of Educational Choice, 49*, 375-398.
- Einarsdóttir, S., Rounds, J., & Su, R. (2010). Holland in Iceland Revisited: An Emic Approach to Evaluating U.S. Vocational Interest Models. *Journal of Counseling Psychology, 57*, 361-367.
- Eurostat (2014). Obtained from: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration\\_and\\_migrant\\_population\\_statistics#Main\\_tables](http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics#Main_tables)
- Feldt, L.S., & Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 105-146). New York: Macmillan.
- Francis, G. (2012). Evidence that publication bias contaminated studies relating social class and unethical behavior. *Proceedings of the National Academy of Sciences, 109*, E1587.
- Fuller, B. E., Holland, J. L., & Johnston, J. A. (1999). The relation of profile elevation in the Self-Directed Search to personality variables. *Journal of Career Assessment, 7*, 111-123.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149-158.
- Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656-672.
- Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality, 76*, 1267-1316.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review, 18*, 694-734.
- Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.
- Gottfredson, L. S. (1980). Construct validity of Holland's occupational typology in terms of prestige, census, department of labor, and other classification systems. *Journal of Applied Psychology, 65*, 697-714.
- Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology, 28*, 545-579.

- Grassi, M., Luccio, R., & Di Blass, L. (2010). CircE: An R implementation of Browne's circular stochastic process model. *Behavioral Research Methods*, *42*, 55-73.
- Gupta, S., Tracey, T. J., & Gore Jr, P. A. (2008). Structural examination of RIASEC scales in high school students: Variation across ethnicity and method. *Journal of Vocational Behavior*, *72*, 1-13.
- Hannover, B., Morf, C. C., Neuhaus, J., Rau, M., Wolfgramm, C., & Zander-Musić, L. (2013). How immigrant adolescents' self-views in school and family context relate to academic success in Germany. *Journal of Applied Social Psychology*, *43*, 175-189.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77-89.
- 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.
- Heller, D., Watson, D., Komar, J., Min, J. A., & Perunovic, W. Q. E. (2007). Contextualized personality: Traditional and new assessment procedures. *Journal of Personality*, *75*, 1229-1254.
- Hirschi, A., & Läge, D. (2007). Holland's secondary constructs of vocational interests and career choice readiness of secondary students: Measures for related but different constructs. *Journal of Individual Differences*, *28*, 205-218.
- Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, *6*, 35-45.
- Holland, J. L. (1985). *Vocational Preference Inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.
- Holtrop, D. J., Born, M. Ph., De Vries, R. E., & De Vries, A. (2014) A matter of context: A comparison of two types of contextualized personality measures. *Personality and Individual Differences*, *68*, 234-240.
- Holtrop, D., Born, M. P., & de Vries, R. E. (2015). Relating the spherical representation of vocational interests to the HEXACO personality model. *Journal of Vocational Behavior*, *89*, 10-20.
- Holtz, B.C., Ployhart, R.E., & Dominguez, A. (2005). Testing the rules of justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. *International Journal of Selection and Assessment*, *13*, 75-86.
- Howard, K. A., Carlstrom, A. H., Katz, A. D., Chew, A. Y., Ray, G. C., Laine, L., & Caulum, D. (2011). Career aspirations of youth: Untangling race/ethnicity, SES, and gender. *Journal of Vocational Behavior*, *79*, 98-109.
- Hubert, L., & Arabie, P. (1987). Evaluating order hypotheses within proximity matrices. *Psychological Bulletin*, *102*, 172-178.
- Hunter, J. E., & Schmidt, F. L. (1978). Differential and single group validity of employment tests by race: A critical analysis of three recent studies. *Journal of Applied Psychology*, *63*, 1-11.
- Hunthausen, J. M., Truxillo, D. M., Bauer, T. B., & Hammer, B. L. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, *88*, 545-551.
- Johnson, J.W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1-19.
- Johnson, M. K., Rowatt, W. C., & Petrini, L. (2011). A new trait on the market: Honesty-Humility as a unique predictor of job performance ratings. *Personality and Individual Differences*, *50*, 857-862.

- Katzell, R. A., & Dyer, F. J. (1977). Differential validity revived. *Journal of Applied Psychology, 62*, 137-145.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York, NY, US: The Guilford Press.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8*, 265-280.
- Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63-82.
- Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of big six interests and big five personality factors. *Journal of Vocational Behavior, 61*, 217-239.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*, 329-358.
- Lee, K., & Ashton, M. C. (2008). The HEXACO personality factors in the indigenous personality lexicons of English and 11 other languages. *Journal of Personality, 76*, 1001-1053.
- Lee, K., Ashton, M.C., & De Vries, R.E. (2005). Explaining workplace delinquency and integrity with the HEXACO and Five-Factor Models of personality structure. *Human Performance, 18*, 179-197.
- Lee, K., Ashton, M. C., Pozzebon, J. A., Visser, B. A., Bourdage, J. S., & Ogunfowora, B. (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of Personality and Social Psychology, 96*, 460-472.
- Lee, K., Ashton, M.C., & Shin, K. H. (2005). Personality correlates of workplace anti-social behavior. *Applied Psychology: An International Review, 54*, 81-98.
- Lee, I. H., & Rojewski, J. W. (2009). Development of occupational aspiration prestige: A piecewise latent growth model of selected influences. *Journal of Vocational Behavior, 75*, 82-90.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*, 79-122.
- Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scores and validity. *Journal of Applied Psychology, 93*, 268-279.
- Liu, M., & Huang, J. L. (2015). Cross-cultural adjustment to the United States: the role of contextualized extraversion change. *Frontiers in Psychology, 6*, no. 1650.
- Long, L., Adams, R. S., & Tracey, T. J. G. (2005). Generalizability of interest structure to China: Application of the personal globe inventory. *Journal of Vocational Behavior, 66*, 66-80.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research how often do they really occur? *Perspectives on Psychological Science, 7*, 537-542.
- Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five, or one in addition? *Personnel Psychology, 60*, 1-34.
- Markt effect BV (2011). NKSO Nationaal studiekeuze onderzoek [National study choice investigation]. Obtained from <http://www.studiekeuzeinformatie.nl/wp-content/uploads/2013/10/NSKO-2011.pdf>
- McCann, C., Lipnevich, A. A., Poropat, A. E., Wiemers, M. J., & Roberts, R. D. (2015). Self- and parent-rated facets of conscientiousness predict academic outcomes: Parent-reports are more

- predictive, particularly for approach-oriented facets. *Learning and Individual Differences*, *42*, 19-26.
- McCrae, R. R., & Costa, P. T. (1994). The stability of personality: Observations and evaluations. *Current Directions in Psychological Science*, *3*, 173-175.
- McKay, D. A., & Tokar, D. M. (2012). The HEXACO and five-factor models of personality in relation to RIASEC vocational interests. *Journal of Vocational Behavior*, *81*, 138-149.
- Meeuwisse, M., Born, M. P., & Severiens, S. E. (2014). The family-study interface and academic outcomes: Differences and similarities between ethnic minority and ethnic majority students. *Cultural Diversity and Ethnic Minority Psychology*, *20*, 401-412.
- 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.
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, *79*, 272-280.
- Murtha, T. C., Kanfer, R., & Ackerman, P. L. (1996). Toward an interactionist taxonomy of personality and situations: An integrative situational-dispositional representation of personality traits. *Journal of Personality and Social Psychology*, *71*, 193-207.
- Nauta, M. M. (2012). Are RIASEC interests traits? Evidence based on self-other agreement. *Journal of Career Assessment*, *20*, 426-439.
- Nelling, A., Kandler, C., & Riemann, R. (2015). Substance and artifact in interest self-reports: A multiple rater twin study. *European Journal of Psychological Assessment*, *31*, 166-173.
- NOA. (2009). *Handleiding Multiculturele Persoonlijheids Test - Big Six* [Manual Multicultural Personality Test - Big Six]. Amsterdam: NOA BV.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*, 116-130.
- Oh, I. S., Le, H., Whitman, D. S., Kim, K., Yoo, T. Y., Hwang, J. O., & Kim, C. S. (2014). The incremental validity of Honesty-Humility over cognitive ability and the Big Five personality traits. *Human Performance*, *27*, 206-224.
- Pace, V. L., & Brannick, M. T. (2010). Improving prediction of work performance through Frame-of Reference consistency: Empirical evidence using openness to experience. *International Journal of Selection and Assessment*, *18*, 230-235.
- Paunonen, S. V., & Ashton, M. C. (2013). On the prediction of academic performance with personality traits: A replication study. *Journal of Research in Personality*, *47*, 778-781.
- Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, *109*, 4086-4091.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*, 153-172.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322-338.
- Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., & Goldberg, L. R. (2010). Psychometric characteristics of a public-domain self-report measure of vocational interests: The Oregon vocational Interest scales. *Journal of Personality Assessment*, *92*, 168-174.



- Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior*, 21, 259-287.
- Prediger, D. J. (1998). Is interest profile level relevant to career counseling? *Journal of Counseling Psychology*, 45, 204-211.
- Pyburn, K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61, 143-151.
- Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, 107, 677-718.
- Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2010). *Increasing the validity of personality questionnaires*. Paper presented at the 25th annual conference of The Society for Industrial and Organizational Psychology, Atlanta, GA, April.
- Reynolds, D. (2007). Restraining Golem and harnessing Pygmalion in the classroom: A laboratory study of managerial expectations and task design. *Academy of Management Learning & Education*, 6, 475-483.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, 138, 353-387.
- Roberts, B. W., Harms, P. D., Caspi, A., & Moffitt, T. E. (2007). Predicting the counterproductive employee in a child-to-adult prospective study. *Journal of Applied Psychology*, 92, 1427-1436.
- Robie, C., & Risavy, S. D. (2016). A comparison of frame-of-reference and frequency-based personality measurement. *Personality and Individual Differences*, 92, 16-21.
- Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of item context specificity on the measurement equivalence of a personality inventory. *Organizational Research Methods*, 3, 348-365.
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The big five personality factors and personal values. *Personality and Social Psychology Bulletin*, 28, 789-801.
- Roe, A., & Klos, D. (1969). Occupation classification. *Counseling Psychologist*, 1, 84-89.
- Roe, A. (1956). *The psychology of occupations*. New York: Wiley.
- Roth, P. L., & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology*, 85, 399-406.
- Rounds, J., & Tracey, T. J. G. (1993). Prediger's dimensional representation of Holland's RIASEC circumplex. *Journal of Applied Psychology*, 78, 875-890.
- Rounds, J., & Tracey, T. J. G. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology*, 43, 310-329.
- Ryder, A. G., Alden, L. E., & Paulhus, D. L. (2000). Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology*, 79, 49-65.
- Sacket, S.A. & Hansen, J.I.C. (1995). Vocational outcomes of college freshmen with flat profiles on the Strong Interest Inventory. *Measurement and Evaluation in Counseling and Development*, 28, 9-24.

- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56*, 302.
- Schmidt, F. L., Berner, J. G., & Hunter, J. E. (1973). Racial differences in validity of employment tests: Reality or illusion? *Journal of Applied Psychology, 58*, 5-9.
- 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.
- 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., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T. Y. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88*, 979-988.
- Schneider, B. (2007). Evolution of the study and practice of personality at work. *Human Resource Management, 46*, 583-610.
- Severiens, S., & Wolff, R. (2008). A comparison of ethnic minority and majority students: Social and academic integration, and quality of learning. *Studies in Higher Education, 33*, 253-266.
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and non-contextualized personality measures. *Personnel Psychology, 65*, 445-494.
- Sherman, R. A., & Serfass, D. G. (2015). The comprehensive approach to analyzing multivariate constructs. *Journal of Research in Personality, 54*, 40-50.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46*, 49-76.
- Sodano, S. M. (2011). Integrating work and basic values into the Spherical Model of Interests. *Journal of Vocational Behavior, 78*, 1-10.
- Sodano, S. M., & Tracey, T. J. G. (2008). Prestige in interest activity assessment. *Journal of Vocational Behavior, 73*, 310-317.
- Starrels, M. E. (1994). Gender differences in parent-child relations. *Journal of Family Issues, 15*, 148-165.
- Staufer, J. M., & Buckley, M. R. (2005). The existence and nature of racial bias in supervisory ratings. *Journal of Applied Psychology, 90*, 586-591.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.
- Stevens, G., & Cho, J. H. (1985). Socioeconomic indexes and the new 1980 census occupational classification scheme. *Social Science Research, 14*, 142-168.
- Su, R., Rounds, J., Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*, 859-884.
- Sverko, I. (2008). Spherical model of interests in Croatia. *Journal of Vocational Behavior, 72*, 14-24.
- Szarota, P., Ashton, M. C., & Lee, K. (2007). Taxonomy and structure of the Polish personality lexicon. *European Journal of Personality, 21*, 823-852.

- Te Nijenhuis, J., & Van Der Flier, H. (2000). Differential prediction of immigrant versus majority group training performance using cognitive ability and personality measures. *International Journal of Selection and Assessment*, 8, 54-60.
- Ten Berge, M. A., & De Raad, B. (1999). Taxonomies of situations from a trait psychological perspective. A review. *European Journal of Personality*, 13, 337-360.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88, 500-517.
- The Economist (2013). Obtained from: <http://www.economist.com/news/business/21575817-can-leaders-be-identified-psychometrics-emotional-breakdown>
- Tracey, T. J. G. (1997). RANDALL: A Microsoft FORTRAN program for randomization test of hypothesized order relations. *Educational and Psychological Measurement*, 57, 164-168.
- Tracey, T. J. G. (2002). Personal globe inventory: Measurement of the spherical model of interests and competence beliefs. *Journal of Vocational Behavior* 60, 113-172.
- Tracey, T. J. G. (2010). Development of an abbreviated Personal Globe Inventory using item response theory: The PGI-Short. *Journal of Vocational Behavior*, 76, 4-15.
- Tracey, T. J. G. (2012). Problems with single interest scales: Implications of the general factor. *Journal of Vocational Behavior*, 81, 378-384.
- Tracey, T. J. G., Allen, J., & Robbins, S. B. (2012). Moderation of the relation between person-environment congruence and academic success: Environmental constraint, personal flexibility and method. *Journal of Vocational Behavior*, 80, 38-49.
- Tracey, T. J. G., & Gupta, S. (2008). Interest assessment in an international context. In Athanasou, J. A., & Van Esbroeck, R. (Eds), *International Handbook of Career Guidance* (pp. 525-537). Springer Science.
- Tracey, T. J. G., & Rounds, J. (1996). The spherical representation of vocational interests. *Journal of Vocational Behavior*, 48, 3-41.
- Van den Broek, A., Wartenbergh, F., Brink, M., Kurver, B., Hampsink, S., & Braam, C. (2013). *Monitor beleidsmaatregelen 2013. De relatie tussen beleidsmaatregelen, studiekeuze en studiegedrag* [The relation between government policies, study choice and study behavior]. Retrieved from the Dutch Government website: <https://www.rijksoverheid.nl/documenten/rapporten/2013/01/31/monitor-beleidsmaatregelen>
- Van der Linden, D., Te Nijenhuis, J., & Bakker, A. B. (2010). 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 Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: The validity of an interest-based selection test in relation to job knowledge, job performance and continuance intentions. *Journal of Applied Psychology*, 96, 13-33.
- Van Iddekinge, C. H., Taylor, M. A., & Eidson Jr., C. E. (2005). Broad versus narrow facets of integrity: Predictive validity and subgroup differences. *Human Performance*, 18, 151-177.
- Vigil-Colet, A., Morales-Vives, F., & Lorenzo-Seva, U. (2013). How social desirability and acquiescence affect the age-personality relationship. *Psicothema*, 25, 342-348.
- Wanous, J. P., & Reichers, A. E. (1996). Estimating the reliability of a single item measure. *Psychological Reports*, 78, 631-634.
- Wasti, S. A., Lee, K., Ashton, M. C., & Somer, O. (2008). The Turkish personality lexicon and the HEXACO model of personality. *Journal of Cross-Cultural Psychology*, 39, 665-684.

- Wasti, S. A., Tan, H. H., Brower, H. H., & Önder, Ç. (2007). Cross-cultural measurement of supervisor trustworthiness: An assessment of measurement invariance across three cultures. *The Leadership Quarterly, 18*, 477-489.
- Watson, D., & Clark, A. E. (1991). Self- versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology, 60*, 927-940.
- Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78*, 546-558.
- Wiechman, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment, 11*, 215-229.
- Wilkins, G. K., Ramkissoon, M., & Tracey, T. J. G. (2013). Structure of interests in a Caribbean sample: Application of the Personal Globe Inventory. *Journal of Vocational Behavior, 83*, 367-372.
- Wolff, R. (2013). Presteren op vreemde bodem: studiesucces van niet-westerse allochtone studenten [Performing on foreign soil: Study success of non-western minority students]. *Tijdschrift voor Hoger Onderwijs, 31*, 165-177.
- Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology, 53*, 1159-1177.
- Zettler, I., & Hilbig, B. E. (2010). Honesty-Humility and a person-situation interaction at work. *European Journal of Personality, 24*, 569-582.
- Zettler, I., Lang, J. W., Hülshager, U. R., & Hilbig, B. E. (2015). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process model to self-and observer reports. *Journal of Personality, in press*.

## Dankwoord

*I'm pretty sure there's a lot more to life than being really, really, ridiculously good looking. And I plan on finding out what that is.*

- Zoolander (2001)

Met deze of mogelijk een vergelijkbare gedachte begon ik aan dit project. Na vier jaar kan ik zeggen dat er inderdaad meer dingen in het leven zijn dan er heel (heel) erg leuk uitzien; alleen vind ik deze dingen vaak een stuk ingewikkelder. Misschien is de leukste conclusie na deze vier jaren dat ik er nog steeds leuk uitzie en een beetje heb ontdekt wat er nog meer te doen is in het leven.

Dat was echter genoeg lof voor mezelf. Een groot aantal personen heeft bijgedragen aan het mooie product dat u nu in handen heeft en zij verdienen ook lof. Om deze reden wil ik graag Marise, Reinout, Remko, en Anita bedanken voor jullie vertrouwen, oneindige steun en (boven alles) passie voor ons vak. Hun geweldige intellect heeft mij doen realiseren dat er oneindig veel te leren is. Zonder deze mensen (en mijn looks) had dit project nooit een succes kunnen zijn of zou ik nooit aan dit project zijn begonnen.

Marise, jij hebt om vrij onduidelijke redenen altijd in me geloofd sinds Janneke mij aan je voorstelde als mogelijke tutor op de EUR. Na een gezamenlijke scriptie en researchmaster was dat vertrouwen er nog steeds en heb je mij gesteund om samen aan dit proefschrift te beginnen. Je wist me elke keer weer op nieuwe dingen te wijzen die belangrijk zijn voor een goed onderzoek(er) en zo het genot van onderzoek op mij over te dragen. Daarnaast kwam je altijd voor me op zodra je het idee had dat dat nodig was. Ik heb het gevoel dat jij en Reinout mijn wetenschappelijke opvoeding niet beter hadden kunnen uitvoeren. Reinout, jij was elke dag beschikbaar voor de meest (on)zinnige vragen die je geduldig beantwoorde. Je rust en precisie gaven mij vaak het overzicht dat ik nodig had om verder te blijven gaan. Dankzij jouw hartelijkheid, thee en gedrevenheid heb ik me vanaf het begin thuis gevoeld op de VU, wat best lastig is als je er maar zo weinig bent! Remko, ik zal altijd dankbaar zijn voor de mogelijkheid om te gaan (proberen om te) promoveren. Ook jouw deur stond altijd open, zelfs op de momenten dat ik dacht dat je hem misschien beter dicht kon houden (wat is het soms druk...). Als ik vragen had over de (on)zin van een onderzoek of gewoon eens wilde sparren liep ik na een tijdje verbaasd je kamer uit: "Hoe doet hij dat toch elke keer?" Tenslotte, Anita, je hebt me vanaf mijn scriptie en jouw eerste onderzoek begeleid. Je enthousiasme over mij heeft mijn eigen (aanzienlijke) enthousiasme altijd overtroffen. Hierdoor heb je mij en de anderen geïnspireerd om samen aan dit project te beginnen. Ik ben er trots op om met

je samen te hebben gewerkt en gelukkig dat we er ook nog een vriendschap aan hebben overgehouden.

Stichting NOA heeft dit project financieel mogelijk gemaakt, hiervoor ben ik ook zeer dankbaar. Ons doel was om vragenlijsten toe te passen op multiculturele vraagstukken. Dit is (ondanks ons harde werk) helaas niet altijd gelukt. Ik streef er wel naar om deze vraagstukken te blijven onderzoeken en hoop de investering van Stichting NOA ook in de toekomst voort te zetten. Nico Bleichrodt en Corine Sonke hebben via de begeleidingscommissie waardevolle bijdragen geleverd aan de onderzoeken in dit proefschrift. Ook al staan hun namen niet bij het begin van elk hoofdstuk, hun gedachten zijn hier weldegelijk in opgenomen. Verder wil ik ook graag de leden van mijn promotiecommissie, prof. dr. Karen van Oudenhoven-van der Zee, prof. dr. Terence Tracey, prof. dr. Filip de Fruyt, prof. dr. Rob Meijer, dr. Arne Evers en dr. Dimitri van der Linden, bedanken voor hun tijd, aandacht en de eer die zij mij aandoen. Moreover, I would also like to thank prof. dr. Terence Tracey for his hospitality and help with several PGI questions.

De sectie Sociale en Organisatiewetenschappen van de VU wil ik graag bedanken voor een werkplek en een groot aantal geweldige (ex-)collega's die het een plezier maakten om op de VU te werken. Janneke en Killian kan ik aan iedereen aanbevelen als geweldige kamergenoten. Heerlijk om met jullie te kletsen over van alles en dankzij jullie heb ik gezien wat er allemaal leeft op een universiteit. Janneke, we hebben samen ook nog eens een leuk bonusonderzoek gedaan en zijn met Reinout en Fabiola naar de V.S. gegaan: een reis die ik nooit zal vergeten. Natuurlijk ben jij ook te vinden onder de wisselende groep squashers, Richard, Jan, Art, Nils, Niels, Hester, en Thomas (ook superleuk om samen met jou en Leander na een zware week wat Mortal Combat te spelen). De andere (ex-)collega's van de VU waar ik altijd bij binnen kon lopen als ik even moest ontspannen, zoals Mariko, Zoi, Nale, Anna, Rutger, Henk, Linda, Seval en Marco, jullie maakten de VU een heerlijke omgeving om te zijn.

Mij (oud-)collega's bij NOA hebben de afgelopen vier jaar allemaal met mij meegeleefd en zich (alweer) een ware familie getoond. Zo door de jaren heen zijn er mensen gekomen en gegaan, ik ben jullie allemaal dankbaar voor jullie teamwork, warmte en betrokkenheid: Dirk, Ruben (we soldaten door), Willem, Sophie, Niels, Olaf, Indra, Edwin, Leandro, Corine (nogmaals), Diana, Marleen, Nicole, Frans, Mira (fijn dat je er was als ik het even niet meer op een rijtje had). Ik ga jullie dadelijk verlaten, maar nooit vergeten!

Verder worden veel mensen en organisaties aan het begin van de empirische hoofdstukken bedankt voor hun hulp en steun. Graag wil ik hierbij nogmaals benadrukken dat zonder jullie hulp deze onderzoeken niet tot stand hadden kunnen komen.

Dank ook aan mijn vrienden die altijd geïnteresseerd zijn geweest in de onderzoeken (er regelmatig ook een beetje mee hielpen) en me op de goede momenten ook aan de andere dingen in het leven lieten denken. Thijs, Peter, Menno (paranimfen doe je ongetwijfeld in stijl), Lena (en GJ), Afraisa, Onno, Arjen, Arno, Jean, Pascal, Sebas, Martin, Marie, Jannec, Alexander, Karsten, Laura, Pokerboys (double or nothing).

Dan de familie: Rob, Miek, Ros, San (geweldige tekenaar en de geestelijk vader van de cover van dit proefschrift), Yo, Gonny, Keith, Teb (en opa), Peter en Thera, ik heb het gehaald! We hebben geweldige tijden gehad die mij en mijn onderzoek regelmatig welverdiende rust gaven. Jullie zijn allemaal een voorbeeld voor me. Ma, je hebt je door de onderzoeksverslagen heen geworsteld en stopte niet tot je begreep wat je zoon uitvoerde. Pa, je hebt me op allerlei manieren gesteund met het onderzoek: je hebt meegedacht en zelfs geholpen met de uitvoering! Jullie hebben samen een hele, hele leuke jongen opgevoed die ook nog geïnteresseerd is in de rest van de wereld. Ik ben jullie dankbaar en enorm trots dat jullie mijn ouders zijn.

Cin, lieverd, het oneindige geduld dat je voor mij hebt gehad kan niet alleen verklaard worden met mijn looks. Wij zijn partners voor het leven en samen gaan we verder het avontuur in (en daarna nog één en nog één en ...). Je zorgt ervoor dat we stil staan bij de momenten waarvan we (van elkaar) moeten genieten. Samen met jou kan ik de wereld aan. Deze is voor jou schat.







