Improving personality and interest measurement for purposes of selection and assessment: Summary and general discussion
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, Hülsheger, & 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.

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

---

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
Summary and general discussion

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 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.