A Trait-Based Perspective on the Assessment of Personality and Personality Pathology in Older Adults

Joke Van den Broeck

Brussel, 2012

Colofon

Cover photo: Klaas Geers

Vrije Universiteit Brussel
Faculty of Psychology and Educational Science
Department of Clinical and Lifespan Psychology

A Trait-Based Perspective on the Assessment of Personality and Personality Pathology in Older Adults

Joke Van den Broeck

A dissertation submitted in fulfillment of the requirements for the degree of Doctor in Psychological Sciences

Brussel, 2012

Promotor	Prof. Dr. G. Rossi, Vrije Universiteit Brussel
Copromotors	Prof. Dr. E. Dierckx, Vrije Universiteit Brussel
	Prof. Dr. B. De Clercq, Universiteit Gent
Chairman	Prof. Dr. L. De Donder, Vrije Universiteit Brussel
Secretary	Prof. Dr. S. P. J. Van Alphen, Vrije Universiteit Brussel, Leerstoel Klinische Ouderenpsychologie
Jury	Dr. E. Hoogenhout, GZ psycholoog i.o., Vitalis (NL)
	Prof. Dr. M. J. M. Kardol, Vrije Universiteit Brussel, Leerstoel Bevordering Active Aging
	Dr. A. Haekens, Ouderenpsychiater Broeders Alexianen, Tienen

Dankwoord

Heel veel mensen hebben – al dan niet bewust - bijgedragen aan de totstandkoming van dit doctoraat. Ik wil al deze mensen heel graag bedanken, zonder hen was het mij nooit gelukt.

Mijn eerste woord van dank gaat uit naar mijn promotors Gina Rossi, Eva Dierckx en Barbara De Clercq. Bedankt Gina voor het vertrouwen dat je me vier jaar geleden schonk, en om al die jaren enthousiast in mij te blijven geloven. Eva, jij ook bedankt voor alle hulp, ik heb heel graag met jou samengewerkt. Af en toe had ik een duwtje in de rug nodig (lees: een schop onder mijn kont), en jij stimuleerde me op die momenten om het onderste uit de kan te halen. Barbara, jij ontfermde je over mij op mijn eerste internationale congres in Brno, waardoor ik er als enige VUB'er al snel werd opgenomen in de kliek van Gent, mooie herinneringen heb ik daaraan. Bedankt voor de tijd en het werk dat je stak in het nalezen van mijn teksten, jouw feedback was altijd heel waardevol voor mij.

Dat het schrijven van een doctoraat niet altijd van een leien dakje loopt en met de nodige ups en downs gepaard gaat, heb ik aan den lijve ondervonden. Dat ik werd omringd met toffe collega's waar ik altijd bij terecht kon om mijn hart te luchten of ervaringen te delen, maakte gelukkig heel veel goed. Bedankt Jan, Inge, Mieke, Daphné, Koen, Jochen, Lize en Natacha… ik ga nog vaak aan jullie denken om 11u45 (of was het 12u10?). Leen, jij stond van alle collega's het dichtst bij mij. Jij weet dan ook als geen ander hoe stormachtig het er soms aan toe ging de voorbije vier jaar. Ik kan je niet genoeg bedanken voor alles wat je voor mij betekend en gedaan hebt.

Bedankt ook Joeri Hofmans, Bas Van Alphen, Toine Thissen en Harry Barendse. Joeri, jouw hulp kwam er op een cruciaal moment en bracht het één en ander in beweging. Het was een plezier om met jou te

kunnen samenwerken. Je enthousiasme en motivatie werkten aanstekelijk, ik heb veel van jou geleerd. Bas, ook de samenwerking met jou was heel leerrijk en inspirerend. Jouw doctoraatsscriptie vormde de basis voor mijn doctoraat en je bent dan ook al die jaren een groot voorbeeld voor mij geweest. Harry en Toine, onze wegen scheidden zich in de loop van mijn doctoraat, maar ik ben jullie vriendelijke, open en bereidwillige houding zeker niet vergeten. Bedankt voor de fijne contacten, en Harry, veel succes met jouw doctoraatsonderzoek.

Ik wil ook graag Julia Villanueva O' Driscoll bedanken, voor de vriendschap, sinds de kleuterklas tot nu. Jul, geen idee wat de toekomst zal brengen, maar onze vriendschap is voor altijd, ook al zit jij aan de andere kant van de wereld. Groot woord van dank ook aan mijn ouders en zussen, gewoon om er te zijn, altijd en overal. Mama, papa, Loesje, Imke, Lientje en Ineke, bedankt voor alles. Mijn laatste dankwoord richt ik aan de twee grootste liefdes in mijn leven. Klaas en Linde, jullie hielpen me relativeren en tegenslag verwerken, en gaven me telkens weer de moed om door te zetten. Ik hou ontzettend veel van jullie!! Tot slot ben ik ook opgelucht dat ik me, na de afronding van dit doctoraat, voor 100% kan wijden aan het volgende ambitieuze project dat zich de laatste vijf maanden in sneltempo ontwikkelde in mijn buik. Sorry lieve schat in wording, voor alle stressmomentjes die jij ongetwijfeld hebt meegevoeld, hopelijk was je stressbestendiger dan ik. Tot gauw!!

Table of contents

Chapter 1:	Introduction	11
Chapter 2:	Age-neutrality of the NEO-PI-R: potential differential item functioning in older versus younger adults *Joke Van den Broeck, Gina Rossi, Eva Dierckx, & Barbara De Clercq*	33
Chapter 3:	Validation of the FFM PD count technique for screening personality pathology in later middle-aged and older adults *Joke Van den Broeck, Gina Rossi, Barbara De Clercq, Eva Dierckx, & Leen Bastiaansen*	59
Chapter 4:	Age-neutrality of the trait facets proposed for personality disorders in DSM-5: A DIFAS analysis of the PID-5 *Joke Van den Broeck, Leen Bastiaansen, Gina Rossi, Eva Dierckx, & Barbara De Clercq*	81
Chapter 5:	Hierarchical structure of maladaptive personality traits in older adults: Joint factor analysis of the PID-5 and the DAPP-BQ *Joke Van den Broeck, Leen Bastiaansen, Gina Rossi, Eva Dierckx, Barbara De Clercq, & Joeri Hofmans*	101
Chapter 6:	Personality assessment in older adults: the value of personality questionnaires unraveled *Joke Van den Broeck, Bas Van Alphen, Gina Rossi, Eva Dierckx, & Daniel Segal*	115
Chapter 7:	General discussion	125

CHAPTER 1: INTRODUCTION
1.1. Introduction 11
1.2. Why study personality (pathology) in later life? 13
1.3. Normal personality: A Five-Factor Model perspective 14
1.4. Personality across the lifespan 15
1.5. Personality disorders: An operational definition 16
1.6. Towards DSM-5: A dimensional model of classification 17
1.7. Personality disorders across the lifespan 19
1.8. Heterotypic continuity 19
1.9. Personality disorders in later life: Measurement issues 20
1.10. Assessing older adults (mal)adaptive personality traits: Challenges and pitfalls 23
1.11. Aim of the current dissertation 24
1.12. Specific research objectives of this dissertation: An overview 24
1.13. References 26
CHAPTER 2: AGE-NEUTRALITY OF THE NEO-PI-R
2.1. Introduction 34
2.2. Method 37
2.3. Results 40
2.4. Discussion 47
2.5. References 52
CHAPTER 3: FFM PD COUNTS
3.1. Introduction 60
3.2. Method 63
3.3. Results 67
3.4. Discussion 71
3.5. References 76
3.6. Appendix 80
CHAPTER 4: AGE-NEUTRALITY OF THE PID-5
4.1. Introduction 82
4.2. Method 83
4.3. Results 86
4.4. Discussion 92
4.5. References 96
CHAPTER 5: PID-5 - DAPP-BQ JOINT HIERARCHICAL STRUCTURE
5.1. Introduction 102
5.2. Method 104
5.3. Results 107

5.4. Discussion 108
5.5. References 112
CHAPTER 6: COMMENTARY
References 121
CHAPTER 7: GENERAL DISCUSSION
7.1. Introduction 125
7.2. Summary of findings 127
7.3. Limitations 131
7.4. Directions for further research 133
7.5. Practical implications 134
7.6. Conclusion 137
7.7. References 137

Chapter 1
Introduction[1]

1.1. Introduction

The past decades, the number of people aged 65 or more has worldwide considerably increased. It is expected that this growth will persist and even increase in the future, mainly due to an improved health care system (Dierckx, 2012). In 2007 in Belgium, for example, the percentage of older people (> 65) amounted to 17%, whereas in 2030 this is estimated to be 22.6% (http://statbel.fgov.be). As a result, health care institutions will be increasingly confronted with the specific demands of older patients, including their need for mental health counseling. This implies several challenges and pitfalls, such as the need for valid and age-appropriate assessment tools for psychopathology. In this context, the current dissertation focuses on the assessment of personality and personality pathology in older adults[2].

---

[1] Part of this introduction is based on:

Van den Broeck, J., Rossi, G., & Dierckx, E. (2010). Diagnostiek van persoonlijkheid en persoonlijkheidspathologie bij ouderen. *Tijdschrift voor Gerontologie en Geriatrie, 41(2)*, 68-78.

Van den Broeck, J., Barendse, H. P. J., van Alphen, S. P. J., Thissen, T. & Rossi, G. (2012). Testdiagnostiek van persoonlijkheid en persoonlijkheidsstoornissen bij ouderen: een praktijkgerichte benadering. *Tijdschrift voor Klinische Psychologie, 42(1),* 24-32.

[2] Traditionally, the age of 65 is used as a demarcation of old age or the later life stage, but it may be clear that this is a very heterogeneous group with

The trajectory of personality disorders (PDs) is an understudied field of interest, especially compared to the amount of studies devoted to other forms of mental disorders (Oltmanns & Balsis, 2011). By extension, little attention has been paid to the psychological assessment of personality and personality disorders in older adults, both in research and clinical practice. It was generally assumed that people's personality mellows or softens with age (Kenan et al., 2000; Paris, 2003), and that older adults with personality difficulties would not benefit from psychotherapy. The past decade however, the interest in this topic and related to this the amount of research programs has substantially grown. The study of personality and personality disorders in later life will become even more important given the growing number of older adults in our Western society in general and in mental care institutions particularly. Despite the growing interest in the assessment of personality and personality disorder pathology in older adults however, research in this field is hampered by both conceptual and methodological issues that mutually affect each other. In short, knowledge about the conceptualization of personality and personality disorders in later life is relatively limited partly due to problematic diagnostic criteria, and this lack of information hampers researchers to thoroughly revise these criteria in order to resolve the conceptual problems (Oltmanns & Balsis, 2011).

---

significant variations in life experiences, physical ability, psychological features and social opportunities. The scientific literature therefore makes a distinction between the "young-old" (between ages 65 and 74), the "old-old" (between ages 75 and 84), and the "oldest-old" (aged 85 and older) (Segal, Coolidge, & Rosowsky, 2006). Throughout this dissertation however, we have chosen to use the general term "older adults" across these sub-groups for reasons of readability, although we do acknowledge the wide diversity and heterogeneity of this age group.

1.2. Why study personality (pathology) in later life?

The relevance of studying personality (pathology) in later life can be understood from several perspectives. First, adaptation to changing circumstances is one of the main functions of our personality, and people with maladaptive personality traits (or personality disorders) may be less able to adequately cope with age-related changes. The transition period from mid to late adulthood is often considered as a turbulent period in which people are confronted with life-changing experiences as retirement, illness, or loss. Most people will adapt successfully to these age-related changes, for example by establishing a more dependent relationship with relatives in case of physical deterioration. However, for people with maladaptive personality traits and inadequate coping styles, these normative changes may initiate or aggravate psychopathology. Think for example of a woman with histrionic personality disorder features who has relied her whole life on her physical attractiveness and sexual provocativeness as a means of gaining attention, but who may feel neglected and abandoned as she ages and loses some of her seductiveness (Molinari & Segal, 2011). Second, it is generally assumed that co-morbid personality disorders may influence the presentation of Axis I symptomatology, impeding the assessment process. For example, disruptive behavior in the nursing home may camouflage the fact that the person is suffering from a depression which, in turn, aggravates premorbid antisocial personality features (Molinari & Segal, 2011). However, although the comorbidity issue between Axis I syndromes and personality disorders have been broadly addressed for younger adults, it has received surprisingly little attention in the geriatric mental health literature. As such, relatively little is known regarding the relation between depression, anxiety, and other mental disorders and personality disorders in an older population (Agronin & Maletta, 2000; Rosowsky, Abrams, & Zweig, 1999; Segal, Coolidge, & Rosowsky, 2006). Third, just as for younger adults, treatment of patients who suffer from a (co-morbid) personality disorder generally takes more time because of the more complex and often chronic psychological symptoms, and the risk of relapse is higher compared to patients who do not suffer from a

personality disorder (Van Alphen, Engelen, Kuin, & Derksen, 2006). Especially in somatic and psychological interventions for older adults, the presence of a personality disorder tends to complicate treatment due to resistance to care, noncompliance or medication abuse, or excessive care demands, depending on the specific personality disorder (Van Alphen, Derksen, Sadavoy, & Rosowsky, 2012).

1.3. Normal personality: A Five-Factor Model perspective

One of the most common approaches to characterize individual differences within psychology is the use of traits (Tackett, Balsis, Oltmanns, & Krueger, 2009). In this respect, personality is generally operationalized as a complex construct that is broadly composed of personality traits and characteristic adaptations (e.g., coping style) owned by a person, and uniquely influencing his or her thoughts, feelings and behaviors. Traits are thought to be stable across time and situations, and to predict future behavior. Arguably the predominant model of normal-range personality traits is the five-factor model (FFM) (Goldberg, 1993). The FFM was derived originally through empirical studies of trait terms within the English language, and subsequent lexical studies have been conducted on many additional languages, all confirming the existence of five broad domains of general personality functioning (Ashton & Lee, 2001). These domains have been identified as neuroticism (or emotional instability), extraversion (or surgency), agreeableness, conscientiousness (or constraint), and openness (or intellect, imagination, or unconventionality) (Widiger & Trull, 2007). The five broad domains have been further differentiated into more specific facets by Costa & McCrae (1992) on the basis of their development of and research with the NEO Personality Inventory-Revised (NEO-PI-R), by far the most commonly used and heavily researched measure of the FFM (Widiger & Mullins-Sweatt, 2009, p. 199).

1.4. Personality across the lifespan

Although personality traits are commonly defined as relatively enduring patterns of thoughts, feelings and behaviors that distinguish individuals from one another, the personality stability issue is the subject of considerable debate in personality research. Based on both longitudinal and cross-sectional studies, it had been previously argued that there is little or no mean level change in personality after the age of 30 (Costa & McCrae, 1992, 1997). In this perspective, personality traits were seen as biologically determined, and not affected by environmental influences. More recent studies however, provided evidence that changes in mean levels of personality traits may occur beyond the age of 30, suggesting that there is no specific age at which personality traits stop changing (e.g., Srivastava et al., 2003; Terracciano et al., 2005; Roberts et al., 2006; Costa & McCrae, 2006). A meta-analyses by Roberts and colleagues (2006) examined mean level changes in traits across discrete age categories over the lifespan ranging from 10 to 101 years. Based on their own cross-sectional and longitudinal studies, Costa and McCrae (2006) reached broadly the same conclusions as Roberts et al. (2006), namely that: a) neuroticism (the inverse of emotional stability) and extraversion decline, whereas agreeableness and conscientiousness increase with age; and that b) openness first increases and then decreases. Additionally, they concluded that c) changes are more pronounced in early adulthood than either before or after; and that d) similar patterns are found for men and woman (McCrae & Costa, 2003). However, despite the consistent findings of slight differences in personality traits across age, the overall trend in the FFM tradition still suggests that personality traits are stable over time (Balsis, Gleason, Woods, & Oltmanns, 2007).

On the contrary, a dynamic context depended-view rejects the notion that traits are sufficient descriptors of personality (Balsis et al., 2007). This view suggests that personality may change as one's situation changes, as a result of complex interactions between biological and socio-cultural influences (e.g., Mischel, 2004; Mischel & Shoda, 1998). Within this tradition, personality is thought to manifest itself differently

across younger and older adults in as much as those age groups represents different meaningful contexts in terms of occupational, social, economic, and physiological aspects (Balsis et al., 2007). Several studies confirm the idea that personality can change, especially when people are confronted with important life events and need to cope with their changing lives (Haan, Millsap, & Hartka, 1986; Maiden, Peterson, Caya, & Hayslip, 2003). For instance, Haan and colleagues (1986) found that personality was unstable during the transitional period between middle adulthood to late adulthood. People in this stage of life often face serious changing life circumstances as retirement, illness, and widowhood, hence this period is commonly seen as a challenging and turbulent period in terms of behavioral and affective expressions (Zarit, Johansson, & Malmberg, 1995). These findings support the notion that personality does change, specifically when confronted with life events that require adaptation. In this view, people maintain a stable personality under stable life circumstances (e.g., a stable marriage, job satisfaction,…), but their personality changes when they are confronted with changing circumstances and try to adapt. Thus, from this perspective, changes in adaptive personality traits may reflect individual's attempts to cope with alterations in their life circumstances (Maiden et al., 2003).

All in all, all of the above suggests that normative maturational changes, as well as environmental contingencies, may affect the trajectory of personality (and personality disorders) over the life-span (Zweig, 2008).

1.5. Personality disorders: An operational definition

According to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.: DSM-IV; American Psychiatric Association, 1994), the most widely used classification taxonomy for personality disorders, personality traits constitute a personality disorder when they are rigid and maladaptive and cause functional impairment or subjective distress. According to the general criteria for personality disorder, "a personality disorder is an

enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture", as evidenced in at least two of the following domains of functioning: cognition, affectivity, interpersonal functioning, and impulse control (American Psychiatric Association, 1994, p. 685). In addition, the diagnosis of a personality disorder requires that the enduring pattern is pervasive and inflexible across a range of situations, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment in important areas of functioning (e.g., the work environment). Also the pattern may not be better accounted for as a manifestation of another mental disorder, and may not be due to the direct physiological effects of substance use or medical illness.

Apart from these general criteria, Axis II of the DSM-IV lists ten specific diagnoses of personality disorder, and presents them as clearly distinct categories. The ten personality disorders are grouped into three clusters: Cluster A, characterized by odd, eccentric traits (i.e., the paranoid, schizoid, and schizotypal PDs); Cluster B, characterized by dramatic, emotional traits (i.e., the anti-social, borderline, narcissistic, and histrionic PDs); and Cluster C, characterized by anxious, avoidant traits (i.e., the avoidant, dependent, and obsessive-compulsive PDs). For individuals who meet the general criteria but not the required threshold for any specific personality disorder, an additional category labeled "PD not otherwise specified" is also available.

1.6. Towards DSM-5: A dimensional model of classification

The categorical conceptualization of personality disorders in the current DSM-IV has been extensively criticized (e.g., Clark, 2007; Widiger & Trull, 2007), and has led to the proposal of a dimensional classification of personality disorders in the upcoming new edition of the DSM (i.e., the DSM-5), that is now scheduled for May 2013. In the course of this PhD-project, the transition from the DSM-IV (American Psychiatric

Association, 1994) towards the DSM-5 increasingly began to take shape. The DSM-5 Personality and Personality Disorders Workgroup proposes a hybrid dimensional-categorical model for personality and personality disorder assessment and diagnosis in which six specific personality disorder types are defined by two fundamental criteria, being impairments in personality functioning and the presence of pathological personality traits. Regarding the latter, a multidimensional maladaptive personality trait system and an associated assessment instrument has been developed by Krueger and colleagues (2012) in which 25 primary traits are organized by five higher-order dimensions (Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism). While constructing this trait model and its associated assessment instrument, the Personality Inventory for DSM-5 (PID-5; Krueger et al., 2012), the DSM-5 Personality and Personality Disorders Workgroup relied on existing models of maladaptive personality traits, such as Harkness's Personality Psychopathology Five model (PSY-5; Harkness, McNulty, & Ben-Porath, 1995), and the Dimensional Assessment of Personality Pathology model (DAPP; Livesley, Jackson, & Schroeder., 1992; Krueger et al., 2012). In particular, they sought to identify traits that encompass the four major bipolar domains of maladaptive personality variation identified by Widiger and Simonsen (2005) as presenting the common ground among 18 existing dimensional models of personality disorder: extraversion versus introversion; antagonism versus compliance; constraint versus impulsivity; and negative affect versus emotional stability. In addition to these four broad domains, a fifth domain of psychoticism was included, to provide coverage of features associated with schizotypal personality disorder (i.e., cognitive or perceptual distortions and eccentricities of behavior) (DSM-IV-TR, APA, 1994; Harkness et al., 1995; Chmielewski & Watson, 2008; see also Krueger et al., 2012). Important to note is that the implementation of a new personality disorder description in the DSM is currently the subject of much controversy and ongoing debate - for the latest information concerning the actual proposal we therefore refer to the DSM-5 website (www.dsm5.org; APA, 2012).

1.7. Personality disorders across the lifespan

Longitudinal studies of personality disorders that extend into old age are scarce. Thus, knowledge about the course of personality disorders over the lifespan is relatively limited (Zweig, 2008), especially when compared to the later life trait literature. Yet some initial developmental inferences can be drawn from cross-sectional comparisons between younger and older participants. Several studies indicate that personality disorders from the A and C cluster remain relatively stable over time, whereas cluster B disorders (especially borderline and antisocial personality disorders) are less prevalent among older people than younger people (Abrams & Horowitz, 1999). In terms of prevalence rates, the overall trend suggests that younger adults are diagnosed with personality disorders more frequently than older adults (e.g., Ames & Molinari, 1994; Casey & Schrodt, 1989; Fogel & Westlake, 1990; Kenan et al., 2000). The reasons for this trend are not entirely clear. Some researchers have suggested that personality disorders mellow or soften with age (Kenan et al., 2000; Paris, 2003), whereas others found that significant interpersonal and functional problems remained, while specific symptoms to meet diagnostic threshold disappeared as people grow older (Moffit, Caspi, Harrington, & Milne, 2002; see also Balsis et al., 2007). However, the apparent lower amount of personality pathology in older adults may as well reflect a measurement artifact. It is possible that personality problems present themselves differently in later life and hence remain undetected when relying on diagnostic criteria that are not attuned to the later life context (Mroczek, Hurt, & Berman, 1999).

1.8. Heterotypic continuity

One way to explain the different presentation of personality disorders through age is via the concept of heterotypic continuity. It refers to the idea that one's basic personality characteristics (or: traits) remain stable with age, while the presentation of these characteristics may change (e.g., Kagan, 1969; Caspi & Bem, 1990; Mrockzek et al., 1999). In this view one's core personality is behaviorally expressed in developmental

congruent ways throughout the lifespan (Hyer, Molinari, Mills, & Yeager, 2008). In other words: the person still possess the underlying disorder or qualities of the disorder, but the manifestation or presentation of the disorder changes with time, as a function of age and/ or changing contexts (Mroczek et al., 1999).

There are at least three ways in which personality disorder features may show heterotypic continuity with age, and combinations of types are possible (for an extensive overview we refer to Balsis et al., 2007, p. 172). First, the presentation of personality disorder features may show natural developmental change. For example, a little girl might behave aggressively by pulling other children by the hair, whereas this aggression may later in life be expressed by verbally insulting her caregivers in the nursing home. Second, the context within which the personality disorder features exist may change, both at the societal and the personal level. For example, physical deterioration in late life may cause paranoid features to come to the surface, when a fearful man no longer possesses the strength and agility he used to have that allowed him to overcome his fear of being mugged and venture outside the safety of him home (Balsis et al., 2007). Third, the opportunity for the presentation of the features may change over time (Mrockzek et al., 1999). For instance, a younger woman with borderline personality disorder features may have many opportunities to exhibit irritability as she frequently encounters conflict in her job for example. When she is retired, her irritability may become less apparent as she encounters less conflict in her daily life.

1.9. Personality disorders in later life: Measurement issues

In recent years, there have been a remarkable increase of knowledge about personality disorders in older adults, yet many questions remain unanswered, largely due to conceptual and methodological quandaries in this controversial study area (Molinari & Segal, 2011). It seems like a

vicious cycle: the conceptualization of maladaptive features of DSM-based personality disorders in older adults is hampered by limitations in its assessment, while improving the current measurement system is limited by the lack of knowledge about the conceptualization of personality pathology in later life. An additional difficulty is that there is no "gold standard" in personality assessment, and certainly not for the assessment of personality disorders in older age groups (Van Alphen, Engelen, Kuin, Hoijtink, & Derksen, 2006; Balsis, Segal, & Donahue, 2009).

With regard to the assessment of personality disorders in older adults, the most fundamental psychometric and conceptual issues concern the applicability of the current DSM-IV nosology in older age groups (Tackett et al., 2009). Several bottlenecks can indeed be listed in relation to the operationalization of both the general and specific criteria for DSM-IV personality disorders in older age groups (Van Alphen, Engelen, Kuin, & Derksen, 2004). With regard to the general criteria, one major issue pertains to the temporal stability. Following the above-mentioned definition, a personality disorder cannot arise in later life, since it is defined as an enduring pattern that is stable over time, and its onset has to be traced back at least to early adulthood. However, the DSM-IV does recognize the possibility that some personality disorders tend to remit with age, or remain undetected until relatively late in life (American Psychiatric Association, 1994). Clearly investigating the condition of temporal stability is a tough task for clinicians working with older adults, since reliable information regarding the patients background is often lacking, or at least questionable (Abrams & Bromberg, 2007). Questions can indeed be asked whether and to what extend an elderly patient or an informant is able to report reliably on a retrospective history covering several decades (Agronin & Maletta, 2000). Also, impairments in the occupational context no longer apply to a retired population. Likewise, impairments in social functioning might rather be due to physical deterioration or experiences of loss in older adults than pointing to personality dysfunction.

Similar reservations apply to the specific DSM-IV criteria for personality disorders. Most of the criteria focus on the living conditions of younger adults, affecting its content validity for use in older adults (Balsis et al., 2009). For example, the criterion "Almost always chooses solitary activities" is intuitively related to schizoid personality disorder pathology in younger adults. In later life however, this item will likely be more readily endorsed, independently from the underlying level of schizoid personality disorder pathology. Older adults may choose solitary activities because of physical limitations or immobility, or a diminishing social network, without having a schizoid personality disorder (Balsis et al., 2009). As another example, the criterion "Irritability and aggressiveness, as indicated by repeated physical fights or assaults" may lead to an underdiagnosis of the antisocial personality disorder in an elderly population, because it does not adequately capture the manifestation of an antisocial personality disorder in later life (Van Alphen, Nijhuis, & Oei, 2007). Aggression in later life may manifest itself in more verbal or passive-aggressive acts, rather than through physical fights and assaulting behavior, even though the latent trait of aggression is equally present (cfr. heterotypic continuity).

Personality pathology might therefore remain undetected by diagnostic criteria that are not designed for older people (Balsis et al., 2007). Moreover, the use of such criteria (or items based on such criteria) in older age groups not only hampers a valid assessment of personality pathology in later life, but also calls into question commonly held beliefs and theories on the conceptualization of personality pathology in later life (Tackett et al., 2009). Given the limitations inherent to the DSM-IV's categorical conceptualization of personality disorders in older adults, the transition towards a new edition of the DSM offers a great opportunity to ameliorate the existing classification taxonomy, especially with regard to a better understanding of the course of personality disorder pathology across the lifespan.

1.10. Assessing older adults (mal)adaptive personality traits: Challenges and pitfalls

Psychologists face several challenges to accurately assessing older adults who may present with maladaptive personality functioning. Arguably the most important challenges derive from difficulties applying the current DSM-IV personality disorder nosology to the evaluation of personality disorders in older adults (Zweig, 2008), as described above. Yet there are several other difficulties in diagnosing personality disorders in late-life that are worth mentioning. Cognitive impairment, for example, or memory and cognitive changes associated with normal aging may hamper a valid personality assessment (Morse & Lynch, 2000). Also technical, abstract or modern language often used in current personality inventories may hamper a valid assessment of those older adults with less formal education (Van Alphen et al., 2006). There may also be a cohort difference in language use contributing the validity issue. Older adults, for example, may be less inclined than younger adults to describe their lives in terms of "problems" or "stress" (Aldwin & Levenson, 1994), or they may tend to under-report personality traits considered socially undesirable (Maier et al., 1991). The majority of the current personality assessment measures have been developed for and validated in mixed-age younger adult samples, and the lay-out, item content, and norms are often not adjusted to the specific context of later life (Van Alphen et al., 2004; Van Alphen, 2006; Zweig, 2008). The work of geriatric health care providers is therefore often hampered by a lack of suitable assessment tools for use in older populations. Furthermore, a clinician's beliefs, expectations, and knowledge regarding personality and aging may certainly influence the assessment process. Given the aging population, more and more psychologists will be seeing older adults in their practices, even though they didn't necessarily have a formal clinical training in geropsychology (Zweig, 2008). As such, persisting stereotypes of the elderly as rigid, dependent, withdrawn, or untreatable, may erroneously incline clinicians to view pathological behavior as being part

of the normal aging process, possibly leading to under-diagnosis (Morse & Lynch, 2000; Zweig, 2008).

In sum, accurate assessment of personality pathology in older adults may be impeded by psychologists' unfamiliarity with this population, diagnostic criteria of uncertain validity for older persons, and significant limitations of current assessment tools. (Zweig, 2008, p. 300)

1.11. Aim of the current dissertation

The aim of this dissertation is twofold. First, we want to investigate the NEO-PI-R's age-neutrality, and examine the psychometric characteristics of a FFM-based system for assessing personality pathology in older adults, building on research conceptualizing personality disorders as maladaptive, extreme variants of general personality traits. Second, we zoom in on the assessment of personality pathology in future editions of the DSM, and aim to investigate the relevance of the proposed DSM-5 trait system for use with older adults.

1.12. Specific research objectives of this dissertation: An overview

An age-neutral measurement system is one of the basic conditions to study the course of personality across the lifespan, both longitudinally and cross-sectionally. To our knowledge, only two personality measures were created with the goal of age neutrality: the *Revised NEO Personality Inventory* (NEO-PI-R; Costa & McCrae, 1992), one of the most widely used personality measures, and the *Personality Assessment Inventory* (PAI; Morey, 1991). During the development of the latter, item response theory was applied to identify and eliminate those items that contained measurement bias across two broad age groups (Oltmanns & Balsis, 2010). In the construction of the NEO-PI-R the later life context was theoretically considered during the item generation and selection phase. However, its age-neutrality has not been empirically investigated. In a

first study **(Chapter 2)** we address this issue and investigate the age-neutrality of the NEO-PI-R items, exploring possible age-related measurement invariance across a younger and an older sample by conducting Differential Item Functioning (DIF) analyses.

In a second study **(Chapter 3)** we aim to evaluate the NEO-PI-R as a possible screening instrument to assess DSM-IV personality disorders from a Five-Factor Model perspective. This was done using the so-called "FFM PD count" technique, which was developed by Miller and colleagues (2005), and previously validated in younger (and middle-aged) adult samples. Five alternative FFM PD counts based upon the NEO-PI-R were computed and evaluated with the Assessment of DSM-IV Personality Disorders Questionnaire (ADP-IV; Schotte et al., 2004), both in terms of convergent and divergent validity. The best working count for each personality disorder was selected, and normative data was gathered, from which cut-off scores were derived. The validity of these cut-off scores and their usefulness as a screening tool was than tested against both a categorical and a dimensional measure of personality pathology (i.e., the DSM-IV and the DAPP-BQ, respectively).

One of the major proposed changes in the fifth edition of the DSM to the conceptualization of personality disorders includes the replacement of current personality disorder categories on Axis II with a taxonomy of dimensional maladaptive personality traits (Tackett et al., 2009). Unfortunately, this dimensional focus detracted attention from another important issue, namely the suitability of the criteria for measuring personality in later life (Oltmanns & Balsis, 2011). Apparently and regrettably, the later life context was not explicitly considered during the development of this new classification system either (Tackett et al., 2009). Analogous to study 1, we therefore set out to empirically investigate the age-neutrality of the *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012), the operationalization of the proposed DSM-5 traits (**Chapter 4**). Subsequently, we investigate its convergent validity by examining the joint hierarchical structure of the 25 proposed DSM-5

personality traits with the 18 dimensions of the Dimensional Assessment of Personality Pathology (DAPP; Livesley et al., 1992) model, a widely recognized and researched model that also focuses on pathological features of personality (**Chapter 5**).

In **Chapter 6** we aim to unravel the specific value of personality questionnaires in clinical geropsychology and geriatric psychiatry from a more contemplative perspective. The role of personality questionnaires is discussed within the broader context of personality assessment in older adults, and some critical reflections are made regarding the use of self versus informant reports, and the choice for an age-neutral versus an age-specific measurement system. Also, some recommendations are provided that should help psychiatrists, clinical geropsychologists, geriatricians and researchers in their search towards a better understanding of personality disorders in later life. Finally, the major findings of the abovementioned studies will be summarized and discussed from a broader perspective in the final chapter (**Chapter 7**), along with recommendations for further research and general conclusions.

1.13. References

Abrams, R. C., & Bromberg, C. E. (2007). Personality disorders in the elderly. *Psychiatric Annals, 37(2),* 123-127.

Abrams, R. C., & Horowitz, S. V. (1999). Personality disorders after age 50: A meta-analytic review of the literature. In E. Rosowsky, R. C., Abrams, & R. A. Zweig (Eds)., Personality disorders in older adults (pp. 55-68). Mahwah: Lawrence Erlbaum Associates.

Agronin, M. E., & Maletta, G.(2000). Personality disorders in later life: Understanding and overcoming the gap in research. *American Journal of Geriatric Psychiatry, 8*, 4-18.

Aldwin, C. M., & Levenson, M. R. (1994). Aging and personality assessment. *Annual Review of Gerontology and Geriatrics, 14*, 182-209.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Ames, A., & Molinari, V. (1994). Prevalence of personality disorders in community-living elderly. *Journal of Geriatric Psychiatry and Neurology, 7*, 189-194.

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

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging, 22(1),* 171-185.

Balsis, S., Segal, D.L., & Donahue, C.B.A. (2009). Revising the personality disorder diagnostic criteria for the diagnostic and statistical manual of mental disorders – fifth edition (DSM-V): Consider the later life context. *American Journal of Orthopsychiatry, 79*(4), 452-460.

Casey, D. A., & Schrodt, C. J. (1989). Axis II diagnoses in geriatric patients. *Journal of Geriatric Psychiatry and Neurology, 2*, 87-88.

Caspi, A., Bem, D. J., (1990). Personality continuity and change across the life course. In *Handbook of personality: Theory and Research,* Pervin, L. A. (ed.). Guilford: New York; 549-569.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58,* 227-257.

Cmielewski, M. & Watson, D. (2008). The heterogeneous structure of schizotypal personality disorder: item-level factors of the schizotypal personality questionnaire and their associations with obsessive-compulsive disorder symptoms, dissociative tendencies, and normal personality. *Journal of Abnormal Psychology, 117*, 364-376.

Costa, P. T. Jr., & McCrae, R. R. (2006). Age changes in personality and their origins: Comments on Roberts, Walton, and Viechtbauer. *Psychological Bulletin, 132(1),* 26-28.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI). Professional manual. Psychological Assessment Resources; Odessa, FL.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1997). Longitudinal stability of adult personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.). Handbook of personality psychology (pp. 269-292). San Diego: Academic Press.

Dierckx, E., (2012). Redactioneel: Vergrijzing: een nieuwe uitdaging voor de gezondheidszorg. *Tijdschrift Klinische Psychologie, 42(1),* 2-8.

Fogel, B. S., & Westlake, R. (1990). Personality disorder diagnoses and age in inpatients with major depression. *Journal of Clinical Psychiatry, 51*, 232-235.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48,* 26-34.

Haan, N., Millsap, R., & Hartka, E. (1986). As time goed by: change and stability in personality over fifty years. *Psychology and Aging, 1*, 220-232.

Harkness, A. R., McNulty, J. L., & Ben-Porath, Y. S. (1995). The personality psychopathology five (PSY-5): constructs and MMPI-2 scales. *Psychological Assessment, 7*, 104-114.

http://statbel.fgov.be/nl/modules/publications/statistiques/bevolking/Bevolking_op_1_jan_2007-2061.jsp

Hyer, L., Molinari, V., Mills, W. L., & Yeager, C. (2008). Personological assessment and treatment of older adults. In T. Millon, & C. Bloom. (Eds.). *The Millon Inventories* (p. 296-326). New York: The Guilford Press.

Kagan, J. (1969). The three faces of continuity in human development. In D.A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 53-65). Chicago: Rand McNally.

Kenan, M. M., Kendjelic, E. M., Molinari, V. A., Williams, W., Norris, M., & Kunik, M. E. (2000). Age-related differences in the frequency of personality disorders among inpatient veterans. *International Journal of Geriatric Psychiatry, 15,* 831-837.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42,* 1879-1890.

Livesley, W. J., Jackson, D. N., & Schroeder, M. L. (1992). Factorial structure of traits delineating personality disorders in clinical and general population samples. *Journal of Abnormal Psychology, 101(3),* 432-440.

Maiden, R. J., Peterson, S. A., Caya, M., & Hayslip, B. (2003). Personality changes in the old-old: a longitudinal study. *Journal of Adult Development, 10(1),* 31-39.

Maier, W., Lichtermann, D. Minges, J., Heun, R., Hallmayer, J., & Klingler, T. (1991). Unipolar depression in the aged: determinants of familial aggregation. *Journal of Affective Disorders, 23*, 53-61.

McCrae, R. R., & Costa, P. T., Jr. (2003). Personality in adulthood: A Five-Factor Theory perspective (2nd. Ed.). New York: Guilford Press.

Miller, J. D., Bagby, R. M., Pilkonis, P. A., Reynolds, S. K., & Lynam, D. R. (2005). A simplified technique for scoring DSM-IV personality disorders with the five-factor model. *Assessment, 12*, 404-415.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology, 55*, 1-22.

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology, 49*, 229-258.

Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology, 14*, 179-207.

Molinari, V., & Segal, D. L. (2011). Personality disorders: Description, aetiology, and epidemiology. In Abou-Saleh, M. T., Katona, C., & Kumar, A.

Principles and Practice of Geriatric Psychiatry (3rd ed.) (p. 649-654). John Wiley & Sons, Ltd.

Morey, L. C. (1991). Personality Assessment Inventory. Odessa, FL: Psychological Assessment Resources, Inc.

Morse, J. Q., & Lynch, T. R. (2000). Personality disorders in late life. *Current Psychiatry Reports, 2*, 24-31.

Mrockzek, D. K., Hurt, S. W., & Berman, W. H. (1999). Conceptual and methodological issues in the assessment of personality disorders in older adults. In E. Rosowsky, R. C., Abrams, & R. A. Zweig, (Eds.), Personality disorders in older adults (1st ed., pp. 135-150). Mahwah: Lawrence Erlbaum Associates.

Oltmanns, T. F. & Balsis, S. *Assessment of personality disorders in older adults.* In: Lichtenberg, P. A. (ed.). Handbook of Assessment in Clinical Gerontology. Academic/Elsevier; 2010. p. 101-122.

Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: Questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology, 7,* 321-349.

Paris, J. (2003). Personality disorders over time: precursors, course, and outcome. *Journal of Personality Disorders, 17,* 479-488.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 3-27.

Rosowsky, E., Abrams, R. C., & Zweig, R. A. (Eds). *Personality Disorders in Older Adults: Emerging Issues in Diagnosis and Treatment.* Mahwah, NJ: Erlbauw; 1999.

Schotte, C. K. W., De Doncker, D. M., Dmitruk, D., Van Mulders, I., D'Haenen, H., & Cosyns, P. (2004). The ADP-IV questionnaire: differential validity and concordance with semi-structured interview. Journal of Personality Disorders, 18(4), 405-419.

Segal, D. L., Coolidge, F. L., & Rosowsky, E. (2006). *Personality disorders and older adults: Diagnosis, assessment, and treatment.* New Jersey: Wiley.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84(5),* 1041-1053.

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: Developmental considerations for the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. *Development and Psychopathology, 21,* 687-713.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20(3),* 493-506.

Van Alphen, S. P. J., Derksen, J. J. L., Sadavoy, J., & Rosowsky, E. (2012). Editorial: Features and challenges of personality disorders in late life. *Aging & Mental Health, 16(7),* 805-810.

Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., & Derksen, J. J. L. (2006). The relevance of a geriatric sub-classification of personality disorders in the DSM-V. *International Journal of Geriatric Psychiatry, 21,* 205-209.

Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., & Derksen, J. J. L. (2004). Persoonlijkheidsstoornissen bij ouderen. Een overzicht. *Tijdschrift voor Psychiatrie, 46*, 145-156.

Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., Hoijtink, H. J. A., & Derksen, J. J. L. (2006). A preliminary study of the diagnostic accuracy of the Gerontological Personality Disorders Scale (GPS). *International Journal of Geriatric Psychiatry, 21,* 862-868.

Van Alphen, S. P. J., Nijhuis, P. E. P., & Oei, T. I. (2007). Antisocial personality disorder in older adults. A qualitative study of Dutch forensic psychiatrists and forensic psychologists. *International Journal of Geriatric Psychiatry, 22*, 813-815.

Van Alphen, S.P.J. (2006). *Diagnostiek van persoonlijkheidsstoornissen bij ouderen: een bijdrage aan de ontwikkeling van een screeningsinstrument*. Academisch proefschrift, Eindhoven: University Press Facilities.

Widiger, T. A., & Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: A proposal for DSM-V. *Annual Review of Clinical Psychology, 5,* 197-220.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: finding a common ground. *Journal of Personality Disorders, 19(2)*, 110-130.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62(2),* 71-83.

Zarit, S. H., Johansson, B., & Malmberg, B. (1995). Changes in functional competency of the oldest old : A longitudinal study. Journal of Aging and Health, 7, 3-23.

Zweig, R. A. (2008). Personality disorder in older adults: assessment challenges and strategies. *Professional Psychology, Research and Practice, 39 (3), 298-305*.

Chapter 2
Age-neutrality of the NEO-PI-R: Potential differential item functioning in older versus younger adults

Joke Van den Broeck, Gina Rossi, Eva Dierckx, & Barbara De Clercq

Abstract

Geriatric researchers and clinicians often have to deal with a lack of valid personality measures for older age groups (e.g., Mroczek, Hurt, & Berman, 1999; Zweig, 2008), which hampers a reliable assessment of personality in later life. An age-neutral measurement system is one of the basic conditions for an accurate personality assessment across the lifespan, both longitudinally and cross-sectionally. In the present study, we empirically investigate the age-neutrality of one of the most widely used personality measures (i.e., the NEO PI-R (Costa & McCrae, 1992)), by examining potential Differential Item Functioning (DIF). Overall, results indicate that the vast majority (92.9% at domain-level and 95% at facet-level) of the NEO PI-R items was similarly endorsed by younger and older age groups with the same position on the personality trait of interest, corroborating the NEO PI-R's age neutrality. However, Differential Test Functioning (DTF) analyses revealed large DTF for Extraversion, and facet A6 (Tender-Mindedness). Results are discussed in terms of their implications for using the current format of the NEO PI-R in older aged samples.

Published in: Journal of Psychopathology and Behavioral Assessment, 34(3), 361-369

2.1. Introduction

Although personality traits are commonly defined as relatively enduring patterns of thoughts, feelings, and behaviors that distinguish individuals from one another, the personality stability issue has been the subject of considerable controversy in personality research. Two lines of research can be distinguished within this debate (Balsis, Gleason, Woods, & Oltmanns, 2007). In the 1980s, the assumption of personality stability throughout adulthood has been systematically put forward by Costa and McCrae (e.g., Costa & McCrae, 1988), and since the late 1980s and early 1990s many personality psychologists opted for the Five Factor Theory (FFT) (e.g., Digman, 1990). In terms of personality traits, the FFT clearly states that traits develop through childhood and stop changing by the age of 30 and that this pattern holds across different cultures (Roberts, Walton, & Viechtbauer, 2006). However, more recent studies examined age trends across the five broad personality factors and found evidence for age differences and systematic age-related changes in personality traits during late adulthood (e.g., Terracciano, McCrae, Brant, & Costa, 2005; Allemand, Zimprich, & Hendriks, 2008). In general, these studies showed that Neuroticism, Extraversion, and Openness to experience tend to remit with age, whereas Agreeableness and Conscientiousness tend to increase (e.g., Costa, McCrae, Zonderman, Barbano, Lebowitz, & Larson, 1986; Costa & McCrae, 1986, 1988; Terraciano et al., 2005). Despite these slight but consistent differences found in both cross-sectional and longitudinal data, the overall trend still suggests that traits are stable over time (Tackett, Balsis, Oltmanns, & Krueger, 2009; Balsis et al., 2007). From a more dynamic context-dependent view, it is assumed that personality may change as a result of complex interactions between biological and socio-cultural influences. From this point of view, personality is not a static construct in adulthood, but represents a constant and active process that extends across the entire life course, and with each age period having its own

developmental agenda (Baltes, Lindenberger, & Staudinger, 1998). Several studies confirm the idea that personality can change, especially when people are confronted with important life events and need to cope with their changing lives (Haan, Millsap, & Hartka, 1986; Maiden, Peterson, Caya, & Hayslip, 2003). However, the life span perspective states that the effects of psychological, social and cultural factors diminish as people grow older, often as a result of selection, optimization, and compensation processes (Baltes et al., 1998).

Whereas these two perspectives (trait stability versus context dependency) were traditionally seen as incompatible, Balsis and colleagues recently suggested that "they may address different, albeit related phenomena that operate simultaneously" (Balsis et al., 2007, p. 180). From this view, the trait tradition addresses the underlying latent structure of personality, whereas the more dynamic context-dependent tradition addresses the changing presentation of personality across situations and time and defines personality as a dynamic construct (Balsis et al., 2007; Mischel, 1969, 2004). The assumption that the manifestation of personality can change while the underlying traits remain stable has also been referred to as 'heterotypic continuity' (e.g., Caspi & Bem, 1990; Kagan, 1969; Mroczek et al., 1999). This so-called 'heterotypic continuity' can appear in at least three different ways (Balsis et al., 2007): Personality trait manifestation may show natural developmental change, the context wherein these traits exist may change, or the opportunities for the presentation of the features may change over time (Mroczek et al., 1999). (For an illustration of each of these possibilities applied to personality disorders, we refer to Balsis et al., 2007).

Given that the context of younger and older adults meaningfully differs in terms of social, occupational, financial, physiological and cognitive aspects, one can assume that the presentation of personality may change as people age (Tackett et al., 2009). However, most of the current personality measures have been developed with younger adults in mind

and have not taken the specific later life context into account, potentially lacking face validity for the assessment of personality in older adults (Agronin & Maletta, 2000; Zweig, 2008; Tackett et al., 2009; Abrams & Bromberg, 2007). As one of the few age-neutral intended personality measurements, the later life context was closely considered during the development of the NEO PI-R (Costa et al., 1986; McCrae & Costa, 1987; see Tackett et al., 2009), leading one to expect that this inventory should measure each personality trait equally well across younger and older age groups.

Although we do acknowledge that the NEO PI-R purports to be age-neutral, we also note that its age-neutrality has, as far as we know, not been empirically investigated. Yet the answer to this question is of important value. Not only for geriatric researchers and clinical practitioners in geriatric settings, whose work is often hampered by a lack of valid personality assessment tools, but also for researchers interested in studying the course of personality, and for those aiming to construct an age-related personality theory (Oltmanns & Balsis, 2010). More specifically, measurement invariance is an important prerequisite for reliable and valid comparisons of personality profiles across age. In case of the NEO PI-R, a lack of measurement invariance at item level would question the comparability of facet and/or domain scores across age (e.g., Church, Alvarez, Mai, French, Katigbak, & Ortiz, 2011).

The present study addresses this issue and empirically investigates the age-neutrality of the NEO PI-R items, exploring the role of specific age-related measurement invariance in response sets. Relying on Differential Item Functioning (DIF; Millsap & Everson, 1993) analyses[3], it will be examined whether there are NEO PI-R items that measure the personality construct of interest differently in younger versus older

[3] DIF analyses are well suited to detect how systematically biased an item is for one group versus an other group, controlling for true group-mean differences (Balsis et al., 2007, p. 172).

adults, after controlling for the overall differences between both age groups (e.g., Holland & Wainer, 1993; Thissen, 2001; Zumbo, 2007). More specifically, if younger and older adults with a similar position on a trait dimension do not have the same probability of endorsing an item, the item is said to exhibit DIF (Edwards & Edelen, 2009). Consequently, test scores based on items exhibiting DIF can lead to potentially misleading group differences (Holland & Wainer, 1993).

Several statistical approaches have been proposed for the analysis of DIF, both within Classical Test Theory (CTT) and Item Response Theory (IRT) (Edwards & Edelen, 2009). Current analyses were conducted using an odds ratio approach (CTT), which is, in contrast to the IRT approach, not hampered by requirements of model fit and large sample sizes, and can be conducted using the easily accessible DIFAS program (Penfield, 2005).

2.2. Method

Participants and Procedure

The sample of younger adults consisted of a Dutch speaking community sample of 411 adults, ranging from 18 to 40 years (M= 28.28, SD= 7.00), and with 44% male participants. Data came primarily from the normative sample of the NEO PI-R gathered in the Netherlands and Flanders (Hoekstra, Ormel, & De Fruyt, 2007), and was extended with data that were collected by college students in return for course credit. The sample of older adults comprised a total of 434 adults ranging from 65 to 92 years (M= 72.17, SD= 5.74), and 55% male participants. A cut-off of 65 was chosen for inclusion in the older age group, because this age is commonly used as the demarcation point for the start of later life in research literature (Segal, Coolidge, & Rosowsky, 2006). Participants were voluntarily recruited from leisure clubs and senior meetings. All participants volunteered their participation and provided a signed informed consent.

Measure

The Dutch authorized version of the Revised NEO Personality Inventory (Hoekstra et al., 2007), a 240-item questionnaire designed to assess five broad domains of personality as conceptualized by the Five Factor Model (FFM) of personality, was used in this study to assess self-reported personality traits. Items are rated on a 5-point Likert format scale, ranging from strongly disagree to strongly agree. The NEO PI-R measures six specific lower-level facets belonging to each of the five overarching domains of the FFM (Neuroticism, Extraversion, Openness to experience, Agreeableness, and Conscientiousness). In the younger sample, internal consistencies of the domain scales were adequate, and ranged from .87 (Extraversion, Openness to experience, Agreeableness) to .92 (Neuroticism) with a median value of .87. Cronbach's α coefficients for the 8-item facet scales ranged from .59 (Tender-Mindedness) to .84 (Anxiety) (median coefficient a = .72). For the domain scales in the older sample, internal consistency reliabilities ranged from .84 (Extraversion) to .91 (Neuroticism) with a median value of .88. Cronbach's a coefficients for the facet scales in this sample ranged from .53 (Excitement-Seeking and Values) to .80 (Anxiety) (median coefficient a = .70). These values are consistent with normative data reported in the NEO PI-R manuals (Costa & McCrae, 1992; Hoekstra et al., 2007). Principal component analysis of the 30 NEO PI-R facet scales, followed by varimax rotation, produced a factor-loading matrix that was highly comparable to the structure obtained in previous studies (e.g., De Fruyt, Mervielde, Hoekstra, & Rolland, 2000; Savla, Davey, Costa, & Whitfield, 2007). In both the younger and older sample, parallel analysis (Horn, 1965) determined that five factors should be retained, explaining 60.14% and 57.49% of the total variance, respectively.

Data Analyses

Differences at the NEO PI-R domain level between the younger and the older age sample were tested with t-tests for independent samples. We used Cohen's *d* as a measure of effect size (Cohen, 1988), with $r \geq .20$ indicating a small effect, $r \geq .50$ a medium effect, and $r \geq .80$ a large effect.

In order to detect possible DIF as a function of age, the Mantel Chi-square, the Liu-Agresti Cumulative Common Log-Odds Ratio (L-A LOR), and the Cox's Noncentrality Parameter Estimator (Cox's B) were used for polytomous items, using the DIFAS 5.0 software program (Penfield, 2007b). The Mantel chi-square statistic (Mantel, 1963; Zwick, Donoghue, & Grima, 1993), an extension of the general Mantel-Haenszel statistic for dichotomous items (Mantel & Haenszel, 1959), is based on a group (2) x response option (5) contingency table for each item. The statistic is distributed as a chi-square with one degree of freedom, with a higher chi-square value of a particular item indicating a higher probability for this item to display DIF. In the current study, the chi-square statistic was used as a first step in the process of detecting DIF, to highlight the items with potential DIF. In addition, two alternative measures of DIF, the L-A LOR (Liu & Agresti, 1996; Penfield & Algina, 2003) and the Cox's B (Camilli & Congdon, 1999) statistics, were consulted to estimate the effect size of the DIF. All three statistics use an item-level (omnibus) approach of DIF evaluation in polytomous items, which addresses item-level invariance and measures the overall effect across all score levels (Gattamorta, 2009). For both the L-A LOR and Cox's B statistic negative values indicate DIF against the reference group (younger adults), and positive values indicate DIF against the focal group (older adults). The following cut-off criteria are available to flag items with large DIF: $|\text{L-A LOR}| > .64$ (Penfield, 2007a), and $|\text{Cox's B}| > .40$ (Camilli & Congdon, 1999). The impact of DIF at scale level was examined by Differential Test Functioning (DTF) analyses. Penfield and Algina (2006) propose to define DIF effect variance as small for $v^2 < .07$, medium for $.07 \leq v^2 < .14$, and large for $v^2 > .14$.

In order to make nuanced statements about the age-neutrality of the NEO PI-R, we conducted DIF and DTF analysis at both domain- and facet-level. At the domain-level, the total domain scores were used as stratifying or matching variables. The stratum size was set at 5, in order to avoid too many empty cells or strata. To reduce the Type I error, we used a Bonferroni corrected critical chi-square value of 12.78 ($p < .05$ corrected to $p < .00035$). The Bonferroni correction was applied across all 48 items for each domain by three test statistics (.05/48*3). Similar stringent Bonferroni cut-offs were applied to the LA-Lor (> 1.08) and Cox's B (> .68) statistics. Similarly, the total facet scores were used as stratifying variables for the facet-level analyses. Here the stratum size was set at 1, which is the default option in DIFAS 5.0. A Bonferroni corrected critical chi-square value of 9.55 ($p < .002$; .05/8*3) was applied, and the following adjusted cut-off criteria to flag items with large DIF were used: |LA-Lor| > .92 and |Cox's B| > .58.

2.3. Results

Descriptive Statistics

Table 1 shows the Cronbach alpha reliability estimates, means, standard deviations and the effect sizes of the domain and facet NEO PI-R scale scores for the younger and older age groups. All mean scale domain scores of the older sample differed significantly ($p < .001$) from the mean scale scores of the younger sample, with three differences showing a moderate effect size ($d > .50$). At the domain level, younger adults scored significantly higher than older adults on the Neuroticism, $t(782) = 4.023$, $p < .001$, Extraversion, $t(756) = 8.692$, $p < .001$, and Openness, $t(774) = 9.028$, $p < .001$ scales. In contrast, when compared with younger adults, older adults displayed higher levels of Agreeableness, $t(771) = -7.235$, $p < .001$, and Conscientiousness, $t(771) = -3.910$, $p < .001$. At facet level, large significant differences were found for Excitement-Seeking (E5) and Values (O6), and moderate significant differences for Impulsiveness (N5), Gregariousness (E2),

Positive Emotions (E6), Fantasy (O1), Feelings (O3), Actions (O4), Straightforwardness (A2), and Dutifulness (C3). Only for the latter two the older adults had higher mean scale scores compared to the younger adults. For all facets, the direction of the effect was the same as for their respective factor.

Table 1 *Descriptive statistics for the NEO PI-R domain and facet scales for the younger (n=411) and older (n=434) sample*

	Coefficient Alpha		Raw Means (SD)		Effect size
Domain	Younger	Older	Younger	Older	Cohen's *d*
Neuroticism	.92	.91	132.45 (22.43)	126.16 (21.28)	.29*
N1: Anxiety	.84	.80	23.10 (5.81)	22.18 (5.48)	ns
N2: Angry Hostility	.74	.70	20.41 (4.67)	20.16 (4.35)	ns
N3: Depression	.80	.73	22.71 (5.39)	21.68 (4.89)	ns
N4: Self-consciousness	.72	.70	22.22 (4.87)	21.36 (4.64)	ns
N5: Impulsiveness	.71	.61	25.17 (4.73)	22.38 (4.18)	.63*
N6: Vulnerability	.80	.77	19.09 (4.77)	19.06 (4.32)	ns
Extraversion	.87	.84	160.08 (18.89)	149.16 (15.84)	.62*
E1: Warmth	.70	.69	28.77 (4.13)	29.50 (3.76)	ns
E2: Gregariousness	.77	.73	27.05 (5.33)	24.07 (5.28)	.56*
E3: Assertiveness	.81	.72	23.18 (5.36)	23.51 (4.99)	ns
E4: Activity	.68	.64	26.50 (4.31)	25.11 (4.24)	.33*
E5: Excitement-Seeking	.63	.53	24.92 (4.78)	19.91 (4.23)	1.11*
E6: Positive Emotions	.77	.73	29.22 (4.95)	26.75 (4.57)	.52*
Openness	.87	.85	159.49 (18.62)	148.02 (16.86)	.64*
O1: Fantasy	.79	.71	26.31 (5.08)	22.91 (4.49)	.71*
O2: Aesthetics	.74	.76	25.56 (5.36)	26.25 (5.58)	ns
O3: Feelings	.69	.64	28.82 (4.20)	26.52 (3.96)	.56*
O4: Actions	.61	.60	23.87 (4.21)	21.09 (4.20)	.66*
O5: Ideas	.77	.68	26.06 (5.26)	25.71 (4.88)	ns
O6: Values	.61	.53	29.06 (3.91)	25.66 (3.86)	.88*
Agreeableness	.87	.88	167.65 (17.40)	176.58 (16.92)	-.52*
A1: Trust	.77	.72	27.85 (4.40)	28.95 (3.95)	-.26*
A2: Straightforwardness	.75	.63	27.61 (5.05)	30.16 (4.22)	-.55*
A3: Altruism	.65	.75	30.08 (3.60)	30.75 (3.80)	ns
A4: Compliance	.68	.66	24.74 (4.54)	26.69 (4.43)	-.43*
A5: Modesty	.74	.75	27.96 (4.85)	29.52 (4.36)	-.34*
A6: Tender-Mindedness	.59	.65	29.29 (3.61)	30.82 (3.83)	-.41*
Conscientiousness	.90	.89	164.58 (19.49)	169.77 (17.53)	-.28*
C1: Competence	.64	.70	28.43 (3.32)	27.87 (3.62)	ns
C2: Order	.67	.56	25.82 (4.55)	26.15 (3.92)	ns
C3: Dutifulness	.61	.71	30.30 (4.04)	33.26 (3.62)	-.77*
C4: Achievement Striving	.78	.71	26.45 (5.11)	26.62 (4.38)	ns
C5: Self-Discipline	.75	.66	27.64 (4.62)	27.98 (3.99)	ns
C6: Deliberation	.79	.76	25.80 (5.07)	27.39 (4.60)	-.33*

Note. * $p < .001$

Differential Item Functioning: Domain Level

To investigate the possible presence of DIF at domain-level, five separate analyses were conducted, with the five total domain-scores serving as stratifying variables. Overall, the initial Mantel Chi-square test of all 240 NEO PI-R items revealed DIF at a stringent cut-off ($p <$ .00035) for 73 items (30.4%). The Bonferroni adjusted L-A Lor (> 1.08) and the Cox's B (> .68) statistics confirmed large DIF for 17 items (7.1%). The majority of these items (11) showed DIF against the younger sample, indicating they were more readily endorsed by older adults, despite equal levels of the underlying personality trait (i.e., Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness, respectively). These items were divided across five facets: E1 (Warmth; 2 items), E3 (Assertiveness; 1 item), O2 (Aesthetics; 4 items), C3 (Dutifulness; 3 items), and C6 (Deliberation; 1 item). Six items exhibited DIF against the older sample, indicating they were more readily endorsed by younger adults. These six items were divided across the following four facets: N5 (Impulsiveness; 1 item), E5 (Excitement-seeking; 3 items), O1 (Fantasy; 1 item), and C1 (Competence; 1 item). We refer to Table 2 for the paraphrased item content and statistical details of the items exhibiting large DIF resulting from these domain-level DIF analyses.

To evaluate the impact of DIF at scale level, DTF was investigated. Weighted v^2 values were .17, .35, .33, .16, and .24 for the Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness domains, respectively. In line with the above-mentioned stringent Bonferroni cut-off criteria for large DIF, we adjusted the more flexible thresholds proposed by Penfield and Algina (2006) to $v^2 < .18$ for small, $.18 \leq v^2 < .35$ for moderate, and $v^2 \geq .35$ for large DIF effect variance[4].

[4] Our rationale for this was the following: in deriving their thresholds Penfield and Algina (2006) argued that a collective large level of DIF in a group of items exist if 25% or more of the items are categorized as having moderate or large magnitudes of DIF based on the ETS classification scheme (i.e. if 25% or

This Bonferroni corrected cut-offs resulted in small DTF for Neuroticism and Agreeableness, moderate DTF for Openness and Conscientiousness, and large DTF for Extraversion.

more of the items have an absolute value of log(αMH) greater than or equal to .43. They also suggest that MH and LA-Lor have similar meanings in terms of DIF magnitude. Because we wanted to reduce the Type I error, a Bonferroni correction was applied and an adjusted LA-Lor cut-off value of 1.08 (instead of .64) was used to flag items with large DIF. In line, we made a similar adjustment for the DTF thresholds. For example: Penfield and Algina consider the variance of DIF effect large when weighted $v^2 > .14$, using an LA-Lor value of .43 as critical value. Since we adhere to a stringent LA-Lor critical value (> 1.08) we adjusted this to $v^2 > .35$ (i.e. .14/.43*1.08).

Table 2 *DIF analyses at domain-level: items meeting Bonferroni adjusted criteria for large DIF*

Item	Paraphrased content	Facet	Mantel X^2	L-A LOR	Cox's B
	NEUROTICISM				
111.	I tend to eat too much.	N5	69.18	1.12	0.51
	EXTRAVERSION				
2.	I really like most people I meet	E1	62.86	-1.12	-0.67
122.	I really enjoy talking to people	E1	44.06	-1.07	-0.69
72.	I have often been a leader of groups	E3	82.30	-1.26	-0.59
82.	Done things just for "kicks" or "thrills".	E5	63.61	1.08	0.52
112.[R]	I tend to avoid movies that are shocking	E5	91.23	1.30	0.53
172.	I love the excitement of roller coasters	E5	102.59	1.38	0.60
	OPENNESS				
33.[R]	Keep thoughts realistic and avoiding flights of fancy	O1	62.11	1.06	0.57
8.[R]	Aesthetic and artistic concerns aren't important	O2	102.05	-1.47	-0.63
98.	Intrigued by the patterns in art and nature	O2	94.76	-1.43	-0.76
128.[R]	Poetry has little or no effect on me	O2	56.73	-1.07	-0.52
188.	Poetry or art can give me a wave of excitement	O2	64.78	-1.15	-0.58
	CONSCIENTIOUSNESS				
185.	I'm a very competent person	C1	113.46	1.62	0.99
75.	I pay my debts promptly and in full	C3	64.64	-1.24	-0.64
105.[R]	Sometimes I cheat when I play solitaire	C3	60.72	-1.09	-0.51
165.	I adhere strictly to my ethical principles	C3	66.56	-1.10	-0.66
210.	I plan ahead carefully when I go on a trip	C6	73.02	-1.16	-0.62

Note. N5= Impulsiveness, E1= Warmth, E3= Assertiveness, E5= Excitement-Seeking, O1= Fantasy, O2= Aesthetics, C1= Competence, C3= Dutifulness, C6= Deliberation. L-A LOR= Liu-Agresti Common Log Odds Ratio. Cox's B= Cox's Noncentrality Parameter Estimator. Negative values indicate DIF against younger adults, positive values indicate DIF against older adults. [R] Reversed scored items.

Differential Item Functioning: Facet-Level

To investigate the possible presence of DIF at facet-level, 30 separate analyses were conducted, with the respective total facet-score serving as stratifying variable in each case. Taking into account our stringent Bonferroni adjusted criteria ($p < .002$; Mantel $X^2 > 9.55$, LA-Lor > .92, and Cox's B > .58) these analyses revealed 12 items (5%) displaying large DIF, divided across ten facets. The majority of these items (10) showed DIF against the younger sample, indicating they were more readily endorsed by older adults matched on the underlying personality trait (in this case: total facet scores). These ten items were divided across eight facets: E3 (Assertiveness; 1 item), E4 (Activity; 1 item), O6 (Values; 1 item), A1 (Trust; 1 item), A2 (Straightforwardness; 1 item), A6 (Tender-Mindedness; 2 items), C1 (Competence; 1 item), C2 (Order; 1 item), and C6 (Deliberation; 1 item). Two items, one from the Aesthetics (O2) and one from the Competence (C1) scale, exhibited DIF against the older sample. As can be noted, facet C1 (Competence) contained two items displaying DIF: one against the younger sample and one against the older sample. Table 3 shows the paraphrased item content and statistical details of the items displaying large DIF at facet-level.

DTF was also investigated at facet-level. We used the following adjusted thresholds to interpret the impact of DIF at facet level: $v^2 < .15$: small; $.15 \leq v^2 < .30$: moderate; and $v^2 \geq .30$: large[5]. DTF was large for facet A6 (Tender-Mindedness; .37), and moderate for the other nine facets (E3, Assertiveness: .19; E4, Activity: .17; O2, Aesthetics: .28 ; O6, Values: .24; A1, Trust: .22; A2, Straightforwardness: .20; C1, Competence: .17; C2, Order: .20; C6, Deliberation: .23).

[5] A similar reasoning was handled at facet-level. Here we used a stringent LA-Lor critical value of .92, leading to an adjusted $v^2 \geq .30$ for large DTF (i.e. .14/.43*.92).

Table 3 *DIF-analyses at facet-level: items meeting Bonferroni adjusted criteria for large DIF*

Item	Paraphrased context	Facet	Mantel X^2	L-A Lor	Cox's B
	EXTRAVERSION				
72.	I often have been a leader of groups	E3	55.6	-1.06	-.64
227.	I am a very active person	E4	41.2	-.95	-.69
	OPENNESS				
38.	Sometimes completely absorbed in music	O2	50.08	.97	.51
58.	Law and social policies should change	O6	52.15	-1.14	-.71
	AGREEABLENESS				
214.	Faith in human nature	A1	30.69	-.84	-.60
69.	I couldn't deceive anyone	A2	50.29	-.98	-.55
29.	Awareness of political leaders for human aspect	A6	44.12	-.97	-.59
59.[R]	Hard-headed and tough-minded attitudes	A6	55.08	1.04	.58
	CONSCIENTOUSNESS				
5.	Known for prudence and common sense	C1	38.89	-.93	-.66
185.	I'm a very competent person	C1	102.24	1.61	1.08
100.	I like to keep everything in it's place	C2	61.72	-1.20	-.69
210.	I plan ahead carefully when I go on a trip	C6	63.62	-1.12	-.61

Note. E3= Assertiveness, E4= Activity, O2= Aesthetics, O6= Values, A1= Trust, A2= Straightforwardness, A6= Tender-Mindedness, C1= Competence, C2= Order, C6= Deliberation. L-A LOR= Liu-Agresti Common Log Odds Ratio. Cox's B= Cox's Noncentrality Parameter Estimator. Negative values indicate DIF against younger adults, positive values indicate DIF against older adults. [R] Reversed scored items.

2.4. Discussion

The primary aim of the current study was to empirically investigate the age-neutrality of the NEO PI-R, one of the most prominent FFM personality inventories. By examining whether the response tendency on the same set of items is different in younger versus older adults, we addressed the question of measurement invariance across age in order to

verify whether the NEO PI-R is an appropriate and reliable measure of personality for use in both younger and older adults. Overall, the present findings corroborate the NEO PI-R's age-neutrality, since more than 92% of the items was similarly endorsed by younger and older adults that share the same position on the underlying personality trait. The percentage of items displaying DIF (7.1% at domain-level and 5% at facet-level) are considerably less than the threshold of 25% of items put forward by Penfield & Algina (2006) to indicate that the instrument as a whole may yield biased results. In general, current findings justify the comparability of NEO PI-R profiles across age.

The different mean scores between older and younger adults on all of the NEO PI-R domain scales are consistent with previous research on mean-level change in personality trait scores, showing a decline in Neuroticism, Extraversion and Openness to experience, and a small increase in Agreeableness and Conscientiousness over the lifespan (e.g., Costa et al., 1986; Costa & McCrae, 1986, 1988; Terraciano et al., 2005). At the facet level, 21 of the 30 facets showed significantly different mean scale scores for younger versus older adults, with eight differences showing a small effect ($d > .20$), eleven differences showing a moderate effect ($d > .50$), and two differences showing a large effect ($d > .80$).

In general, parallels can be drawn between current results and those obtained by Terracciano et al. (2005). They examined age trends in the five domains and 30 facets of the NEO-PI-R by means of Hierarchical Linear Modeling (HLM) analyses and found gradual personality changes across the lifespan. As concerns the Openness to experience facets, for example, we found a large mean-level difference between younger and older adults on the Openness to Values (O6) scale, confirming the idea that older adults are less willing to re-examine social, political, and religious values (Krosnick & Alwin, 1989; see Terracciano et al., 2005, p. 9). Additionally, no differences were found for the Openness to Aesthetics (O2) and Ideas (O5) facets, underscoring the findings of

Terracciano et al. (2005) that these traits remain relatively stable from age 30 to 90.

Mean-level change, or whether a group of people increases or decreases on trait dimensions over time, is often assimilated with normative change in personality (Roberts et al., 2006). Shared maturational or historical processes, or engagement in normative life tasks and roles are thought to induce these mean-level changes. This explains the increases found in traits associated with psychological maturity, such as Agreeableness, Conscientiousness and Emotional stability (Roberts et al., 2006). Despite these generalizable patterns of personality development at group-level, the trait model assumes stability in terms of latent personality traits at the individual level. Response sets reflect the position of a person on the latent trait being measured, and should not be affected by age. Therefore, developing an item set and not taking into account the fact that the externalization of latent traits can differ across the life course (cfr. 'heterotypic continuity'), leads to poor face validity and can in turn affect the content validity of the entire scale (Tackett et al., 2009). To address this issue, DIF and DTF were conducted at both domain- and facet-level.

Analyses at domain-level revealed 17 items exhibiting large DIF between the two age groups. Although this number of DIF items comprise only a minority of the total NEO PI-R item set, these findings may raise questions about the metric equivalence of these particular items across age groups (Tackett et al., 2009) and, perhaps more importantly, their impact on the validity of the scales they represent. To investigate in more detail the magnitude of the DIF effect variance at scale level, we performed DTF analyses for each domain separately, revealing large DTF for the Extraversion domain. Items exhibiting DIF in this domain were mainly from the Excitement-Seeking (E5) and Impulsiveness (E1) facets (e.g. 'I love the excitement of roller coasters' and 'I have done things just for kicks or thrills'). With three of the eight items displaying DIF against older adults, the Excitement-seeking scale (E5) may lack

face validity for assessing Extraversion in an older sample. These findings support the notion that older adults may have fewer opportunities to manifest reckless and impulsive behavior, due to physical age-related impairment (Abrams & Bromberg, 2006; Segal et al., 2006). Poor health, financial issues and mobility problems may be related to this natural decline of impulsivity and excitement-seeking with advancing age (Maiden et al., 2003). Although just beneath the large DTF threshold, five items of the Openness to experience domain display DIF, stemming mainly from the Aesthetics (O2) facet. Those items (e.g., 'Poetry has a great effect on me', 'Art can give me a wave of excitement' or 'Intrigued by the patterns in art and nature') showed DIF against younger adults, suggesting that older adults might be in general more interested and moved by art, poetry, and beauty than younger adults with the same position on the Openness trait. This finding may point to a generation gap with regard to cultural experience. Adolescents and young adults may be more interested and influenced by popular culture, coming in the form of, for example, music or television, whereas older adults might be in general more interested in traditional culture (van den Broek & De Haan, 2000). Overall, current DIF effect variance analyses suggest that domain-level comparisons between younger and older adults may reveal potentially misleading group differences on Extraversion, and, to a lesser extent, Openness, with a v^2 value of .33, just below the .35 threshold. At facet-level, Tender-Mindedness (A6) was the only facet that displayed large DTF. Older adults more readily endorsed two items belonging to this facet ('Awareness of political leaders for human aspect' and the reversed scored item 'Hard-headed and tough-minded attitudes') exhibited large DIF. Caution is however recommended in drawing premature conclusions about age differences based on these above-mentioned personality traits.

From a methodological perspective, it is important to note that the presence of DIF in itself is not problematic (Baer, Samuel, & Lykins, 2011). In interpreting DIF, a distinction should be made between item

bias and item impact (Ackerman, 1992). Because DIF is a required, but not sufficient condition for item bias, caution is warranted in drawing premature conclusions about the possibility of age-bias in some of the NEO PI-R domains based on current findings. The above-mentioned DIF results may point to possible bias within the items (i.e., the item measures something else than the construct of interest), but additional investigations are necessary to clarify the role of potential linguistic and/or cultural influences. The presence of DIF may also indicate an item with high impact, due to real differences in the manifestation of the underlying trait being measured across age-groups (Ackerman, 1992). As such, current group differences could be due to real age-differences, other than inherent to the construct of interest being measured (Ackerman, 1992). Further research is therefore needed to replicate current results and to more thoroughly explore the possible causes of the present age DIF and its impact.

A number of limitations should be kept in mind when considering the present findings. As mentioned above, it is difficult to explain the underlying causes of DIF. The cross-sectional methodology makes it impossible to ascertain the extent to which the present results are influenced by cohort effects rather than real age-related differences. Also, because of the item –level (omnibus) approach of DIF testing used in this study, results do not inform us about which specific score levels are manifesting DIF. Therefore, further research is needed to detect *differential step functioning,* or the manifestation of DIF at any particular score level, for example using the graded response model (Cohen, Kim, & Baker, 1993) within an IRT framework, or a common log odds ratio approach (Penfield, 2007b). Another limitation are the broad age ranges in the samples of the current study. We are well aware of the fact that: "There are wide variations in life experiences, physical challenges, psychological experiences, and social opportunities between the "young-old" (usually defined as those between the ages of 65 and 74), the "old-old" (between the ages of 75 and 84) and the "oldest-old" (85 years of age and older)" (Segal et al., 2006, p. 2). Moreover, the data

gathered in the current study are samples of convenience. Further research should therefore take into account the heterogeneity of this population by comparing smaller subgroups that are homogeneous in terms of age, preferably in stratified samples. Further large-scale research is also needed, to thoroughly examine the presence and consequences of DIF in the NEO PI-R using different age groups, in order to fully explore its age-neutrality and its usefulness in successive age groups.

In sum, the present study is the first to empirically validate the NEO PI-R's age neutrality. DIF and DTF analyses indicate that the majority of the NEO PI-R items are equally endorsed by younger and older adults with the same level of underlying personality trait. Only a small number of items display DIF, mainly stemming from the Excitement-Seeking (E5), Impulsiveness (E1), Aesthetics (O2), and Tender-Mindedness (A6) facets. A more elaborated study of these differently behaving items and the possible causes of current age DIF is warranted, as this will further contribute to the research on the course of personality throughout the lifespan.

2.5. References

Abrams, R. C., & Bromberg, C. E. (2006). Personality disorders in the elderly: A flagging field of inquiry. *International Journal of Geriatric Psychiatry, 21,* 1013-1017.

Abrams, R. C., & Bromberg, C. E. (2007). Personality disorders in the elderly. *Psychiatric Annals, 37(2),* 123-127.

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*, 67-91.

Agronin, M. E., & Maletta, G. (2000). Personality disorders in late life: understanding and overcoming the gap in research. *American Journal Geriatric Psychiatry, 3,* 4-18.

Allemand, M., Zimprich, D., & Martin, M. (2008). Long-term correlated change in personality traits in old age. *Psychology and Aging, 23(3),* 545-557.

Baer, R. A., Samuel, D. B., & Lykins, E. L. B. (2011). Differential item functioning on the Five Facet Mindfulness Questionnaire is minimal in demographically matched mediators and nonmediators. *Assessment, 18(1),* 3-10.

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging, 22(1),* 171-185.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (1998). Life-span theory in developmental psychology. In Damon, W., & Lerner, R. M. (Eds.)*, Handbook of Child Psychology. Theoretical Models of Human Development. Fifth edition.* (p. 1029-1143).

Camilli, G., & Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. *Journal of Educational and Behavioral Statistics, 24,* 323-341.

Caspi, A., & Bem, D. J. (1990). Personality continuity and change across the life course. In Pervin, L. A. (Ed.), *Handbook of personality: Theory and research* (p. 549-569). New York: Guilford.

Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the Revised NEO Personality Inventory. *Journal of personality and social psychology, 101(5),* 1068-1089.

Cohen, A. S., Kim, S. H., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement, 17,* 335-350.

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

Costa, P. T. Jr., & McCrae, R. R. (1986). Personality stability and its implications for clinical psychology. *Clinical Psychology Review, 6,* 407-423.

Costa, P. T. Jr., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54,* 853-863.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Professional Manual: Revised NEO Personality Inventory (NEO PI-R) and NEO Five-factor Inventory (NEO-FFI).* Odessa, FL: Psychological Assessment Resources.

Costa, P. T. Jr., McCrae, R. R., Zonderman, A. B., Barbano, H. E., Lebowitz, B., & Larson, D. M. (1986). Cross-sectional studies of personality in a national sample: 2. Stability in neuroticism, extraversion, and openness. *Psychology and Aging, 1,* 144-149.

De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J-P. (2000). Assessing adolescents' personality with the NEO PI-R. *Assessment, 7(4),* 329-345.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417-440.

Edwards, M. C., & Edelen, M. O. (2009). Special topics in item response theory. In Millsap, R. E., & Maydeu-Olivares, A. (Eds.), *The SAGE handbook of quantitative methods in psychology.* (p. 178-185). London: SAGE Publication Ltd.

Gattamorta, K. A. (2009). *A comparison of adjacent categories and cumulative DSF effect estimators.* A dissertation submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Coral Gables, Florida: University of Miami.

Haan, N., Millsap, R., & Hartka, E. (1986). As time goes by: Change and stability in personality over fifty years. *Psychology and Aging, 1,* 220-232.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (2007). *NEO PI-R en NEO-FFI. Handleiding.* Amsterdam: Hogrefe.

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning.* Hillsdale, NJ: Erlbaum.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Kagan, J. (1969). The three faces of continuity in human development. In Goslin, D. A. (Ed.), *Handbook of socialization theory and research* (p. 53-65). Chicago: Rand McNally.

Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratio with a stratified ordinal response. *Biometrics, 52,* 1223-1234.

Maiden, R. J., Peterson, S. A., Caya, M., & Hayslip, B. (2003). Personality changes in the old-old: A longitudinal study. *Journal of Adult Development, 10(1),* 31-39.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Mantel, N. (1963). Chi-square tests with one degree of freedom. Extension of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690-700.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17,* 297-334.

Mischel, W. (1969). Continuity and change in personality. *American Psychologist, 24,* 1012-1018.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology, 55,* 1-22.

Mroczek, D. K., Hurt, S. W., & Berman, W. H. (1999). Conceptual and methodological issues in the assessment of personality disorders in older adults. In Rosowsky, E., Abrams, R. C., & Zweig, R. A. (Eds.), *Personality Disorders in Older Adults. 1st ed.* (p. 135-150). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Oltmanns, T. F. & Balsis, S. *Assessment of personality disorders in older adults.* In: Lichtenberg, P. A. (ed.). Handbook of Assessment in Clinical Gerontology. Academic/Elsevier; 2010. p. 101-122.

Penfield, R. D. (2005). DIFAS: Differential item functioning analysis system. *Applied Psychological Measurement, 29,* 150-151.

Penfield, R. D. (2007a). DIFAS 4.0. Differential item functioning analysis system: user's manual. Unpublished manuscript.

Penfield, R. D. (2007b). An approach for categorizing DIF in polytomous items. *Applied Measurement in Education, 20(3),* 335-355.

Penfield, R. D., & Algina, J. (2003). Applying the Liu-Agresti estimator of the cumulative common odds ratio to DIF detection in polytomous items. *Journal of Educational Measurement, 40,* 353-370.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43(4),* 295-312.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 3-27.

Savla, J., Davey, A., Costa, P. T. Jr., Whitfield, K. E. (2007). Replicating the NEO PI-R factor structure in African-American older adults. *Personality and Individual Differences, 43,* 1279-1288.

Segal, D. L., Coolidge, F. L., & Rosowsky, E. (2006). *Personality Disorders and Older Adults. Diagnosis, Assessment, and Treatment.* New Jersey: John Wiley & Sons, Inc.

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: Developmental considerations for the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. *Development and Psychopathology, 21,* 687-713.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T. Jr. (2005). Hierarchical linear modeling analyses of NEO PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20(3),* 493-506.

Thissen, D. (2001). *IRTLRDIF v.2.0b: software for the computation of the statistics involved in item response theory likelihood ratio tests for differential item functioning.* Chapel Hill: University of North Carolina, L. L. Thurstone Psychometric Laboratory.

Van den Broek, A., & de Haan, J. (2000). *Cultuur tussen competitie en competentie. Contouren van het cultuurbereik in 2030.* Amsterdam: Boekmanstudies.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4(2),* 223-233.

Zweig, R. A. (2008). Personality disorder in older adults: Assessment challenges and strategies. *Professional Psychology: Research and Practice, 39(3),* 298-305.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30,* 233-251.

Chapter 3
Validation of the FFM PD count technique for screening personality pathology in later middle-aged and older adults

Joke Van den Broeck, Gina Rossi, Barbara De Clercq, Eva Dierckx, & Leen Bastiaansen

Abstract

Research on the applicability of the Five Factor Model (FFM) to capture personality pathology coincided with the development of a FFM Personality Disorder (PD) count technique, which has been validated in adolescent, young, and middle-aged samples. This study extends the literature by validating this technique in an older sample. Five alternative FFM PD counts based upon the Revised NEO Personality Inventory (NEO PI-R) are computed and evaluated in terms of both convergent and divergent validity with the Assessment of DSM-IV Personality Disorder Questionnaire (ADP-IV). For the best working count for each PD normative data are presented, from which cut-off scores are derived. The validity of these cut-offs and their usefulness as a screening tool is tested against both a categorical (i.e., the Diagnostic and Statistical Manual of Mental Disorders – Fourth edition – Text Revision; DSM-IV-TR), and a dimensional (i.e., the Dimensional Assessment of Personality Pathology; DAPP) measure of personality pathology. All but the Antisocial and Obsessive-Compulsive counts exhibited adequate convergent and divergent validity, supporting the use of this method in older adults. Using the ADP-IV and the Dimensional Assessment of Personality Pathology – Screening Form (DAPP-SF) as validation criteria, results corroborate the use of the FFM PD count technique to screen for PDs in older adults, in particular for the Paranoid, Borderline, Histrionic, Avoidant and Dependent PDs. Given the age-neutrality of the NEO PI-R and the considerable lack of valid personality assessment tools, current findings appear to be promising for the assessment of pathology in older adults.

Published in Aging & Mental Health, 2012

3.1. Introduction

Based on an extensive body of research, it is nowadays commonly assumed that personality disorders (PDs) can be understood as maladaptive variants of general personality traits (Miller, Reynolds, & Pilkonis, 2004). In addition, growing consensus exists that normal and abnormal personality variation can be described within a single, unified structural framework (Markon, Krueger, & Watson, 2005). Among dimensional trait models of personality, the Five Factor Model (FFM) of personality has been most frequently applied to study the relations between PD constructs and general personality functioning (Miller, Bagby, Pilkonis, Reynolds, & Lynam, 2005). The FFM includes five broad domains of personality (Neuroticism, Extraversion, Openness to experiences, Agreeableness, and Conscientiousness) that are recoverable across age groups throughout the lifespan (e.g., children – Markey, Markey, Tinsley, & Ericksen, 2002; adolescents – Parker & Stumpf, 1998; adults – Costa & McCrae, 1990; and older adults – Weiss et al., 2005). The FFM is most commonly assessed using self-reports or other reports on the Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992) (Miller et al., 2010).

As one of the few age-neutral intended personality measures, the later life context was closely considered during the NEO PI-R's development (Costa et al., 1986; McCrae & Costa, 1987; see Tackett, Balsis, Oltmanns, & Krueger, 2009), and its age neutrality has recently been empirically validated by Van den Broeck and colleagues (2012). Moreover, separate norms for adults aged 50 or more are available for the Dutch/Flemish adaptation of the NEO PI-R (Hoekstra, Ormel, & De Fruyt, 2007). Despite its primary aim to capture general trait variance, the FFM has proven quite successful in representing the ten DSM-IV PD constructs (e.g., Saulsman & Page, 2004).

Increasing research on the applicability of the FFM to capture personality pathology coincided with the development of an easy-to-use FFM PD count technique (Miller et al., 2005), which was a simplification of a more complex prototype-matching method developed earlier (Miller, Reynolds, & Pilkonis, 2004). Based on expert-generated FFM prototypes (Lynam and Widiger, 2001), prototypically low and high facets for each PD are identified and summed. Accordingly, Lynam and Widiger's (2001) prototypes describe the 10 DSM-IV PDs based upon 7 (Schizotypal) to 17 (Antisocial) out of the 30 FFM facets (Miller et al., 2010). For example, the FFM PD count for the Borderline PD would involve a summation of the facets: Anxiety (N1), Angry Hostility (N2), Depression (N3), Impulsiveness (N5), Vulnerability (N6), Openness to feelings (O3), Openness to actions (O4), and the reverse scored facets of Compliance (A4) and Openness to values (O6).

It is important to note that normative data are crucial to determine the relative level of elevation of a specific score, because individual counts are difficult to interpret and of limited clinical use without a standard to compare with (Miller et al., 2008). Miller and colleagues (2008) presented data from normative samples from the United States, France, and Belgium-Netherlands that can be used as norms for the FFM PD counts based on Lynam and Widiger's prototypes (2001) in the respective countries.

Although there is a growing consensus that the FFM can be used to represent PD pathology, most of this research is based on younger adult samples (Tackett et al., 2009). In particular, the validation of the FFM PD count technique, relies – to our knowledge – almost exclusively on adolescent (Decuyper, De Clercq, De Bolle, & De Fruyt, 2009) and younger adult samples with mean ages ranging from 25.8 (Miller et al., 2008) to 41.4 (Miller et al., 2005). Only recently, a study of Lawton and colleagues (Lawton, Shields, & Oltmanns, 2011) extended this literature by validating the FFM PD count technique in a large community-

dwelling sample of later middle-aged adults with ages ranging between 55 and 64 years. Their results indicated adequate convergent validity for the Schizoid, Borderline, Histrionic, Narcissistic, and Avoidant PDs, with adequate discriminant validity for the latter four. The validation of the FFM PD count technique has – to our knowledge – not been investigated in an older adult sample.

All previous FFM count studies made use of the expert-generated prototypes of Lynam and Widiger (2001), although other FFM prototypes are available and can easily be converted into alternative FFM counts. In a recent study on the validity of FFM PD counts, Bastiaansen, Rossi, and De Fruyt (*in press*) proposed four alternative FFM counts, based on the theoretically derived prototypes of Trull and Widiger (1997), and Widiger, Trull, Clarkin, Sanderson, and Costa (2002), the clinician-generated prototypes of Samuel and Widiger (2004), and those based on the meta-analytic FFM profiles of Samuel and Widiger (2008). The results of their study, conducted in a clinical adult sample, supported the use of alternative FFM prototypes. The Paranoid, Schizoid, Histrionic, Narcissistic, and Obsessive-Compulsive FFM counts performed better when they were based on other prototypes than those derived by Lynam and Widiger (2001).

Current Study

The current study aims to address the applicability of the FFM PD counts for personality pathology screening purposes in older adults, since this has not been empirically investigated thus far. Given the significant lack of valid personality measurement tools for older adults, and consequently the poor understanding of PDs in later life (Balsis, Gleason, Woods, & Oltmanns, 2007), the answer to this question may have important clinical and empirical value. If current results underscore the use of a FFM PD count technique for screening personality pathology in older adults, they may contribute the ongoing study on relevant techniques for describing PDs throughout the lifespan. Also,

they may yield an important screening tool for geriatric health care providers, enabling the elaboration of appropriate diagnostic assessment procedures. Five alternative FFM PD counts will be evaluated in terms of both convergent and divergent validity, and the best working count for each PD will be selected. Normative data will be presented for these counts, from which cut-off scores can be derived in order to screen for PDs. The validity of these benchmarks will be tested not only against a categorical (i.e., the DSM-IV; APA, 2000) index, but also against a dimensional measure that specifically attempts to include pathological features of personality (i.e., the Dimensional Assessment of Personality Pathology [DAPP]; (Livesley et al., 1992).

3.2. Method

Participants and Procedure

Participants were 272 Dutch-speaking community-dwelling later middle-aged and older adults recruited by undergraduate psychology students of the Vrije Universiteit Brussel and Lessius Antwerp. Students were requested to recruit at least one person aged 50 or older. Participants were asked to fill out three self-report questionnaires (see Measures) and an additional demographic information form, all administered by means of paper-and-pencil. Participants' age ranged between 50 and 88 years (M = 68.18; SD = 6.79), with 45.2% male participants. Twenty-one percent of the participants reported that they had previously received psychological treatment or counseling in an ambulant setting, and 3.3% reported that they had previously been hospitalized for psychological or psychiatric problems. The study was approved by the ethical board of the Vrije Universiteit Brussel's and all participants volunteered their participation and provided a written informed consent.

Measures

NEO PI-R. The Dutch authorized translation of the NEO PI-R (Costa & McCrae, 1992; Hoekstra, Ormel, & de Fruyt, 2007) was used in this study to assess self-reported FFM personality traits. The NEO PI-R consists of 240 items to be rated on a 5-point Likert scale, ranging from *strongly disagree* to *strongly agree*. Items are hierarchically organized into five broad domains as conceptualized by the FFM, with each of them comprising six facets including eight items. A comprehensive body of studies supports its validity across cultures and languages (McCrae & Terracciano, 2005). In the current sample, internal consistency reliabilities for the five domains ranged from .83 (Extraversion) to .90 (Neuroticism), with a median value of .85. The internal consistency reliabilities of the facet scales ranged from .42 (C3, Dutifulness) to .82 (N1, Anxiety) (median coefficient alpha = .70). These values are consistent with normative data reported in the NEO PI-R manuals (Costa & McCrae, 1992; Hoekstra et al., 2007).

ADP-IV. The Assessment of DSM-IV Personality Disorders Questionnaire (ADP-IV), a 94-item Dutch self-report inventory developed by Schotte et al. (2004), measures personality pathology as conceptualized by the DSM-IV criteria for the ten recognized personality disorders. Each item measures both "trait" as well as "distress/impairment" characteristics of a DSM-IV criterion. Both trait and distress scales are necessary to assign a categorical PD diagnosis, but only the trait scales were used here. In the current sample, internal consistency reliabilities ranged from .54 (Antisocial) to .81 (Avoidant) with a median value of .75, in line with the values reported in previous studies (e.g., Schotte, de Doncker, Vankerckhoven, Vertommen, & Cosyns, 1998).

DAPP-SF. The Dutch authorized translation of the Dimensional Assessment of Personality Pathology – Short Form (DAPP-SF; van Kampen & de Beurs, 2009), a screening version of the DAPP-Basic Questionnaire (DAPP-BQ; Livesley & Jackson, 2001), was used to measure personality pathology. The DAPP-SF comprises 136 items of

the original 290 items. The items are rated on a 5-point Likert scale, ranging from *very unlike me* to *very like me*. Like the DAPP-BQ, the DAPP-SF covers 18 personality disorder trait-based dimensions fitting into four broad higher order factors (Emotional Dysregulation, Dissocial Behavior, Inhibition, and Compulsivity). The other psychometric characteristics of the original DAPP-BQ are preserved in the shortened DAPP-SF as well (de Beurs, Rinne, van Kampen, Verheul, & Andrea, 2009). Cronbach alpha coefficients for the 18 maladaptive trait facets in the current sample ranged from .66 (Conduct Problems) to .88 (Insecure Attachment) with a median value of .80. These values are broadly consistent with normative data reported in the manual (van Kampen & de Beurs, 2009).

Normative Older Sample

In order to compute FFM PD count benchmarks in an independent normative sample, a second sample of NEO PI-R self-reports in older adults was collected. This sample consisted of 659 older men (47.3%) and women (52.7%) voluntarily recruited from leisure clubs and senior meetings. Participants were recruited by a "snowball" technique in which volunteers already participating invited their friends and family to join in. Participants' mean age was 66.45 (*SD*= 8.76), ranging from 50 to 92 years. This sample has been partly used in previous research (Van den Broeck, Rossi, Dierckx, & De Clercq, 2012).

FFM Counts

For each of the ten DSM-IV PDs, five different FFM counts were calculated, based on the prototypes defined by Lynam and Widiger (2001), Widiger et al. (2002), Trull and Widiger (1997), Samuel and Widiger (2004), and Samuel & Widiger (2008), respectively. Concerning Samuel and Widiger's (2008) meta-analysis, facets with a correlation larger than .20 were considered prototypical, in line with the authors' viewpoint (this procedure was also applied by Bastiaansen et al., *in press*).

Facets that are considered prototypically low in relation to a given PD were reverse scored in the direction of maladaptivity for each PD.

Statistical Analyses

Bivariate correlations between the different FFM PD counts and their corresponding ADP-IV scale were computed, to evaluate the convergent and divergent validity of each count. Each FFM count was assumed to have its highest significant correlation with its corresponding PD scale, in order to conclude adequate convergent validity. In addition, mean discriminant correlations were computed for each FFM count, and compared to the respective convergent correlation (after r-to-z transformations). Adequate divergent validity would be evidenced by at least a small effect size ($q \geq .10$; Cohen, 1988), with a higher q indicating better divergent validity. Following these steps, the best working count for each PD was selected, and these ten counts were then subjected to one-way analyses of variance (ANOVAs), in order to evaluate their screening capacity for PDs. According to a statistical deviance model, individuals scoring 1.5 standard deviation above average were considered to have "extreme" scores, possibly reflecting problematic levels of the PD in question (Miller et al., 2008). Thus, T-scores were computed and used as benchmarks. Next, individuals with scores at or above a T65 cut-off were compared to those scoring below this cut-off on the ADP-IV and DAPP-SF using ANOVAs (Welch correction applied when required). We used Cohen's d as a measure of effect size (Cohen, 1988), with $d \geq .20$ indicating a small effect, $d \geq .50$ a medium effect, and $d \geq .80$ a large effect. Concerning the between-group analyses with the DAPP-SF, only those facets for which a difference was expected were included. Therefore we relied upon the most recently hypothesized relationships between DAPP-BQ scales and specific DSM-IV personality disorders described in the DAPP-BQ manual (Livesley & Jackson, 2009, p. 65).

Table 1 *Convergent and mean divergent correlations between ADP-IV Axis II scales and five FFM PD count prototypes in an older adult sample*

	FFM PD count														
	Lynam & Widiger (2001)			Widiger et al. (2002)			Trull & Widiger (1997)			Samuel & Widiger (2004)			Samuel & Widiger (2008)		
ADP-IV	con	div	*q*	con	div	*q*	con	div	*q*	con	div	*q*	con	div	*q*
PAR	.55	.35	.21	.64	.33	**.32**	.46	.35	.11	.61	.38	.23	.56	.45	.11
SZ	.46	.20	.26	.50	.23	.27	.41	.05	**.36**	.49	.23	.27	.47	.32	.15
ST	.38	.32	.06	.39	.31	.08	.35	.25	.10	.21	.08	**.13**	.42	.41	.01
AS	.21	.18	.04	.23	.36	-.12	.15	.13	.03	.19	.14	.05	.23	.39	-.16
BDL	.68	.29	**.39**	.77	.39	.38	.61	.25	.36	.72	.37	.35	.74	.42	.32
HIS	.07	-.18	.25	.14	-.09	.22	-.02	-.20	.18	.23	-.01	**.24**	-.09	-.23	.14
NAR	.37	.19	.19	.40	.30	.11	.26	.04	.23	.35	.08	**.27**	.45	.34	.11
AV	.50	.22	.28	.58	.29	.29	.56	.26	**.30**	.53	.26	.27	.63	.33	**.30**
DEP	.43	.09	**.34**	.29	.01	.28	.25	-.04	.29	.53	.22	.32	.60	.31	.29
OC	.05	-.07	.12	.05	-.06	.11	.19	.05	.14	.05	-.07	.13	-.08	-.17	.09

Note. The correlations shown are r-to-z (Fisher) correlations. PAR = Paranoid, SZ = Schizoid, ST = Schizotypal, AS = Antisocial, BDL = Borderline, HIS = Histrionic, NAR = Narcissistic, AV = Avoidant, DEP = Dependent, OC = Obsessive-Compulsive. Con = convergent correlation, div = mean discriminant correlation, *q* = Cohen's *q* (Cohen, 1998). All correlations > .13 are significant at $p < .05$. The counts that had their highest (significant) correlation with their corresponding PD ($p < .001$) are underlined, for each of these counts, the highest Cohen's *q* value is bold-faced. The eight selected counts are grey-shaded.

3.3. Results

Convergent and Divergent Validity

Table 1 presents the convergent and mean discriminant Fisher z correlations across the five alternative FFM PD count prototypes. Each ADP-IV scale significantly correlated with at least one of the five alternative FFM PD count models. Convergent Pearson correlations ranged from .14 (Narcissistic count based on the prototypes by Widiger et al., 2002) to .77 (Borderline count based on the prototypes by Widiger et al., 2002), with a median *r* of .53. In order to examine the discriminant validity, mean divergent Fisher z correlations were computed for each FFM count, and subtracted from the respective Fisher z convergent

correlations. We relied on the highest *q* value for selecting the specific FFM count that showed the best discriminant validity. Results indicated that the FFM counts based on Lynam and Widiger's expert-generated prototypes (2001) worked best for the Borderline and Dependent counts, whereas the counts based on the prototypes proposed by Widiger and colleagues (2002) were most adequate for the Paranoid count. The Schizoid and Avoidant counts performed best when based on Trull and Widiger's prototypes (1997). The counts based on Samuel and Widiger's prototypes (2004) worked best for the Schizotypal, Histrionic, and Narcissistic counts. Finally, none of the Antisocial and Obsessive-compulsive counts showed sufficient discriminant validity (as evidenced by $q < .10$ and/or having one or more divergent correlations exceeding the convergent correlation), so these counts were excluded from further analyses. The eight selected counts are highlighted in Table 1, and their facet contents are listed in Appendix A. Based on the above mentioned criteria, the Avoidant count performed equally well based on the prototypes proposed by Trull and Widiger (1997) and those derived from Samuel and Widiger's meta-analysis (2008) (i. e., Cohen's *q* of .30 in both cases). For reasons of parsimony, we decided to retain the Trull and Widiger (1997) count for further analyses as both counts contain exactly the same facets, with six additional facets in the Samuel and Widiger (2008) count.

Table 2 *Descriptive statistics and benchmarks for normative data of Flemish older adults NEO-PI-R self-reports (N = 659)*

FFM PD count	mean	SD	T-score 50	T-score 65
PAR Widiger	46.60	12.51	47	61
SZ Trull	103.68	15.02	104	135
ST Samuel	79.18	11.93	79	103
BDL Lynam	125.40	23.70	125	163
HIS Samuel	112.50	14.39	113	146
NAR Samuel	127.84	21.63	128	166
AV Trull	168.74	29.45	169	220
DEP Lynam	115.8	18.83	116	151

Note. PAR = Paranoid, SZ = Schizoid, ST = Schizotypal, BDL = Borderline, HIS = Histrionic, NAR = Narcissistic, AV = Avoidant, DEP = Dependent; Lynam = Lynam & Widiger (2001); Widiger = Widiger et al. (2002a); Trull = Trull & Widiger (1997); Samuel = Samuel & Widiger (2004).

Validating the FFM PD Count Benchmarks

Table 2 presents means and standard deviations for the eight best working FFM PD counts from our Flemish older adult normative data set. Based on these data, raw counts were converted to T-scores and used as benchmarks. Next, between-group differences were computed, comparing the mean ADP-IV scores between individuals scoring at or above the T65 cut-off and those scoring below this benchmark, using one-way ANOVAs (see Table 3). Individuals scoring at or above the FFM PD count cut-off scored significantly higher on their corresponding PD scale for the Paranoid, the Borderline, and the Avoidant count. The differences between high and low scorers were not significant for the Schizoid, Schizotypal, Narcissistic, Histrionic and Dependent PDs.

Finally, the validity of the FFM PD count cut-off scores was evaluated against the pathological personality dimensions of the DAPP-SF. As with the ADP-IV, mean scores on the DAPP-SF facets of individuals

scoring at or above the FFM PD count cut-offs (displayed in Table 3) were compared to those scoring below these thresholds. The vast majority of the hypothesized group differences on DAPP-SF facets were supported, including those for the Paranoid, Borderline, Histrionic, Avoidant, and Dependent counts. For the remaining PD counts, only a minority of the hypothesized differences were confirmed, compromising the differential validity of these counts. Detailed information on these analyses can be found in Table 4.

Table 3 *Between-group differences for ADP-IV scores based upon FFM PD counts in an older adult sample*

Note. PAR = Paranoid, SZ = Schizoid, ST = Schizotypal, BDL = Borderline, HIS = Histrionic, NAR = Narcissistic, AV = Avoidant, DEP = Dependent; Lynam = Lynam & Widiger (2001); Widiger = Widiger et al. (2002a); Trull = Trull & Widiger (1997); Samuel = Samuel & Widiger (2004); [a] Welch F (asymptotically F distributed); * $p \leq .001$; † $p \leq .01$.

FFM PD count	< T65		≥ T65			
	Mean ADP-IV (*SD*)	n	Mean ADP-IV (*SD*)	n	F	Cohen's *d*
PAR Widiger	14.01 (5.05)	240	20.40 (7.57)	32	21.50[a]*	.99
SZ Trull	17.15 (6.01)	263	19.78 (5.07)	9	1.67	.60
ST Samuel	18.17 (6.61)	264	20.90 (8.78)	8	1.30	.35
BDL Lynam	19.36 (5.94)	259	33.93 (12.89)	13	16.46[a]†	1.45
HIS Samuel	15.65 (5.57)	270	24.86 (14.34)	2	.82[a]	.85
NAR Samuel	17.14 (5.62)	265	20.00 (8.66)	7	1.72	.39
AV Trull	15.41 (5.90)	254	22.94 (6.46)	18	27.01*	1.22
DEP Lynam	16.64 (5.93)	257	19.84 (9.32)	15	1.73[a]	.41

Table 4 *Between-group differences for DAPP-SF scores based upon FFM PD counts in an older adult sample*

		< T65		≥ T65			
FFM PD count	DAPP-SF	Mean DAPP-SF (*SD*)	*n*	Mean DAPP-SF (*SD*)	*n*	F	Cohen's *d*
PAR Widiger	Suspic	13.65 (4.35)	240	19.72 (5.83)	32	50.37***	1.18
SZ Trull	Lowaffil	12.29 (4.39)	263	14.89 (2.85)	9	3.10	.70
	Restrexp	21.99 (5.73)	263	27.44 (3.13)	9	24.60***	1.18
	Intimacy	19.50 (6.13)	263	20.22 (3.60)	9	.12	.14
ST Samuel	Cogndys	11.45 (4.14)	264	10.13 (3.52)	8	.80	.34
	Suspic	14.28 (4.90)	264	16.88 (5.96)	8	2.14	.48
	Restrexp	22.07 (5.74)	264	25.25 (5.42)	8	2.39	.60
	Lowaffil	12.24 (4.33)	264	16.75 (3.69)	8	8.48**	1.12
BDL Lynam	Insecatt	14.39 (5.79)	259	19.56 (7.05)	13	9.66**	.80
	Selfharm	7.27 (2.66)	259	9.97 (6.43)	13	2.27[a]	.55
	Anxious	13.42 (4.94)	259	22.13 (5.84)	13	37.78***	1.61
	Afflab	18.58 (5.75)	259	28.83 (6.77)	13	38.61***	1.63
	Cogndys	11.21 (3.96)	259	15.57 (5.35)	13	14.52***	.93
	Identity	11.16 (4.53)	259	17.25 (6.83)	13	10.12**[a]	1.05
HIS Samuel	Narciss	15.96 (5.55)	270	28.35 (3.74)	2	9.93**	2.62
	Submiss	18.02 (5.58)	270	26.00 (5.66)	2	4.07*	1.42
	Afflab	18.96 (6.07)	270	34.50 (2.12)	2	13.07***	3.42
NAR Samuel	Narciss	15.97 (5.59)	265	19.14 (6.62)	7	2.17	.52
AV Trull	Lowaffil	12.00 (4.13)	254	17.72 (4.34)	18	32.15***	1.35
	Anxious	13.37 (4.88)	254	20.44 (6.79)	18	33.40***	1.20
DEP Lynam	Submiss	17.78 (5.41)	257	23.27 (6.52)	15	14.28***	.92
	Insecatt	17.78 (5.41)	257	18.46 (5.77)	15	6.88**	.12

Note. PAR = Paranoid, SZ = Schizoid, ST = Schizotypal, BDL = Borderline, HIS = Histrionic, NAR = Narcissistic, AV = Avoidant, DEP = Dependent; Lynam = Lynam & Widiger (2001); Widiger = Widiger et al. (2002a); Trull = Trull & Widiger (1997); Samuel = Samuel & Widiger (2004); Suspic = Suspiciousness, Afflab = Affective Lability, Cogndys = Cognitive Dysregulation, Identity = Identity Problems, Narciss = Narcissism, Submiss = Submissiveness, Insecatt = Insecure Attachment, Opposite = Oppositionality, Lowaffil = Low Affiliation, Anxious = Anxiousness, Restrexp = Restricted Expression, Intimacy = Intimacy Problems, Selfharm= Self-harm.

3.4. Discussion

The current study empirically investigates the applicability of the FFM PD count technique for personality pathology screening purposes in

later middle-aged and older adults and extends existing evidence on the validity of the FFM PD count technique (e.g., Miller et al., 2005, 2008, 2010; Decuyper et al., 2009; Lawton et al., 2011; Bastiaansen et al., *in press*) in adolescent, young, and middle-aged adult populations towards older age groups. Moreover, following Bastiaansen et al. (*in press*), we evaluated and compared five alternative FFM PD counts, instead of exclusively focusing on one specific FFM count technique as was done in previous studies. Overall, all but the Antisocial and Obsessive-Compulsive counts loaded highest and significantly on their corresponding ADP-IV scale and displayed adequate mean divergent validity, hence indicating their usefulness as a screening tool an older adults.

The inability of each of the five Obsessive-Compulsive prototypes to screen for Obsessive-Compulsive PD is consistent with findings from previous research in adolescent (Decuyper et al., 2009), (young) adult (Miller et al., 2008), and middle-aged adult (Lawton et al., 2011) samples. A possible explanation for this observation relates to recent research literature concerning the inability of the NEO-PI-R to adequately capture maladaptivity at the high ends of the Conscientiousness dimension (Haigler & Widiger, 2001). The Obsessive-Compulsive PD, conceptually associated with high Conscientiousness, may therefore not be captured well by this operationalization of the FFM.

Also, neither of the five alternative prototypes worked well for the Antisocial PD. Parallel to the study of Lawton and colleagues (2011) in middle-aged community-dwelling adults, current results showed higher correlations between the Antisocial count and the Narcissistic PD scale of the ADP-IV instead of the Antisocial PD scale. They used the phenomenon of heterotypic continuity (i.e., the idea that the manifestation of personality can change while the underlying traits remain stable) as a possible explanation for this finding, suggesting that "antisocial behaviors shift towards a more narcissistic presentation as the person approaches later life" (Lawton et al., 2011, p. 289). Having

replicated similar findings in our older sample strengthens this idea suggesting that older adults may have fewer opportunities to manifest reckless or aggressive behavior (related to the Antisocial PD), and instead shift to more subtle actions as manipulation, characteristic for the Narcissistic PD. However, the high co-morbidity between the Antisocial and Narcissistic PDs should not be overlooked when interpreting these findings. The DSM-IV conceptualization of both PDs certainly shows considerable overlap, like a lack of empathy and exploiting behavior towards others, although this cannot fully explain current findings, given that the Narcissistic count did not display the same pattern of discriminative failure towards the Antisocial PD. An alternative explanation for the Antisocial FFM count's failure was proposed by Bastiaansen and colleagues (*in press*). They stated that, "in DSM-IV, the Antisocial PD is largely defined in terms of specific behavior, mostly criminal activities, which are less directly translated into more abstract personality traits" (Bastiaansen et al., *in press*, p. 17), and hence captured less well by the FFM.

A unique contribution of the current study is that it offers cut-off scores for the best working FFM PD counts derived from an independent older adult sample, which allows these counts to be used as a screener for personality pathology in older adults. The current FFM PD count cut-off scores were validated against both a categorical and dimensional measure of personality pathology. By taking the ADP-IV as a validation criterion, the validity of the cutoff- scores for the Paranoid, Borderline, and Avoidant counts could be confirmed. Results for the Schizoid, Schizotypal, Narcissistic, and Dependent counts were not significant, although the trends were in the expected direction.

Given the considerable amount of criticism regarding the conceptualization of personality pathology within a categorical framework, and the resulting shift towards a more dimensional personality trait model in the latest DSM-5 proposal (e.g., Krueger, Eaton, Derringer, Markon, Watson, & Skodol, 2011), we additionally

validated the FFM cut-off scores against the DAPP-SF. These analyses confirmed the validity of the cut-off scores for the Paranoid, Borderline, Histrionic, Avoidant, and Dependent counts. At this moment of writing, the DSM-5 proposal for diagnosing PDs withholds only six specific PD types, namely the Antisocial, Avoidant, Borderline, Narcissistic, Obsessive-Compulsive, and Schizotypal. These types will be assessed according to five criteria, including a constellation of pathological personality traits descriptive of the disorder (i.e., Criterion B). Evaluating the FFM PD counts against the proposed maladaptive DSM-5 traits (Krueger et al., 2011) will be an interesting topic for further research. For now, current results are interesting for clinicians familiar with the traditional DSM-IV terminology, because they offer an opportunity to assess the well-known PDs constructs in a dimensional way, warranting the continuity with the current classification format. Moreover, the use of FFM prototypes may be of particular value to examine personality pathology in later life, considering the age-neutrality of the NEO PI-R (e.g., Van den Broeck et al., 2012).

Limitations

Despite considerable strengths, a number of limitations of the current study should be considered. One limitation concerns the relative low occurrence of personality pathology in the current sample. Since data was gathered in the general population, few people showed deviant or extreme scores on the FFM PD counts. As such, we were unable to investigate the differential validity for some of the FFM PD counts. Clearly, it is necessary for clinical purposes to replicate current findings in a clinical older sample, and to provide normative data from clinical elderly samples as well. A related drawback refers to the recruitment procedure applied in this study, whereby students were asked to select at least one person aged 50 or older. As a result of this procedure, it is possible that helpful, cooperative older people are overrepresented in the current sample. Another limitation pertains to the relatively young lower-bound of age that is used in this study to demarcate later life.

Considerable differences may exist between "young-olds" and "old-olds" in terms of physical, psychological and social functioning (Segal, Coolidge, & Rosowsky, 2006). Yet we decided to take age 50 as lower bound to ensure a sufficiently high number of participants, necessary for the between-group comparisons carried out in this study. Besides, participants' mean age was 68.2, and 65.4% was aged 65 or older, so speaking of an older adult sample seemed fair enough to us, although we do acknowledge that follow up studies should better take this heterogeneity into account. A final limitation of the current study concerns the exclusive reliance on self-report data. The reported personality features could be susceptible to a variety of distortions such as limited insight or fake good tendencies. Lawton and colleagues (2011) demonstrated that self- and informant-reported FFM PD counts worked equally well, but also that informant reports added significant predictive utility for the Schizoid, Antisocial, Borderline, Histrionic, and Narcissistic PDs. Especially within an elderly population, where an increased risk of cognitive decline exists due to a degenerative disease or as a result of normal aging, further research exploring the possibilities of informant-ratings should be encouraged. Also, given the drawbacks related to self-report, replicating current findings using face-to-face semi-structured diagnostic interviews is an important avenue for further research.

Conclusion

In sum, eight FFM counts exhibited adequate convergent validity combined with adequate divergent validity, supporting the use of this method in older adults. Given the lack of valid measures for use in older adults, and the recently empirically validated age-neutrality of the NEO PI-R (Van den Broeck et al., 2012), these results may have important clinical value. They offer a valid screening tool for the assessment of pathological personality traits in older adults, enabling a valid and comprehensive description of an older patient's personality difficulties. The normative data and the derived 1.5 standard deviation cut-offs

enable practitioners and researchers to use FFM scores for PD screening purposes in older adults, in particular to screen for Paranoid, Borderline, Histrionic, Avoidant and Dependent PDs. After comparing the patients scores for each count with the norms provided in this paper, a more fine-tuned advice regarding the presence/absence of personality pathology can be formulated, resulting in treatment plans with more realistic therapeutic goals.

3.5. References

American Psychiatric Association (2000). DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, 4th edn., text revision. Washington DC: APA.

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. Psychology and aging, 22(1), 171-185.

Bastiaansen, L., Rossi, G., De Fruyt, F. Comparing five sets of Five-Factor Model personality disorder counts. European Journal of Personality, in press.

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

Costa, P. T., & McCrae, R. R. (1990). Personality Disorders and The Five-Factor Model of Personality. Journal of Personality Disorders, 4(4), 362-371. Guilford Publications Inc.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory: Professional manual. Odessa, FL: Psychological Assessment Resources.

De Beurs, E., Rinne, T., van Kampen, D., Verheul, R., & Andrea, H. (2009). Reliability and validity of the Dutch Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF), a shortened version of the DAPP-Basic Questionnaire. Journal of personality disorders, 23(3), 308-26.

Decuyper, M., De Clercq, B., De Bolle, M., & De Fruyt, F. (2009). Validation of FFM PD counts for screening personality pathology and psychopathy in adolescence. Journal of personality disorders, 23(6), 587-605.

Haigler, E. D., & Widiger, T. A. (2001). Experimental manipulation of NEO-PI-R items. Journal of personality assessment, 77(2), 339-58.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (2007). NEO PI-R en NEO-FFI. Handleiding. Amsterdam: Hogrefe.

Krueger, R. F., Eaton, N. R., Derringer, J., Markon, K. E., Watson, D., Skodol, A. E. (2011). Personality in DSM-5: Helping delineate personality disorder content and framing the metastructure. Journal of Personality Assessment, 93(4), 325-331.

Lawton, E. M., Shields, A. J., & Oltmanns, T. F. (2011). Five-factor model personality disorder prototypes in a community sample: Self- and informant-reports predicting interview-based DSM diagnoses. Personality disorders: Theory, research and treatment, 2(4), 279-292.

Livesley, W. J., & Jackson, D. N. (2009). DAPP-BQ: Dimensional Assessment of personality pathology - Basic questionnaire. Technical manual. Sigma Assessment systems. Inc.

Livesley, W. J., Jackson, D. N., & Schroeder, M. L. (2001). Manual for the dimensional assessment of personality pathology. Port Huron: Sigma Press.

Livesley, W. J., Jackson, D. N., & Schroeder, M. L. (1992). Factorial structure of traits delineating personality disorders in clinical and general population samples. Journal of Abnormal Psychology, 101(3), 432-440.

Lynam, D. R., & Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. Journal of Abnormal Psychology, 110, 401-412.

Markey, P. M., Markey, C. N., Tinsley, B. J., & Ericksen, A. J. (2002). A Preliminary Validation of Preadolescents' Self-Reports Using the Five-Factor Model of Personality. Journal of Research in Personality, 36(2), 173-181.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the Structure of Normal and Abnormal Personality: An Integrative Hierarchical Approach. Journal of Personality and Social Psychology, 36, 139-157.

Miller, J. D., Reynolds, S. K., & Pilkonis, P. A. (2004). The validity of the Five-Factor Model prototypes for personality disorders in two clinical samples. Psychological Assessment, 16, 310–322.

Miller, J. D., Bagby, R. M., Pilkonis, P. A., Reynolds, S. K., & Lynam, D. R. (2005). A simplified technique for scoring DSM-IV personality disorders with the Five-Factor Model. Assessment, 12(4), 404-415.

Miller, J. D., Lynam, D. R., Rolland, J.-P., De Fruyt, F., Reynolds, S. K., Pham-Scottez, A., Baker, S. R., et al. (2008). Scoring the DSM-IV personality disorders using the Five-Factor Model: development and validation of normative scores for North American, French, and Dutch-Flemish samples. Journal of personality disorders, 22(5), 433-50.

Miller, J. D., Maples, J., Few, L. R., Morse, J. Q., Yaggi, K. E., & Pilkonis, P. A. (2010). Using clinician-rated five-factor model data to score the DSM-IV personality disorders. Journal of personality assessment, 92(4), 296-305.

Parker, W. D., & Stumpf, H. (1998). A validation of the five-factor model of personality in academically talented youth across observers and instruments. Personality and Individual Differences, 25, 1005-1025.

Samuel, D. B., & Widiger, T. A. (2004). Clinicians' personality descriptions of prototypic personality disorders. Journal of personality disorders, 18(3), 286-308.

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: a facet level analysis. Clinical psychology review, 28(8), 1326-42.

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. Clinical Psychology Review, 23(8), 1055-1085.

Schotte, C. K. W., De Doncker, D. a M., Dmitruk, D., Van Mulders, I., D'Haenen, H., & Cosyns, P. (2004). The ADP-IV questionnaire: differential

validity and concordance with the semi-structured interview. Journal of personality disorders, 18(4), 405-19.

Schotte, C. K., de Doncker, D., Vankerckhoven, C., Vertommen, H., & Cosyns, P. (1998). Self-report assessment of the DSM-IV personality disorders. Measurement of trait and distress characteristics: the ADP-IV. Psychological medicine, 28(5), 1179-88.

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: developmental considerations for the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. Development and psychopathology, 21(3), 687-713.

Trull, T. J., & Widiger, T. A. (1997). SIFFM: Structured Interview for the Five-Factor Model of Personality, professional manual. Odessa, FL: Psychological Assessment Resources.

Van den Broeck, J., Rossi, G., Dierckx, E., & De Clercq, B. (2012). Age-neutrality of the NEO-PI-R: potential differential item functioning in older versus younger adults. Journal of Psychopathology and Behavioral Assessment, DOI: 10.1007/s10862-012-9287-4.

Van Kampen, D., & De Beurs, E. (2009). DAPP-BQ: Dimensionale assessment van persoonlijkheidspathologie inclusief screeningsversie. Amsterdam: Hogrefe.

Weiss, A., Costa, P. T., Karuza, J., Duberstein, P. R., Friedman, B., & McCrae, R. R. (2005). Cross-sectional age differences in personality among medicare patients aged 65 to 100. Psychology and aging, 20(1), 182-5.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa, P. T., Jr. (2002). A description of the DSM-IV personality disorders with the five-factor model of personality. In P. T. Costa, & T. A. Widiger (Eds.), Personality disorders and the five-factor model of personality (2nd ed., pp. 89–99). Washington, DC: American Psychological Association.

3.6. Appendix

PD	Source	FFM Facets
Paranoid	Widiger	N2, A1r, A2r, A4r
Schizoid	Trull	N2r, N4r, E1r, E2r, E4r, E6r, O3r
Schizotypal	Samuel	E1r, E2r, E5r, O1, O5
Borderline	Lynam	N1, N2, N3, N5, N6, O3, O4, C6r
Histrionic	Samuel	N1, N5, E2, E5, O1, O3, C6
Narcissistic	Samuel	N4r, E3, E4, E5, A1r, A2r, A3r, A4r, A5r, A6r
Avoidant	Trull	N1, N3, N4, N6, E2r, E3r, E4r, E5r, O4r, A5, C1r
Dependent	Lynam	N1, N4, N6, E3r, A1, A4, A5

Note: DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th edition; PD = personality disorder; FFM = Five-Factor Model of personality; Lynam = Lynam & Widiger (2001); Widiger = Widiger et al. (2002); Trull = Trull & Widiger (1997); Samuel = Samuel & Widiger (2004); N1 = Anxiousness; N2 = Angry Hostility; N3 = Depression; N4 = Self-consciousness; N5 = Impulsiveness; N6 = Vulnerability; E1 = Warmth; E2 = Gregariousness; E3 = Assertiveness; E4 = Activity: E5 = Excitement Seeking; E6 = Positive Emotions; O1 = Fantasy; O3 = Feelings; O4 = Actions; O5 = Ideas; A1 = Trust; A2 = Straightforwardness; A3 = Altruism; A4 = Compliance; A5 = Modesty; A6 = Tendermindedness; C1 = Competence; C6 = Deliberation; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; r = reversed.

Chapter 4
Age-neutrality of the trait facets proposed for personality disorders in DSM-5: A DIFAS analysis of the PID-5

Joke Van den Broeck, Leen Bastiaansen, Gina Rossi, Eva Dierckx, & Barbara De Clercq

Abstract

An age-neutral measurement system is one of the basic conditions for an accurate personality assessment across the lifespan, both longitudinally and cross-sectionally. In this study the age-neutrality of the *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012) was investigated. Potential Differential Item Functioning (DIF) was examined for the 25 trait facets in older versus younger adults. Overall, 33 items displayed large DIF, according to the adjusted Bonferroni corrected cutoffs (Mantel Chi-square, Liu-Agresti Cumulative Common Log-Odds Ratio [L-A LOR], and Cox's Noncentrality Parameter Estimator [Cox's B]). In a next step, the implications of the item level DIF across age groups was investigated on scale (i.e., facet) level. These Differential Test Functioning (DTF) analyses revealed large DTF for four of the 25 PID-5 facets (i.e., Withdrawal, Attention Seeking, Rigid Perfectionism and Unusual Beliefs). Current initial results show that most PID-5 traits are measured equally well across age, however, further research is needed to further refine this instrument and make it entirely age-neutral.

Submitted for publication in Journal of Psychopathology and Behavioral Assessment

4.1. Introduction

The transition from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994) towards the DSM-5 is well underway. The DSM-5 Personality and Personality Disorders Work Group (http://www.dsm5.org/MeetUs/Pages/PersonalityDisorders.aspx) proposed a hybrid dimensional-categorical model for personality and personality disorder assessment and diagnosis in which six specific personality disorder types are defined by two fundamental criteria, being impairments in personality functioning and the presence of pathological personality traits. Regarding the latter, a multidimensional maladaptive personality trait system has been developed, in order to represent individual differences in personality disorder expression (Krueger, Derringer, Markon, Watson, & Skodol, 2012; Wright et al., 2012). In this model, 25 primary traits are organized by 5 higher-order dimensions: Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism. The *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012) was developed to operationalize these DSM-5 traits. This inventory is publicly available for research purposes, with the goal of encouraging additional refinement and development prior to the finalization of the DSM-5 (Krueger et al., 2012).

In preparing the upcoming shift towards DSM-5, the Work Group's main focus has been on the transition from a categorical to a dimensional classification system. Concerns about the categorical conceptualization of personality disorders in the current DSM-IV and the rationale to switch to a dimensional approach are extensively documented elsewhere (e. g., Widiger & Trull, 2007). Unfortunately, this dimensional focus detracted attention from another important issue, namely the suitability of the criteria for measuring personality in later life (Tackett et al., 2009; Oltmanns & Balsis, 2011). This is regrettable, since "even if a dimensional shift is made, there will be continued psychometric and conceptual problems if the criteria do not closely consider the presentation of personality in later life" (Tackett et al.,

2009, p. 14). It is commonly assumed that the current DSM-IV criteria for personality disorders are not adequately attuned to the living situations and experiences of older people (Agronin & Maletta, 2000; Segal et al., 2000). Based on item response theory analyses in a large, cross-sectional study of 37.000 participants, Balsis and colleagues (2007) concluded that 29% of the DSM-IV Axis II criteria lack face validity in older age groups, hence possibly leading to over- or underdiagnosis of personality pathology in old age populations. Unfortunately, the presentation of later life was not explicitly considered in the construction of the PID-5 either. Yet an age-neutral measurement system is one of the basic conditions for an accurate personality assessment across the lifespan, both longitudinally and cross-sectionally.

Since the PID-5's age-neutrality has, to our knowledge, not been empirically investigated thus far, we set out to detect possible Differential Item Functioning (DIF) in older versus younger adults. An item is said to exhibit DIF when younger and older adults with a similar position on the underlying trait of interest do not have the same probability of endorsing that item (Edwards & Edelen, 2009). If DIF occurs, the assumption of measurement invariance is violated, leading to possibly flawed interpretations of observed between-group differences (Millsap, 2011). DIF analyses can be done in both a Classical Test Theory (CTT) and an Item Respons Theory (IRT) framework. The current analyses were conducted using an odds ratio approach (CTT), which is, in contrast to the IRT approach, not hampered by requirements of model fit and large sample sizes, and can be conducted using the easily accessible DIFAS program (Penfield, 2005).

4.2. Method

Participants and Procedure

A total of 464 participants were included in the current study, subdivided into a younger and an older sample. The younger sample

consisted of 288 undergraduate psychology students with ages ranging from 17 to 40 (*M*= 21.05, *SD*= 3.70, 27% male). Participants in the older sample were 176 Dutch-speaking community-dwelling adults recruited by undergraduate psychology students. Students were requested to recruit at least one person aged 60 or older. No other specifications or conditions were provided. As a return for participation, the students received course credits. Participants' age in the older sample ranged between 61 and 99 years (*M* = 72.73; *SD* = 6.09), with 40% male participants. All participants volunteered their participation and provided a written informed consent.

Measure

The Dutch authorized translation of the *Personality Inventory for DSM-5* (PID-5; Krueger et al., *in press*; De Clercq et al., 2011), a 220-item self-report questionnaire, was used to measure the proposed DSM-5 traits. Items are rated on a 4-point Likert format scale, ranging from very false or often false to often true or very true. The PID-5 has 25 primary lower-order scales or facets that load onto five higher-order personality pathology dimensions (Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism). Lower-order scale internal consistencies (Cronbach's alpha) ranged from .68 (Suspiciousness) to .95 (Eccentricity) in the current younger sample, and from .25 (Suspiciousness) to .91 (Eccentricity) in the older sample (Mdn= .82 in both samples). These values are largely in line with data reported in previous research (e.g., Wright et al., 2012; Hopwood et al., 2012), however the internal consistency of Suspiciousness is remarkably low in the current older sample (see Table 1).

Statistical Analyses

Between-group differences between the younger and older age group on the PID-5 facets were computed with t-tests for independent samples. Cohen's *d* was used as a measure of effect size (Cohen, 1988), with $r \geq$

.20 indicating a small effect, $r \geq .50$ a medium effect, and $r \geq .80$ a large effect.

Next, it was investigated whether there are PID-5 items that measure the personality trait of interest differently in younger versus older adults, after controlling for the overall level of underlying trait between both age groups. To detect possible DIF as a function of age, the Mantel Chi-square (Mazor et al., 1992), the Liu-Agresti Cumulative Common Log-Odds Ratio (L-A LOR; Liu & Agresti, 1996), and the Cox's Noncentrality Parameter Estimator (Cox's B; Camilli & Congdon, 1999) were used for polytomous items, using the DIFAS 5.0 software program (Penfield, 2007). The Mantel chi-square statistic is based on a group (2) x response option (4) contingency table, distributed as a chi-square with one degree of freedom. The higher the chi-square value, the higher the probability the item displays DIF. In line, the L-A LOR considers the log odds ratio of one group endorsing a response option relative to another. The Cox's B statistic is similar to the Mantel-Haenszel statistic but uses the hypergeometric mean. For both the L-A LOR and Cox's B statistic negative values indicate DIF against the reference group (younger adults), and positive values indicate DIF against the focal group (older adults). The following cut-off criteria are available to flag items with large DIF: |L-A LOR| > .64 (Penfield, 2007a), and |Cox's B| > .40 (Camilli & Congdon, 1999). The impact of DIF at scale level was examined by Differential Test Functioning (DTF) analyses. Penfield and Algina (2006) propose to define DIF effect variance as small for $v^2 < .07$, medium for $.07 \leq v^2 \leq .14$, and large for $v^2 > .14$.

DIF and DTF analyses were conducted at facet level, so the total facet scores were used as stratifying variables. The stratum size was set at 1, which is the default option in DIFAS 5.0. To reduce the Type I error, we used a Bonferroni corrected critical chi-square value, ranging from 8.28 to 10.83 (depending on the number of items per scale; for example for Anhedonia the Bonferroni correction was applied across all 8 items by three test statistics [.05/8*3], leading to a critical chi-square value of

9.55). Similar stringent Bonferroni cut-offs were applied to the L-A-LOR (ranging from .85 to .99), and Cox's B (ranging from .53 to .62) statistics to flag items with large DIF (also depending on the number of items per scale)[6].

4.3. Results

Descriptive Statistics

Overall, for 16 of the 25 facets, the mean scale scores of the older sample differed significantly ($p \leq .05$) from the mean scale scores of the younger sample (see Table 1). Large significant differences ($d > .80$) were found for Intimacy Avoidance and Risk Taking, and moderate significant differences ($d > .50$) for Hostility, Attention Seeking, Deceitfulness, Manipulativeness, Distractibility, and Irresponsibility. For all but the Intimacy Avoidance scale, mean scale scores for these facets were significantly higher for younger compared to older adults ($p < .001$).

[6] The exact cutoff values for each of the three DIF indicators for each analysis can be obtained on requested from the first author.

Table 1 *Descriptive statistics for the PID-5 primary traits for the younger (n= 288) and older (n= 176) sample.*

Facet	Cronbach Alpha		Raw Means (SD)		Effect size
	Young	Old	Young	Old	Cohen's *d*
Anhedonia	.83	.74	4.97 (3.60)	5.73 (3.88)	-.20*
Anxiousness	.88	.85	11.05 (5.59)	8.44 (5.61)	.47***
Depressivity	.89	.88	7.25 (6.01)	6.18 (6.45)	ns
Emotional Lability	.88	.85	9.24 (4.94)	7.17 (4.86)	.42***
Hostility	.81	.78	10.86 (4.88)	7.76 (5.13)	.62***
Perseveration	.78	.74	8.56 (4.22)	7.32 (4.49)	.28**
Rigid Perfectionism	.87	.85	9.93 (5.77)	11.02 (6.10)	ns
Separation Insecurity	.74	.72	9.02 (3.87)	8.23 (4.19)	.20*
Submissiveness	.76	.73	4.18 (2.24)	3.94 (2.76)	ns
Suspiciousness	.68	.24	7.02 (3.15)	7.39 (2.64)	ns
Withdrawal	.87	.87	4.66 (4.52)	6.34 (5.67)	-.33***
Attention Seeking	.86	.86	8.13 (4.39)	5.34 (4.83)	.60***
Callousness	.81	.77	5.58 (4.34)	5.60 (5.24)	ns
Deceitfulness	.84	.84	7.20 (4.61)	4.74 (4.97)	.51***
Grandiosity	.72	.83	2.83 (2.43)	2.57 (3.26)	ns
Manipulativeness	.83	.82	5.07 (3.13)	2.91 (3.12)	.69***
Intimacy Avoidance	.81	.68	2.15 (2.74)	5.79 (3.67)	-1.12***
Restricted Affectivity	.82	.70	5.69 (3.97)	6.01 (3.62)	ns
Distractibility	.89	.84	10.33 (5.42)	7.10 (5.29)	.60***
Eccentricity	.95	.91	9.05 (8.01)	6.33 (7.00)	.36***
Perceptual Disregulation	.79	.86	5.13 (4.51)	4.97 (5.53)	ns
Risk Taking	.89	.74	18.64 (6.87)	13.27 (5.83)	.84***
Unusual Beliefs	.80	.81	2.88 (3.42)	3.57 (4.03)	ns
Impulsivity	.80	.72	6.57 (3.26)	5.06 (3.44)	.45***
Irresponsibility	.71	.71	4.91 (3.05)	3.08 (3.13)	.59***

Note. * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Differential Item and Test Functioning

Twenty-five separate DIF and DTF analyses were conducted for each of the primary traits of the PID-5, whereby the respective total facet-score served as stratifying variable. Overall, DIF analyses revealed 30 items

showing significant DIF, divided across 15 facets. Table 2 and 3 show the paraphrased item content and statistical details of these items, and the facets they belong to. In order to evaluate the impact of these DIF items at scale (i.e., facet) level, additional DTF analyses were conducted. According to the Bonferroni corrected cut-off (> .35)[7], DTF analyses revealed large DTF for Withdrawal, Attention Seeking, Rigid Perfectionism, and Unusual Beliefs (with weighted v^2 values of .64, .65, .37, and .44, respectively). Withdrawal contained five items displaying significant DIF. Two of them displayed DIF against the older age group ("I keep to myself" and "I keep my distance from people"), indicating they were more readily endorsed by younger adults with the same level of underlying personality trait (i.e., Withdrawal). Three items showed DIF against younger adults ("I don't like spending time with others", "I'm not interested in making friends", and "I say as little as possible when dealing with people"). Attention Seeking contained four items displaying significant DIF, of which two displayed DIF against older ("I do things so that people just have to admire me", and "I crave attention"), and two against younger adults ("I love getting attention", and "I like standing out in a crowd"). Rigid Perfectionism contained one item displaying DIF against older ("I simply won't put up with things being out of their proper places"), and one item displaying DIF against

[7] In deriving their thresholds Penfield and Algina (2006) argued that a collective large level of DIF in a group of items exist if 25% or more of the items are categorized as having moderate or large magnitudes of DIF based on the ETS classification scheme (i.e. if 25% or more of the items have an absolute value of $\log(\alpha_{MH})$ greater than or equal to .43. They also suggest that MH and L-A LOR have similar meanings in terms of DIF magnitude. Because we wanted to reduce the Type I error, a Bonferroni correction was applied and adjusted L-A LOR cut-off values were used to flag items with large DIF (depending on the number of items per scale). In line, we made a similar adjustment for the DTF thresholds. For example: Penfield and Algina consider the variance of DIF effect large when weighted $v^2 > .14$, using an L-A LOR value of .43 as critical value. Since we adhere to stringent L-A LOR critical values (for example > .92 for Anhedonia) we adjusted this to $v^2 > .35$ (i.e., .14/.43*.92).

younger adults ("I focus too much on minor details"). Similarly, Unusual Beliefs contained one item displaying DIF against older ("I believe that some people can move things with their minds"), and two displaying DIF against younger adults ("Other people seem to think my behavior is weird", and "I see unusual connections between things"), indicating that the latter two are more readily endorsed by younger adults, matched on underlying personality trait.

Table 2 *Items meeting Bonferroni adjusted criteria for large DIF*

Facet	Item	Paraphrased context	Mantel χ^2	L-A LOR	Cox's B
Items displaying DIF against younger adults (older > younger)					
Anxiousness	110	I worry about almost everything	9.72	-.69	-.61
Anxiousness	174	I'm fearful about bad things that might happen	10.98	-.71	-.58
Emotional lability	165	I get emotional over every little thing	18.54	-1.01	-.79
Restricted affectivity	167	I never show emotions to others	11.9	-.72	-.60
Withdrawal	136	I don't like spending time with others	12.58	-.98	-.78
Withdrawal	146	I'm not interested in making friends	17.42	-1.16	-.66
Withdrawal	147	I say as little as possible when dealing with people	27.64	-1.20	-.95
Attention seeking	43	I do things so that people just have to admire me	24.61	-1.24	-.84
Attention seeking	191	I crave attention	35.56	-1.36	-1.03
Callousness	207	I don't see the point in feeling guilty about things	17.29	-1.23	-.94
Irresponsibility	201	I skip appointments if I'm not in the mood	10.63	-.84	-.62
Rigid perfectionism	196	I simply won't put up with things being out of their proper places	37.40	-1.35	-.85
Risk taking	195	I don't think about getting hurt when I'm doing things that might be dangerous	16.36	-.98	-.66
Unusual beliefs	143	People can move things with their minds	12.60	-.99	-.68

Note. L-A LOR= Liu-Agresti Common Log Odds Ratio. Cox's B= Cox's Noncentrality Parameter Estimator. ^{R} Reversed scored items. Facets with large DTF are given in bold.

Table 3 *Items meeting Bonferroni adjusted criteria for large DIF*

Facet	Item	Paraphrased context	Mantel χ^2	L-A LOR	Cox's B
Items displaying DIF against older adults (younger > older)					
Hostility	28	I snap at people when they irritate me	12.58	.78	.67
Hostility	32	I can be mean when I need to be	32.11	1.20	.77
Intimacy avoidance	89	I keep romance out of my life	11.46	1.02	.71
Withdrawal	20	I keep to myself.	14.40	.87	.63
Withdrawal	82	I keep my distance from people	39.40	1.61	1.13
Attention seeking	74	I love getting attention	12.74	.79	.60
Attention seeking	111	I like standing out in a crowd	18.93	.97	.76
Emotional lability	18	My emotions change for no good reason	24.86	1.04	.74
Callousness	200	I enjoy making people in control look stupid	12.12	.86	.64
Deceitfulness	134	I don't hesitate to cheat if it gets me ahead	10.37	.79	.64
Deceitfulness	214	Lying comes easily to me	10.60	.77	.61
Manipulativeness	125	Sweet-talking others helps me get what I want	8.91	.68	.62
Distractibility	132	I am easily distracted	15.18	.92	.78
Rigid perfectionism	49	I focus too much on minor details	61.62	1.74	1.03
Unusual beliefs	24	Others think my behavior is weird	15.06	.93	.63
Unusual beliefs	194	I see unusual connections between things	14.31	.96	.67

Note. L-A LOR= Liu-Agresti Common Log Odds Ratio. Cox's B= Cox's Noncentrality Parameter Estimator. ^{R} Reversed scored items. Facets with large DTF are given in bold.

4.4. Discussion

The primary aim of this study was to investigate the age-neutrality of the PID-5 facets by examining potential DIF for older versus younger adults. According to the stringent Bonferroni corrected cutoffs, analyses revealed a total of 33 items displaying significant DIF, divided across 15 facets. The impact hereof at scale level was relatively small. Large DTF was confirmed for four facets, namely Withdrawal, Attention Seeking, Rigid Perfectionism and Unusual Beliefs.

Withdrawal. Three items showed negative DIF, indicating they were more readily endorsed by older adults with similar levels of the latent personality trait. These items focus on (the absence of) close relationships (e.g., 'I'm not interested in making friends'). In this respect, it is possible that endorsing these items does not reflect personality pathology, but rather dealing with the death of loved ones, or, for example, isolation caused by physical illness (Van Alphen et al., 2006). Also, as people age, they tend to engage in selective social interaction, maintaining only the most rewarding contacts to satisfy their emotional needs (Carstensen, 1991). Notably, two other items from the Withdrawal facet were more readily endorsed by younger adults ("I keep to myself" and "I keep my distance from people"). Although at first sight substantially very similar to the items displaying negative DIF, these findings might reveal a difference in the interpretation of these items that varies with age. It is not unlikely that older adults are less inclined to endorse these items that probe social isolation and withdrawal, because the diminishing of a social network is more common in later life, due to loss experiences or physical deterioration. Both younger and older participants probably compared themselves to peers when filling out the PID-5 questionnaire, thus an item as "I keep to myself" may lead to less extremely high scores in an older population, because it is not so much seen as "deviant behavior", but rather as a normative phenomenon related to aging.

Attention Seeking. The item "I love getting attention" was more readily endorsed by younger adults, whereas the item "I crave attention" was more readily endorsed by older adults. Although the contents of these items show considerable overlap, it seems as though there really is a difference between "love attention" and "crave attention". Craving attention might be a more desirable statement for those who are lonely, and, supposing elderly people might in general be more lonely (e.g., Holmén & Furukawa, 2002), this might explain the differences found for this item.

Rigid Perfectionism. Younger adults scored generally higher on the item "I focus too much on minor details". This item is possibly more related to an occupational context, and therefore less relevant (and less readily endorsed) by retirees. The other DIF item in this facet ("I simply won't put up with things being out of their proper places") was more readily endorsed by older adults, suggesting that older adults might be in general more orderly and/or rigid, and less resistant against changes in their personal habitat than younger adults. Another explanation might lie in the cognitive decline and memory problems associated with advancing age. Older adults are possibly more prone to compensate with order and regularity, as a way to cope with their forgetfulness.

Unusual beliefs. Three items displayed DIF, of which two were more readily endorsed by younger adults ("Other people think my behavior is weird", and "I see unusual connections between things"). These items possibly reflect the tendency of young people to challenge traditional values and norms and their striving to be seen as unique, independent individuals. Also, young people may be more preoccupied and focused on what others think about them, leading them to think they act weird in the eyes of others.

Overall, the current initial results validate the comparison of mean facet scores across younger and older age groups for 21 of the 25 PID-5 traits. Given the lack of measurement invariance for Withdrawal,

Attention Seeking, Rigid Perfectionism, and Unusual Beliefs, caution is warranted in the interpretation of age differences based on these particular traits. Concerning the current between-group differences on mean facet scores, largest mean differences (as evidenced by moderate and large effect sizes) were found for Hostility, Attention Seeking, Deceitfulness, and Manipulativeness, all four belonging to the higher-order trait domain of Antagonism. Younger adults had significantly higher mean level scores for these traits. Accordingly, younger adults scored also significantly higher on Risk Taking, Distractibility, and Irresponsibility, three facets of the Disinhibition domain. When subjected to a hierarchical structure analysis, it is demonstrated that Antagonism and Disinhibition are both split-offs of a higher-level Externalizing factor (Wright et al., 2012). These findings support the notion that externalizing personality traits tend to remit with age. Due to physical changes associated with aging and consequently a reduced mobility and slower pace, elderly people are less likely to act impulsively or manifest risky, irresponsible behavior (Roberts et al., 2006). Furthermore, the PID-5 five-factor structure shows clear resemblance to the structure of normal personality as represented by the FFM, whereby Antagonism is the pathological variant of (low) Agreeableness, and Disinhibition the pathological variant of (low) Conscientiousness (Thomas, Yalch, Krueger, Wright, Markon, & Hopwood, *in press*). Parallels can thus be drawn between current findings and established research on age-related mean-level changes in general personality trait scores, describing an increase of Agreeableness and Conscientiousness traits throughout the lifespan (e.g., Terraciano et al., 2005; Roberts et al., 2006). Older adults, in turn, scored significantly higher on Intimacy Avoidance, a trait facet belonging to the higher-order domain Detachment, the pathological variant of (low) Extraversion. Again, these results corroborate the decline in Extraversion with advancing age.

Limitations

A few limitations should be considered. First of all, the cross-sectional design of the current study makes it difficult to distinguish real age effects from cohort effects, hampering the interpretation of the DIF results. It was not our intention however, to provide clear explanations for the current DIF results, but only to detect for possible DIF as a function of age in the recently proposed maladaptive personality traits for DSM-5. In line, the possibility cannot be ruled out that some of our findings are due to other factors beyond age that define differences between the current subsamples (e.g., gender or education level). However, these initial results clearly point out the need for further research. Future studies should explore the possible underlying causes of DIF, and their consequences for the assessment of personality pathology across the lifespan. Replicating the current findings within clinical samples is another important avenue for further research, since the PID-5 was primarily designed to identify personality pathology, a clinically-relevant phenomenon. Another limitation pertains to the Suspiciousness scale. Although no significant differences in mean scale scores were found between the younger and older age group for this trait, the internal consistency of this scale was clearly low in the current older sample and warrants further investigation. A last limitation considered here is the lower bound of age 60 as inclusion criteria for the older age group. The heterogeneity of this older age group should not be underestimated, since considerable differences might exist between, for example, 60-65 and 80-85 year-olds. With the current software used to investigate DIF we were limited to the comparison of two age groups, but investigating measurement invariance across different age groups covering the whole lifespan and using smaller age-ranges might reveal interesting findings about more nuanced age-related changes in personality traits.

Conclusions

Despite not having explicitly considered the later life context during its development, current initial results show that most PID-5 traits are

measured equally well across both a younger and an older age group. These results are promising in light of the growing awareness that an age-neutral measurement is crucial for a valid assessment of personality pathology throughout the lifespan. Additional research is certainly needed however to further refine this instrument and make it entirely age-neutral, since 33 items appeared to display large DIF, resulting in four scales exhibiting significant DTF. To this end, a set of alternative items could be written that works equally well for younger and older adults, regardless of their somewhat different living conditions. This set of items can then be tested for DIF across important demographic groups (e.g., gender, age, ethnic status, etc.), to finally reach a scale that contain no measurement artifacts (Oltmanns & Balsis, 2011). We do realize that creating such items is a challenging task, however we hope that the current exploratory analyses point out the need and inspire further researchers towards developing an age-neutral measurement system.

4.5. References

Agronin, M. E., & Maletta, G. (2000). Personality disorders in later life. Understanding and overcoming the gap in research. *American Journal of Geriatric Psychiatry, 8*, 4-18.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed). Washington, DC: Author.

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging, 22(1),* 171-185.

Camilli, G., & Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. *Journal of Educational and Behavioral Statistics, 24,* 323-341.

Carstensen, L. L. (1991). Selectivity theory: Social activity in life-span context. *Annual Review of Gerontology and Geriatrics, 11*, 195-217.

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

De Clercq, B., De Fruyt, F., Mervielde, I., Krueger, R. F., & Markon, K. (2011). Dutch
translation of the Personality Inventory for DSM-5 (PID5).

Edwards, M. C., & Edelen, M. O. (2009). Special topics in item response theory. In Millsap, R. E., & Maydeu-Olivares, A. (Eds.). *The SAGE handbook of quantitative methods in psychology.* (p. 178-185). London: SAGE Publication Ltd.

Holmén, K., & Furukawa, H. (2002). Loneliness, health and social network among elderly people – a follow-up study. *Archives of Gerontology and Geriatrics, 35,* 261-274.

Hopwood, C. J., Thomas, K. M., Markon, K. E., Wright, A. G. C., & Krueger, R. F. (2012). DSM-5 personality traits and DSM-IV personality disorders. *Journal of Abnormal Psychology, 121(2),* 424-432.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42,* 1879-1890.

Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratio with a stratified ordinal response. *Biometrics, 52,* 1223-1234.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *Journal of Personality and Social Psychology, 88(1),* 139-157.

Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement, 52,* 443-451.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance.* New York: Routledge.

Mroczek, D. K., Hurt, S. W., & Berman, W. H. (1999). Conceptual and methodological issues in the assessment of personality disorders in older adults. In Rosowsky, E., Abrams, R. C., & Zweig, R. A. (Eds.), *Personality*

Disorders in Older Adults. 1st ed. (p. 135-150). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Penfield, R. D. (2005). DIFAS: Differential item functioning analysis system. *Applied Psychological Measurement, 29,* 150-151.

Penfield, R. D. (2007). DIFAS 4.0. Differential item functioning analysis system: user's manual. Unpublished manuscript.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43(4),* 295-312.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 3-27.

Segal, D. L., Coolidge, F. L., & Rosowsky, E. (2000). Personality disorders. In Krauss, S. (ed). *Psychopathology in Later Adulthood.* (p. 89-115). John Wiley & Sons Inc.: New York.

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: Developmental considerations for the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. *Development and Psychopathology, 21,* 687-713.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T. Jr. (2005). Hierarchical linear modeling analyses of NEO PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20(3),* 493-506.

Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., & Derksen, J. J. L. (2006). The relevance of a geriatric sub-classification of personality disorders in the DSM-V. *International Journal of Geriatric Psychiatry, 21,* 205-209.

Van den Broeck, J., Rossi, G., & Dierckx, E. (2010). Diagnostiek van persoonlijkheid en persoonlijkheidspathologie bij ouderen. [Assessment of personality and personality pathology in older adults.] *Tijdschrift voor Gerontologie en Geriatrie, 41(2)*, 68-78.

Venables, P., & Bailes, K. (1994). The structure of schizotypy, its relation to subdiagnoses of schizophrenia and to sex and age. Britisch Journal of Clinical psychology, 33, 277-294.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, *62*, 71-83.

Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology,* doi: 10.1037/a0027669.

Zweig, R. A. (2008). Personality disorder in older adults: Assessment challenges and strategies. *Professional Psychology: Research and Practice, 39(3),* 298-305.

Chapter 5
Hierarchical structure of maladaptive personality traits in older adults: Joint factor analysis of the PID-5 and the DAPP-BQ

Joke Van den Broeck, Leen Bastiaansen, Gina Rossi, Eva Dierckx, Barbara De Clercq, & Joeri Hofmans

Abstract

The DSM-5 proposal for the diagnosis of a personality disorder is based on two fundamental criteria, being impairments in personality functioning (criterion A) and the presence of pathological personality traits (criterion B). In the maladaptive trait model that has been developed to operationalize criterion B, 25 pathological traits are organized by five higher-order dimensions. In the current study, we focused on the convergence of the proposed DSM-5 model (as measured by the PID-5) with the Dimensional Assessment of Personality Pathology (DAPP) model (as measured by the DAPP-BQ) in older people. A joint hierarchical factor analysis showed clear convergence between four PID-5 dimensions (*Negative Affect*, *Detachment*, *Antagonism*, *Disinhibition*) and conceptually similar DAPP-BQ components. Moreover, the PID-5 and DAPP-BQ showed meaningful associations on different levels of their joint hierarchical factor structure as well. Methodological and theoretical implications for the conceptualization of personality pathology are discussed.

Submitted for publication in Journal of Personality Disorders

5.1. Introduction

The transition from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994) to the DSM-5 is currently being intensively prepared. According to the latest proposal, the diagnosis of a personality disorder is based on five criteria (criterion A through E) with the two essential features being impairments in personality functioning (criterion A) and the presence of pathological personality traits (criterion B). Regarding the latter, a multidimensional maladaptive personality trait system has been proposed (Krueger et al., 2012). In this model, 25 primary traits are organized by 5 higher-order dimensions: *Negative Affect*, *Detachment*, *Antagonism*, *Disinhibition*, and *Psychoticism*. While constructing this trait model and its associated assessment instrument, the *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012), the DSM-5 Personality and Personality Disorders Work Group relied on existing models of maladaptive personality traits, such as Harkness's Personality Psychopathology Five model (PSY-5; Harkness & McNulty, 1994), and the Dimensional Assessment of Personality Pathology model (DAPP; Livesley et al., 1992; Krueger et al., 2011).

Recently, the hierarchical structure of the DSM-5 personality trait model has been examined by applying Goldberg's (2006) "bass-ackward" analytic strategy on PID-5 data (Wright et al., 2012). At the fifth and final level of their analysis, the PID-5 five-factor structure established by Krueger et al. (2012) was replicated. At previous levels of the hierarchy, the unfolding of the 25 proposed traits revealed structures that closely connected with common personality pathology models. At the second level, an Internalizing component (mainly marked by Depressivity, Anxiousness, and Withdrawal), and an Externalizing component (Manipulativeness, Risk Taking, and Attention Seeking) emerged from a general Personality Pathology factor. The Internalizing component then split into Detachment (Withdrawal, Anhedonia, and Restricted

Affectivity) and Negative Affect (Emotional Lability, Anxiousness, and Perseveration), after which the Externalizing component split into Antagonism (Manipulativeness, Grandiosity, and Callousness) and Disinhibition (Impulsivity, Risk Taking, and Distractibility). At the final level, a Psychoticism component (high loadings of Eccentricity, Perceptual Dysregulation, and Unusual Beliefs) emerged, which had no pronounced roots in any of the fourth level's components.

In the present study, we set out to investigate the convergent validity of the PID-5's hierarchical structure by means of a joint hierarchical factor analysis with the DAPP-BQ. Recently, Kushner and colleagues (2011) delineated the hierarchical structure of the DAPP-BQ. At their sixth and lowest level, five factors showed conceptual resemblance to PID-5 higher order dimensions: Emotional Dysregulation (*Negative Affect*), Inhibitedness (*Detachment*), Compulsivity (the opposite of *Disinhibition*), Dissocial Behavior/Disagreeable (*Antagonism*), Dissocial Behavior/ Externalizing (*Disinhibition*), and Need for Approval. This last component, marked by high loadings of Insecure Attachment, Submissiveness, and Narcissism, has no clear counterpart in the PID-5 higher-order domains, but conceptually connects to some lower-order scales of *Negative Affect* (Separation Insecurity and Submissiveness), and *Antagonism* (Attention Seeking and Grandiosity). Although the hierarchical structures of the PID-5 and DAPP-BQ show considerable conceptual overlap, especially from level one through four, this has - to our knowledge - not yet been empirically tested. This study will do so by examining the joint hierarchical structure of the 25 proposed DSM-5 personality traits and the 18 DAPP dimensions. Because previous studies on the hierarchical structure of the DAPP-BQ (Kushner et al., 2011) and the PID-5 (Wright et al., 2012) focused on young adults and students, we extend this literature by focusing on an older adult sample. By doing so, we aim to additionally contribute to the (sparse) research literature on the conceptualization of personality pathology in later life (e.g., Oltmanns & Balsis, 2011). No a priori predictions were made about the exact unfolding of the joint PID-5/ DAPP-BQ structure (i.e.,

at which level each higher-order component would appear or split). However, we did expect the conceptually related PID-5 and DAPP-BQ traits to dovetail together in a formation parallel to their original unfolding. For example, at the fourth level of the hierarchy, we anticipated the PID-5 and DAPP-BQ scales to coincide into the established "Big Four" dimensions (Widiger & Simonsen, 2005) as follows: PID-5 Negative Affect with DAPP-BQ Emotional Dysregulation, PID-5 Detachment with DAPP-BQ Inhibitedness, PID-5 Antagonism with DAPP-BQ Dissocial Behavior, and PID-5 Disinhibition with (reversed) DAPP-BQ Compulsivity. On the other hand, we expected unique traits (e.g., PID-5's Psychoticism-related traits) to show up as a separate component in the unfolding procedure.

5.2. Method

Participants and Procedure

A total of 173 Dutch-speaking community-dwelling adults, recruited by undergraduate psychology students, participated. Ages ranged between 61 and 99 years (M = 72.72; SD = 6.08), with 39.3% males. All participants provided a written informed consent.

Measures

PID-5. The Dutch authorized version of the *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012; De Clercq et al., 2011) was used to measure the DSM-5 traits. The PID-5 has 25 primary lower-order scales or facets that load onto five higher-order personality pathology dimensions (Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism). In the current sample, lower-order scale internal consistencies ranged from .25 (Suspiciousness) to .91 (Eccentricity) (*Mdn*= .82).

Dimensional Assessment of Personality Pathology – Basic Questionnaire (DAPP-BQ; Livesley & Jackson, 2009). The Dutch translation of the DAPP-BQ (van Kampen & de Beurs, 2009) was used to measure personality pathology. It covers 18 personality disorder trait-based dimensions, which fit into four higher order factors (Emotional Dysregulation, Dissocial Behavior, Inhibition, and Compulsivity). Cronbach's alpha coefficients ranged from .68 (Restricted Expression) to .95 (Selfharm) with a median value of .87.

Statistical Analyses

To examine the joint hierarchical structure of the PID-5 and the DAPP-BQ, the 25 primary DSM-5 traits and the 18 lower-order dimensions of the DAPP-BQ were subjected to a series of varimax rotated PCAs with an increasing number of factors. To decide on the maximal number of factors, we relied on parallel analysis, prior theory and interpretability. Following Goldberg's (2006) "bass-ackward" method, we computed regression-based factor scores on each level of the hierarchy, and these factor scores were subsequently correlated to compute path coefficients between the different hierarchical levels.

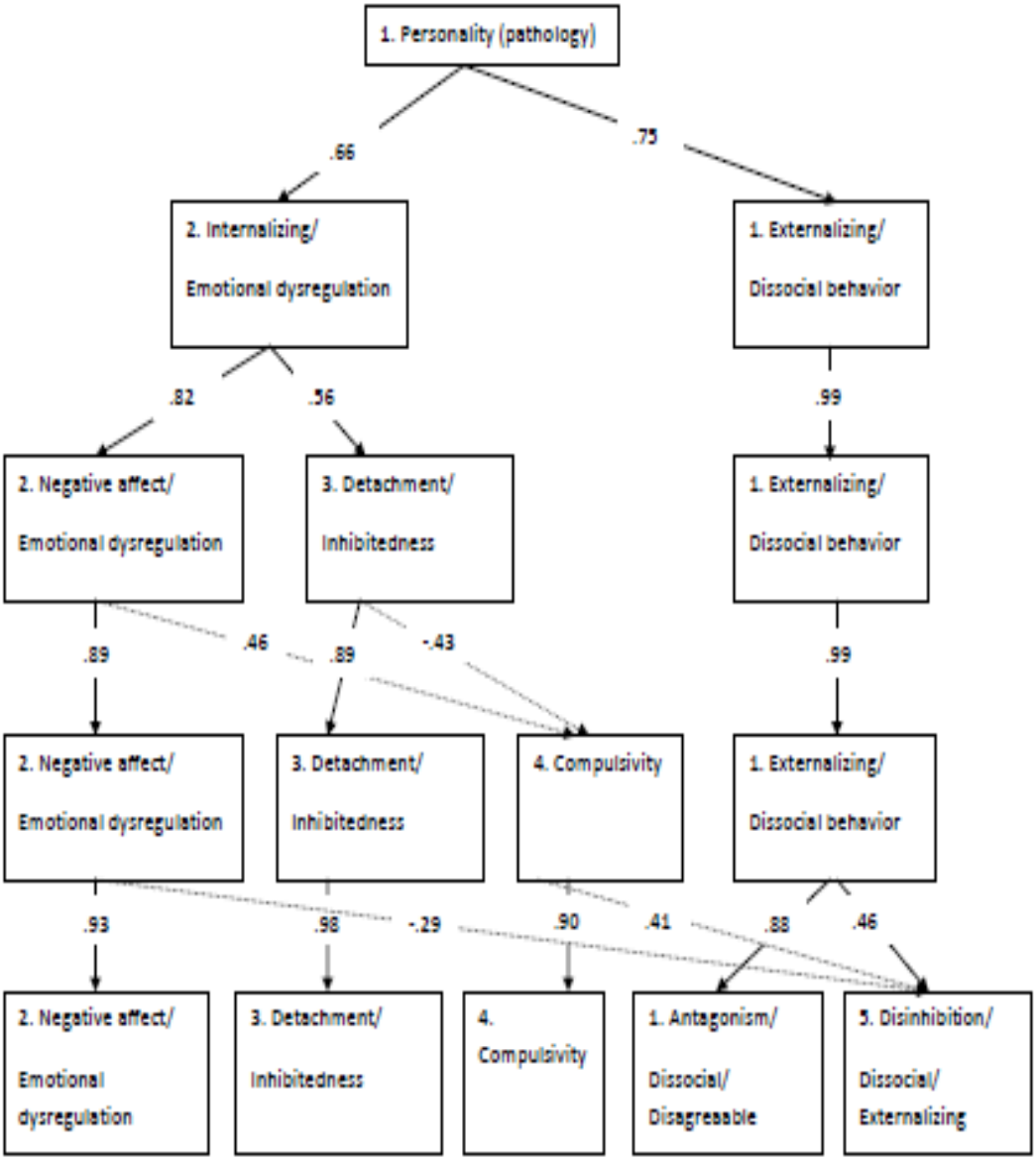

Figure 1: PID-5 - DAPP-BQ Joint Hierarchical Factor Structure

5.3. Results

Parallel analysis indicated the existence of four factors; however, for the fifth one the difference between the actual and the random eigenvalues was .01. Because of this reason, and because of a better interpretability, we decided to stop at the fifth level. In what follows, we will discuss each level of the joint hierarchical structure (see Figure 1[8]).

Level 1. In the one-factor solution all of the PID-5 and DAPP-BQ traits showed factor loadings > .40, with the exception of Intimacy Avoidance (.35) and Risk Taking (.20) for the PID-5, and Intimacy (.05) and Compulsivity (.24) for the DAPP-BQ. This component thus seemed to represent overall "Personality Pathology".

Level 2. The general "Personality Pathology" component subdivided into two components, labeled "Internalizing/ Emotional Dysregulation" and "Externalizing/ Dissocial Behavior". The "Internalizing/ Emotional Dysregulation" component was defined primarily by high loadings of the Anxiousness, Submissiveness, Depressivity, Emotional Lability, Separation Insecurity, and Anhedonia traits (PID-5) on one hand, and by high loadings of the Anxious, Affective Lability, Submissiveness, Suspiciousness, Low Affiliation, and Identity problems dimensions (DAPP-BQ) on the other hand. Scales with salient loadings (> .40) on the "Externalizing/ Dissocial Behavior" component were Grandiosity, Deceitfulness, Callousness, Manipulativeness, Attention Seeking (PID-5), and Callousness, Rejection, Narcisissm, and Stimulus Seeking (DAPP-BQ).

Level 3. The "Internalizing/ Emotional Dysregulation" component split into two subcomponents, "Detachment/ Inhibitedness" and "Negative

[8] Path coefficients < .25 are not shown. The factor solutions used in the analysis of the PID-5 – DAPP-BQ hierarchy can be obtained on request from the first author.

Affect/ Emotional Dysregulation", while the "Externalizing/ Dissocial Behavior" component maintained its structure. PID-5 traits and DAPP-BQ dimensions that loaded highest on the "Detachment/ Inhibitedness" component were Withdrawal, Anhedonia, Intimacy Avoidance, Depressivity, and Restricted Affectivity (PID-5), and Identity Problems, Intimacy Problems, and Restricted Expression (DAPP-BQ). The component "Negative Affect/ Emotional Dysregulation" was mainly marked by high loadings for Anxiousness, Emotional Lability, Separation Insecurity, and Perseveration of the PID-5, and Affective Lability, Submissiveness, Insecure Attachment, Low Affiliation, and Anxiousness of the DAPP-BQ.

Level 4. The three components from the previous level were largely replicated, and a fourth component emerged. This new component was marked by PID-5 Rigid Perfectionism and DAPP-BQ Compulsivity, along with a negative loading of Intimacy Problems (DAPP-BQ). This component was labeled "Compulsivity".

Level 5. At the fifth level of the hierarchy the component "Externalizing/ Dissocial Behavior" split to form two subcomponents, labeled "Antagonism/ Disagreeable" and "Disinhibition/ Externalizing". PID-5 Manipulativeness, Grandiosity, Hostility, Attention Seeking, Callousness, Deceitfulness, and DAPP-BQ Rejection, Callousness, and Conduct Problems loaded strongest on the "Antagonism/ Disagreeable" component. The "Disinhibition/ Externalizing" component was strongly marked by PID-5 Impulsivity and Distractibility, and by DAPP-BQ Stimulus Seeking.

5.4. Discussion

The goal of this study was to unravel the conceptual relations between the DSM-5 maladaptive personality traits and the DAPP-BQ's personality disorder trait-based dimensions. On levels one through three, the hierarchical structures of the PID-5 and DAPP-BQ coincided

in expected ways, thereby mirroring the findings from both Wright et al. (2012) and Kushner et al. (2011). At the second level, the two broad Internalizing and Externalizing dimensions originated from the general Personality Pathology component, replicating the broadly recognized Internalizing-Externalizing dichotomy of psychopathology (e.g., Achenbach, 1966; Krueger, 2002). At the third level of the hierarchy, three dimensions emerged that link to the "Big-Three" model of temperament (i.e., "Negative Affectivity/ Emotional Dysregulation", "Detachment/ Inhibitedness", and "Externalizing/Dissocial"; Clark & Watson, 2008; Wright et al., 2012). The components at the fourth level of the hierarchy represented the established "Big Four", with "Negative Affect/Emotional Dysregulation", "Externalizing/Dissocial Behavior", "Detachment/Inhibitedness", and "Compulsivity" as major dimensions.

Although we expected the PID-5 Disinhibition scales to represent the opposite pole of the Compulsivity component, hence reproducing Widiger and Simonsen's "Constraint vs. Impulsivity" bipolarity, they instead loaded primarily onto the "Externalizing/Dissocial Behavior" component. At the next and fifth level, the Disinhibition scales even split off to form a separate component in their own, together with some Externalizing DAPP-BQ scales. Although this finding was rather unexpected, it is in line with the fact that the position of Disinhibition/Impulsivity versus Compulsivity has been subject to controversy before. For example, in the initial DSM-5 proposal, Disinhibition and Compulsivity were considered separate structural components, with Disinhibition being conceptually linked to DAPP Dissocial Behavior (Krueger et al., 2011). Although both components were later unified into one bipolar domain (labeled "Disinhibition"; Krueger et al., 2012), the current findings rather connect with the initial proposal. Future research is thus needed to resolve this obscurity.

In contrast to Kushner et al. (2011), where Compulsivity split off from the Dissocial component, the origins of Compulsivity in this study were located in "Negative Affect/ Emotional Dysregulation" and (reversely)

in "Detachment/ Inhibitedness". One possible reason for this discrepancy may be that the Compulsivity component in the present study was somewhat broader (i.e., it included Rigid Perfectionism (PID-5), Compulsivity (DAPP-BQ), and Intimacy Problems (DAPP-BQ; negative loading)). Rigid Perfectionism is a (reversed) facet of Disinhibition in the PID-5, but it also shows considerable conceptual similarity to Perseveration, a facet of Negative Affect, hence possibly explaining its roots in this particular component. The negative loading of Intimacy Problems on Compulsivity is counter-intuitive as it is assumed that the more structured and organized a person is, the more likely (s)he is to be reserved and avoid intimacy. The reversal of this relationship in our study may therefore reveal a measurement bias; both the Intimacy Avoidance scale (PID-5) and the Intimacy Problems scale (DAPP-BQ) focus mainly on intimate relationships and sex, which may be a less valid indicator of intimacy in an older sample.

At the fifth and final level of the hierarchy, there was a bifurcation of the broad Externalizing dimension into Antagonism and Disinhibition. DAPP-BQ's Rejection, Callousness, and Conduct Problems loaded highest on the former, and Stimulus Seeking on the latter. Counter to our expectations, a separate "Psychoticism" component, as established in the PID-5 five-factor structure (Krueger et al., 2012; Wright et al., 2012), did not emerge from our data[9]. In contrast, the PID-5 Psychoticism scales loaded highest on the Antagonism (Eccentricity and Cognitive and Perceptual Dysregulation) and Disinhibition (Unusual Beliefs) components. Cognitive Dysregulation loaded highest onto "Negative Affect/ Emotional Dysregulation". This facet of the DAPP-

[9] Because one can wonder whether the absence of a separate Psychoticism domain might be an artefact of factor analyzing the PID-5 together with the DAPP-BQ, in which Psychoticism content may be underrepresented, we also performed a hierarchical factor analysis on the PID-5 itself. In this analysis, no separate Psychoticism component emerged either (the fifth level reproduced the same five components as in the joint analysis).

BQ captures disorganized thinking and could therefore be expected to cluster together with the PID-5 Psychoticism scales. Hence, this finding reveals a conceptual difference between the PID-5 Psychoticism scales, which focus more on odd thought processes in various sensory modalities and therefore tap more into schizotypal features, and the DAPP-BQ Cognitive Dysregulation scale, which is rather a marker of transient thought disturbances and feelings of confusion resulting from extreme anxiousness and distress (Livesley & Jackson, 2009).

When considering the unfolding of the hierarchy, it also became clear that the pathways of the PID-5 Psychoticism traits differed from those in Wright et al. (2012). In particular, in their study, Eccentricity and Perceptual Dysregulation originated from Detachment and Negative Affect, respectively (both derivatives of the Internalizing component), while Unusual Beliefs stemmed from the Externalizing component. In our study, all three facets stemmed from the Externalizing component. Thus, although not corroborating the existence of a separate Psychoticism component, the current results may nevertheless reveal an interesting finding concerning the structural hierarchy of Psychoticism-related traits across age, namely that they are associated with Internalizing traits in younger adults, and with Externalizing traits in older adults.

Despite its methodological (i.e., joint hierarchical factor analysis) and substantive (i.e., testing the hierarchical convergence of the PID-5 and the DAPP-BQ) contributions, our study is also subject to a number of limitations. First, the amount of participants per variable was relatively small, impeding the generalizability of the current findings and making it difficult to distinguish real age effects from possible sample bias. Second, the low internal consistency of the PID-5's Suspiciousness scale in the current older sample warrants further investigation. Finally, further research is needed to provide conclusive evidence regarding the placement of Psychoticism features within a maladaptive trait model, the negative loading of Intimacy Problems on Compulsivity, and the

structural relationship between Compulsivity and Disinhibition. However, despite a few irregularities, the results of this study are especially valuable in that they corroborate the idea of a common hierarchical structure underlying personality pathology (Krueger et al., 2011; Widiger & Simonsen, 2005), and generally support the fact that the PID-5 allows to capture this common structure. As such, our study adds to previous studies on the validity of the PID-5 by not only showing that the DSM-5 traits relate to the DAPP-BQ's dimensions, but also that the PID-5 and DAPP-BQ show meaningful associations on different levels of their hierarchical factor structure.

5.5. References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: a factor-analytic study. Psychological Monographs, 80, (No. 615).

American Psychiatric Association. (1994). DSM-IV: Diagnostic and Statistical Manual (4th Ed.). Washington, DC: Author.

De Clercq, B., De Fruyt, F., Mervielde, I., Krueger, B., & Markon, K. (2011). Dutch
translation of the Personality Inventory for DSM-5 (PID5).

Clark, L. A., & Watson, D. (2008). Temperament: An organizing paradigm for trait psychology. In O. P. John, R. W. Robins, and L. A. Pervins (Eds.). Handbook of personality: Theory and research (3rd ed., p. 265-286). New York: Guilford Press.

Goldberg, L. R. (2006). Doing it all Bass-Ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality, 40*, 347-358.

Harkness, A. R., & McNulty, J. L. (1994). The personality psychopathology five (PSY-5): Issue from the pages of a diagnostic manual instead of a dictionary. In S. Strack & M. Lorr (Eds.), Differentiating normal and abnormal personality New York: Springer.

Krueger, R. F. (2002). Psychometric perspectives on comorbidity. In J. E. Helzer & J. J. Hudziak (Eds.). *Defining psychopathology in the 21st centur: DSM-V and beyond* (pp. 41-54). Washington, DC: American Psychiatric Publishing.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42(9),* 1879-1890.

Krueger, R. F., Eaton, N. R., Clark, L. A., Watson, D., Markon, K. E., Derringer, J., Skodol, A., & Livesley, W. J. (2011). Deriving an empirical structure of personality pathology for DSM-5. *Journal of Personality Disorders, 25(2)*, 170-191.

Kushner, S. C., Quilty, L., C., Tackett, J. L., & Bagby, M. (2011). The hierarchical structure of the dimensional assessment of personality pathology (DAPP-BQ). *Journal of Personality Disorders, 25(4),* 504-516.

Livesley, W. J., & Jackson, D. N. (2009). DAPP-BQ: Dimensional Assessment of personality pathology- Basic questionnaire. Technical Manual. Sigma Assessment Systems. Inc.

Livesley, W. J., Jackson, D. N., & Schroeder, M. L. (1992). Factorial structure of traits delineating personality disorders in clinical and general population samples. *Journal of Abnormal Psychology, 101(3), 432-440.*

Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: Questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology, 27(7), 321-349.*

Van Kampen, D., & De Beurs, E. (2009). DAPP-BQ: Dimensionale assessment van persoonlijkheidspathologie inclusief screeningsversie. Amsterdam: Hogrefe.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: finding a common ground. *Journal of Personality Disorders, 19(2),* 110-130.

Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological

personality traits. *Journal of Abnormal Psychology.* Advance online publication. doi: 10.1037/a0027669.

Chapter 6
Personality assessment in older adults: The value of personality questionnaires unraveled

Joke Van den Broeck, Bas Van Alphen, Gina Rossi, Eva Dierckx, & Dan Segal

Submitted for publication in Clinical Gerontologist

The main goal of personality assessment in clinical settings is to paint a picture of a patient's personality characteristics, both in terms of personality strengths and deficits, and to relate them to the referral question, in order to diagnose personality disorder, set up treatment plans, and/or evaluate treatment outcomes. This process involves the collection and evaluation of various sources of information about individuals, such as clinical interviews, biographical material, self-report questionnaires, and behavioral observations (Wiggins, 2003; Weiner & Green, 2008). Administering self-report personality questionnaires makes up an important part of this data gathering, as it is seen as an inexpensive, reliable, relatively quick, and easy way to collect data. On the other hand, self-report inventories also have their drawbacks (McDonald, 2008), especially with regard to older adults. In this article we aim to unravel the specific value of personality questionnaires in clinical geropsychology and geriatric psychiatry.

Despite a long and rich history of personality assessment in psychology more generally, little attention has been given to personality assessment in older adults (> 65 years old), both in research and clinical practice (e.g., Segal, Coolidge, & Rosowsky, 2006). During the past decade, however, interest in this topic and consequentially the amount of studies and scientific publications has increased steadily. Despite the recent

growing interest in the assessment of personality and personality pathology in older adults, research in this field is hampered by both conceptual and methodological issues (Clark, 2007).

It seems like a vicious cycle: the conceptualization of maladaptive features of DSM-based personality disorders in older adults is hampered by limitations in its assessment, while improving the current measurement system is limited by the lack of knowledge about the conceptualization of personality pathology in later life. An additional difficulty is that there is no "gold standard" in personality assessment, and certainly not for the assessment of personality disorders in older age groups (Van Alphen, Engelen, Kuin, Hoijtink, & Derksen, 2006; Balsis, Segal, & Donahue, 2009). The most fundamental issues among older adults concern the applicability and relevance of the current DSM-IV nosology (Tackett, Balsis, Oltmanns, & Krueger, 2009). Many current DSM-IV Axis II criteria do not take into account the age-specific changes in behavior and interpersonal functioning, therefore lacking face validity for use in later life (e.g., Van Alphen et al., 2006; Balsis, Gleason, Woods, & Oltmanns, 2007). In addition, the majority of the current personality assessment measures have been developed for and validated in mixed-age younger adult samples, and the lay-out, item content, and norms are often not attuned to the specific context of later life (Van Alphen, 2006; Zweig, 2008). Up till now, researchers and practitioners are confronted with the lack of valid and appropriate personality measurements in older adults, and researchers in this area are facing the challenging task of filling this gap, as a first step towards a better understanding of personality pathology in later life.

An important issue in this regard is whether to advocate for an age-specific or an age-neutral measurement system? Clearly, both approaches have their advantages and disadvantages (Rosowsky & Segal, 2010). For research purposes, the advantages of an age-neutral measurement system are obvious. Think of investigators interested in studying the course of personality longitudinally, or those investigating

(mal)adaptive personality features cross-sectionally among younger and older individuals. In both cases, researchers can certainly benefit from an age neutral measure that works equivalently well across all age groups (Balsis et al., 2007; Tackett, Balsis, Oltmanns, & Krueger, 2009). For example, epidemiologic studies suggest that Extraversion tends to remit with age, whereas recently it has been empirically demonstrated that the Extraversion domain, as measured by the *Revised NEO Personality Inventory* (NEO-PI-R; Costa & McCrae, 1992), contains several items that are more readily endorsed by younger adults compared to older adults (e.g., "I have done things just for kicks or thrills"; Van den Broeck, Rossi, Dierckx, & De Clercq, 2012). Such findings call into question the reliable comparability of personality constructs across age and underscore the importance of an age-neutral measurement system.

In clinical practice the merits of an age-neutral measure are ostensibly worthwhile, as it would enable clinicians to rely on valid assessment instruments, without having to adjust items to assess their older patients (Zweig, 2008, Tackett et al., 2009). It would also be conducive for comparability after retesting, for example when a 70-year-old patient has been hospitalized and tested, the current results could be easily compared with previous test results of this patient. On the other hand, one might argue that from a practical view, a first and foremost requirement is a valid instrument in order to screen and/or diagnose personality disorder within a specific population, whether it is age neutral or not. As a matter of fact, an age-neutral measure is no guarantee for practical usefulness.

Especially when working with older adults, some clinicians may prefer an age-specific measure, one that is specifically developed and validated for older people with items that probe the specific aging context. A major advantage of such an age-specific measurement instrument is that it can take into account the specific diagnostic difficulties related to the assessment of personality disorders in older adults, and anticipate practical bottlenecks such as length and complexity. The MMPI-2 for

example, one of the most widely used personality inventories in clinical psychology , consists of 567 items, some of which require a certain level of literacy (e.g., " Most people will use somewhat unfair means to gain profit or an advantage rather than to lose it"), which can obviously overtax cognitively impaired older patients. Other problems are most relevant when testing old-old populations (aged 75 years or older) (Aldwin & Levenson, 1994). For example, many older adults in this group are unfamiliar with test situations in which they have to assign numbers to their experiences on to a rating scale, making them reluctant to participate in such assessment procedures. Furthermore, technical, abstract or modern language often used in current personality inventories may hamper a reliable assessment of those older adults with less formal education (Van Alphen et al., 2006). There may also be a cohort difference in language use contributing the validity issue. Older adults, for example, may be less inclined than younger adults to describe their lives in terms of "problems" or "stress" (Aldwin & Levenson, 1994).

Besides the fact that many practitioners prefer multi-method assessments (e.g., Spitzer, 1983), most personality research still relies solely on self-reports, such as the MMPI-2 (Vazire, 2006). Yet the shortcomings of self-report measures, especially for the assessment of personality disorders, are extensively described in the research literature (e.g., Klonsky, Oltmanns, & Turkheimer, 2002; McDonald, 2008). One important drawback relates to the limited insight in self and interpersonal relations, inherent to the ego-syntonic nature of a personality disorder. Another major limitation pertains to the fact that self-report provides only one viewpoint, whereas, ideally, personality assessment involves the gathering and evaluation of various sources of information (Klonsky et al., 2002).

Informant reports may alleviate some of the above-mentioned shortcomings of self-report inventories. Especially within an older adult population, where an increased risk of cognitive decline exists as a result

of normal aging or due to the higher prevalence of degenerative diseases, informants may play a crucial role in the process of data collection (American Psychological Association, 2004). Indeed, research shows that using peer-ratings adds a unique perspective in the description of personality disorder features, and that informants are able to provide the clinician with a more nuanced picture of the patient (Lawton, Shields, & Oltmanns, 2011). Interestingly, the comparison between self and other reports often reveals a paradox, such as when people who are rated by others as being paranoid and suspicious rather describe themselves as being angry and hostile (Clifton, Turkheimer, & Oltmanns, 2004). Or, from the opposite perspective, people who describe themselves as being paranoid are often seen by others as being cold and unfeeling. According to Oltmanns and Balsis (2010) however, it is fair to state that "*utilizing information from both sources may help a clinician gain a more comprehensive picture of a client's personality disorder than if the clinician were to rely solely on one source of information"* (Oltmanns & Balsis, 2010, p. 111).

However, many fundamental questions still remain unanswered (e.g., Klonsky et al., 2002) as to how one should handle discrepancies between self- and other-reports, who should be selected as an informant, and how this selection might influence the results. Typically, a patient is asked to suggest a friend or family member who knows the patient well, and mostly patients select someone they like and whom they assume like them back. Research shows that these selected informants are more inclined to provide overly positive ratings, and that the value of these selected informant ratings depends on the type of problem being assessed. For example, selected informants report lower levels of narcissism, paranoia, and antisocial personality disorder compared to scores provided by other informants (not selected by the patient) (Oltmanns & Turkheimer, 2006). Specifically with regard to older adult samples, some specific considerations need to be addressed, such as who can provide the most reliable information (e.g., clinicians, spouses or adult children), and which instructions to give to the informant, in

terms of the reported time period (e.g., report on the whole life, or the past ten years, etc…). Regarding this latter issue, it is important to denote whether one is interested in the present or in premorbid personality characteristics, and depending on this choice one has to decide which kind of informant is best qualified to provide the most useful information. Clearly more research is needed to understand fully how informant reports should be incorporated into the personality assessment process, especially among older adults who often have longer and more complicated personal histories.

From all the above, it may be clear that gathering data through self-report questionnaires is only one part of a much broader diagnostic process (De Bruyn, Claes, & Bijttebier, 2006). However, taken into account their relative but important role in the diagnostic process, pertinent questions are what weight should be given to these personality test outcomes and how should clinicians and researchers best use them. In general, there are several options when drawing inferences from test data, roughly distinguishable into a nomothetic and an ideographic approach (Weiner & Greene, 2008). In a nomothetic approach, the assessor relies on empirical and statistical rules for drawing conclusions from test data. An individual's test response is compared to norms and statements are made about how much the individual resembles those in the norm group. However, administering tests, counting scores, and computing formulas alone are not sufficient, especially not when one has to rely on measures with doubtful validity as is often the case in clinical geropsychology. Information about a person's prior experiences, sociocultural background, and current life circumstances are essential features that need to be taken into account when interpreting test data and formulating treatment plans. This person-specific information plays a central role within an ideographic approach that focuses on the unique richness of an individual's experiences. Ideally, both approaches complement each other, and should be used simultaneously. When assessing older adults however, the importance of unique lifetime

experiences and the age-specific context play even a more important role because norms are often not available for this specific age group.

Clinical geropsychologists should therefore always interpret the results of a personality questionnaire in light of the actual, specific context. For example, test results may indicate that a patient is fearful, clinging and helpless, raising the idea of a dependent personality disorder at first sight. In contrast, an entirely different interpretation can be given to these results when it is discovered that this patient is actually suffering from a recent loss of a long-time partner, on whom the patient had a healthy dose of interdependence. Due to increased health problems and an increased frequency of transitions, later life is commonly seen as a turbulent period in terms of behavioral and affective expressions (Oltmanns & Balsis, 2011). Thus integrating adaptive and maladaptive personality traits into a more holistic framework that takes into account a patient's life story is a valuable and rewarding challenge. Along with the integration of various sources of information this strategy will lead to a better understanding of personality disorders in later life.

References

Aldwin, C. M., & Levenson, M. R. (1994). Aging and personality assessment. *Annual Review of Gerontology and Geriatrics, 14*, 182-209.

American Psychological Association. (2004). *Guidelines for psychological practice with older adults, 59*, 236-260.

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging, 22(1),* 171-185.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58,* 227-257.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-factor Inventory (NEO-FFI).* Odessa, FL: Psychological Assessment Resources.

De Bruyn, E., Claes, L., & Bijttebier, P. (2006). Het psychodiagnostische proces. In Claes, L., Bijttebier, P., Vercruysse, T., Hamelinck, L., & De Bruyn, E. (Eds). Tot de puzzel past. Psychodiagnostiek in methodiek en praktijk. Leuven, Acco.

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9,* 300-311.

Lawton, E. M., Shields, A. J., & Oltmanns, T. F. (2011). Five-factor model personality disorder prototypes in a community sample: Self- and informant-reports predicting interview-based DSM diagnoses. *Personality Disorders: Theory, Research, and Treatment, 2(4),* 279-292.

McDonald, J. D. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire, 1(1)*, 1-19.

Oltmanns, T. F., & Turkheimer, E. (2006). Perceptions of self and others regarding pathological personality traits. In Krueger, R. F., & Tackett, J. L. *Personality and psychopathology*. New York, London: The Guilford Press.

Oltmanns, T. F., & Balsis, S. (2010). Assessment of personality disorders in older adults. In P. A. Lichtenberg (Ed.). Handbook of Assessment in Clinical Gerontology (pp. 101-122). San Diego: Academic Press/Elsevier.

Rosowsky, E., & Segal, D. (2010). Personality disorders in later life. In N. A. Pachana, K. Laidlaw, & Knight, B. G. (Eds.). *Casebook of clinical geropsychology: International perspectives on practice* (pp. 195-211). Oxford: University Press.

Segal, D. L., Coolidge, F. L., & Rosowsky, E. (2006). *Personality disorders and older adults: Diagnosis, assessment, and treatment.* Hoboken, New Jersey: Wiley.

Spitzer, R. L. (1983). Psychiatric diagnosis: are clinicians still necessary? *Comprehensive Psychiatry, 24, 399-411.*

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: Developmental considerations for the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. *Development and Psychopathology, 21,* 687-713.

Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., Hoijtink, H. J. A., & Derksen, J. J. L. (2006). A preliminary study of the diagnostic accuracy of the Gerontological Personality disorder Scale (GPS). *International Journal of Geriatric Psychiatry, 21,* 862-868.

Van Alphen, S.P.J. (2006). *Diagnostiek van persoonlijkheidsstoornissen bij ouderen: een bijdrageaan de ontwikkeling van een screeningsinstrument*. Academisch proefschrift, Eindhoven: University Press Facilities.

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality, 40,* 472-481.

Weiner, I. B., & Greene, R. L. (2008). *Handbook of personality assessment*. Hoboken, NJ: John Wiley & Sons, Inc.

Wiggins, J.S. (2003). Paradigms of Personality Assessment New York: Guilford Press

Zweig, R. A. (2008). Personality disorder in older adults: Assessment challenges and strategies. *Professional Psychology: Research and Practice, 39(3),* 298-305.

Chapter 7
General discussion

7.1. Introduction

The topics covered in this dissertation address the growing interest in the study of personality disorders in older adults, and are related to the apparent need for valid and age-appropriate personality assessment tools. As discussed previously, many DSM-IV diagnostic criteria for personality disorders do not fully apply to the later life context, hampering a valid assessment of personality disorders in older adults and creating substantial measurement issues for the field (Oltmanns & Balsis, 2011). There are two alternative approaches to address these measurement issues, namely the development of an age-specific or an age-neutral measurement system.

To date, several age-specific measurement instruments exist, specifically developed and validated for personality assessment in older adults (Van Alphen, Derksen, Sadavoy, & Rosowsky, 2012), such as the Gerontological Personality Disorder Scale (GPS; Van Alphen, Engelen, Kuin, Hoijtink., & Derksen, 2006), the Hetero-Anamnestic Personality Questionnaire (HAP; Barendse, Thissen, Oei, Rossi, & Van Alphen, *in press*), and a hybrid PD scale of 100 items (Balsis, 2009). The GPS was designed by Van Alphen and colleagues (2006) with the intention to contribute the psychodiagnostic process of elderly people in the ambulant mental health setting. This short test is based on the general diagnostic criteria of the DSM-IV-TR (APA, 2000); specific personality disorders cannot be diagnosed with the GPS. Yet, a high score can be indicative for personality pathology. The HAP (Barendse et al., *in press*) is an informant questionnaire originally developed to assess premorbid personality characteristics in the elderly. It is not designed to identify specific personality disorders either, but it can as well be useful in clinical practice as a screener for personality pathology. A third age-

specific measure that was developed is a hybrid PD scale (Balsis, 2009). This measure differs from the GPS and the HAP in that it was specifically created to improve upon the current diagnostic DSM-IV Axis II criteria. New items were generated based on clinician experiences that would better capture personality disorders in later life. Although requiring further research and refinement, this kind of development is of major interest for the field, as it underscores a fundamental issue of personality pathology assessment in older adults: the importance of taking into account the specific aging context.

Although the age-specific measures described above may work well to assess personality disorder pathology in older adults, investigators might prefer to rely on an age-neutral measure to study personality disorder pathology longitudinally into later life, or cross-sectionally among younger and older individuals. Thus, another approach to address the measurement issues associated with personality pathology assessment in older adults is the development of an age-neutral measurement system that works equivalently well across all age groups (Tackett, Balsis, Oltmanns, & Krueger, 2009; Oltmanns & Balsis, 2010; Oltmanns & Balsis, 2011). As discussed in Chapter 1, at least two personality measures were created with the goal of age-neutrality: the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) and the Personality Assessment Inventory (PAI; Morey, 1991).

All in all, taking into account the issues of personality disorder assessment as they relate to later life forces researchers to develop and/or apply more sophisticated measurement models (Oltmanns & Balsis, 2011). Unfortunately the later life context was not considered in the development of a new personality disorder classification system for DSM-5 either. According to the latest DSM-5 proposal, the description of personality disorders will undergo substantial revision, such as the inclusion of a dimensional trait model for personality pathology. It is important to note however that no decisions have yet been formalized regarding the conceptualization of personality disorders in DSM-5

(Krueger, Derringer, Markon, Watson, & Skodol, 2011). As of this writing, a unipolar trait paradigm and corresponding instrument has been proposed that encompasses elements of extreme and maladaptive personality variation in order to capture dispositional features of personality disorders (APA, 2012). Nonetheless, there is compelling evidence that normal-range personality traits also provide clinically useful information (e.g., Samuel, 2011), and the proposal to adopt a unipolar instead of a bipolar trait system is questioned and critized by several leading researchers in the field. However, most of the work on the validity and utility of dimensional models in the assessment of personality pathology has focused on young and middle-aged adult populations, little is known about their applicability in later life.

The four empirical studies presented in this dissertation addressed the abovementioned issues, namely the age-neutrality of the NEO-PI-R and the PID-5 (Chapters 2 and 4, respectively), the suitability of the NEO-PI-R to assess personality pathology in later life (Chapter 3), and the convergent validity of the proposed maladaptive trait model (Chapter 5). In this final chapter, we summarize and discuss the major findings of these studies from a broader perspective. We also dwell upon some practical implications of the current work, and conclude by briefly discussing general limitations and directions for further research.

7.2. Summary of findings

Study 1, reported in Chapter 2, was the first to empirically investigate the NEO-PI-R's proclaimed age-neutrality. Differential Item Functioning (DIF) analyses identified several poor-performing items, although the vast majority of items (92 to 95%) were similarly endorsed by younger and older adults with the same level of underlying personality trait. Differential Test Functioning (DTF) analyses were then performed to explore the impact of the DIF items on the validity of their respective scale. These analyses revealed large DTF for Extraversion (E) at the domain-level, and large DTF for Tender-

Mindedness (A6) at the facet-level. Overall, the percentage of items displaying DIF was considerably lower than the threshold of 25% put forward by Penfield and Algina (2006) as an indicator of biased results. Therefore, the conclusion was drawn that the present findings corroborate the NEO-PI-R's age-neutrality in measuring personality traits.

These results formed the basis for our second study (Chapter 3), in which we addressed the applicability of the NEO-PI-R to screen for personality pathology in older adults, by using the FFM PD count technique. Overall, eight out of ten FFM PD counts (all but the antisocial and obsessive-compulsive counts) exhibited adequate convergent and divergent validity, supporting the use of this method in older adults. A major contribution of this study was that cut-off scores were computed and validated against both a categorical and a dimensional measure of personality pathology. However, testing whether the presented cut-offs can effectively distinct between non-disordered and disordered patients in clinical samples is necessary to investigate their ultimate validity as a screener for personality pathology in older adults on one hand, and their possible applicability as a diagnostic tool for this age group on the other hand.

The link between these first two studies is situated within the adaptive versus maladaptive trait literature. There is a considerable amount of research demonstrating that personality disorders represent maladaptive, extreme variants of general personality traits (e.g., Markon, Krueger, & Watson, 2005; Clark, 2007; Widiger & Mullins-Sweatt, 2009), and compelling evidence suggests that normal-range personality traits also provide clinically useful information (Samuel, 2011). From this perspective, both NEO-PI-R studies complement each other and offer evidence for the NEO PI-R as a valid screening tool for the assessment of (mal)adaptive personality traits in older adults, enabling a valid and comprehensive description of both an older patient's personality difficulties *and* strengths. Some of these strengths can conveniently be

used to set up a treatment plan, such as agreeableness indicating an engagement in group therapy (Widiger & Mullins-Sweatt, 2009). By administering the NEO-PI-R and applying the proposed PD counts, clinical gerontologists will be able to assess the well-known personality disorder constructs in a dimensional way.

A dimensional approach toward the conceptualization and assessment of personality disorders is one of the major proposed changes in the upcoming fifth edition of the DSM. Numerous leading researchers in the field argue for the implementation of a dimensional model that encompasses the full range of both normal and abnormal functioning (e.g., Widiger & Mullins-Sweatt, 2009; Samuel, 2011). Although the DSM-5 Personality and Personality Disorder Workgroup recognizes that the dimensional domains are bipolar when considering both adaptive and maladaptive aspects of personality, they are convinced that the features of personality disorders tend to be concentrated specifically at the maladaptive poles of these domains (i.e., detachment, antagonism, disinhibition, and negative affectivity) and therefore proposed a maladaptive trait paradigm for the conceptualization of personality disorders in DSM-5. In the two last studies, reported in Chapter 4 and 5, we aimed to investigate the validity of this proposed DSM-5 trait model and its associated assessment instrument, the Personality Inventory for DSM-5 (PID-5), for use in older adults. In first instance, the age-neutrality of the PID-5's maladaptive traits was investigated, parallel to the analyses on the NEO-PI-R's adaptive traits reported in Chapter 2. The results of the DIF and DTF analyses on PID-5 data were somewhat less straightforward. Although still under the 25% threshold proposed by Penfield & Algina (2006), 33 items appeared to display large DIF, and substantially impacted at the scale level in four of the 25 cases (i.e., large DTF was found for Withdrawal, Attention Seeking, Rigid Perfectionism, and Unusual Beliefs). A facet-level comparison with the NEO-PI-R study forced us to draw a more nuanced conclusion on the current PID-5 analyses. In the case of the NEO-PI-R, 1 out of 30 facets displayed large DTF (3.3%), in the case of the PID-5 this ratio amounted 4 out of

25 (16%). It was therefore concluded that although initial results revealed that the majority of the PID-5 traits are measured equally well across age, additional research is needed to further refine this instrument in terms of age-neutrality. The PID-5 was actually proposed as a research tool, with the goal of encouraging refinement and development prior to the finalization of the DSM-5 (Krueger, Derringer, Markon, Watson, & Skodol, 2011). This offers an ideal opportunity to make work of a well-thought age-neutral measure, in which these initial results may stimulate further research.

In a second PID-5 study, reported in Chapter 5, we focused on the convergence of the proposed DSM-5 trait model and its corresponding assessment instrument with the Dimensional Assessment of Personality Pathology (DAPP) model in older adults. Several authors argue that the proposed DSM-5 trait structure does not correspond to the established "Big Four" domains of introversion, antagonism, impulsivity, and emotional dysregulation (Pincus, 2011; Widiger, 2011a; Widiger, 2011b). Yet in their rationale for the proposed changes to the personality disorder classification in DSM-5, the DSM-5 workgroup state that the overall structure of the 5 domain/25 facet system does correspond to the "Big Four" domains characterizing other trait models (such as the DAPP), with compulsivity representing the opposite pole of a bipolar domain of disinhibition (APA, 2012). In order to unravel the relations between the maladaptive trait dimensions of both models, the joint hierarchical structure of the PID-5 and the DAPP-BQ was examined. In general, the results of this study corroborate the idea of a common hierarchical structure underlying personality pathology. Interestingly, the hierarchical unfolding of trait dimensions in the current older sample largely resembled the individual DAPP-BQ and PID-5 hierarchical structures established in samples of younger adults (Kushner et al., 2011; Wright et al., 2012). The joint components at the fourth level of the hierarchy represented the established "Big Four", with "Negative Affect/Emotional Dysregulation", "Externalizing/Dissocial Behavior", "Detachment/Inhibitedness", and "Compulsivity" as major dimensions.

Yet some deviations were discussed warranting further investigation, such as the structural relationship between Compulsivity and Disinhibition, and the integration of Psychoticism features within a maladaptive trait model.

7.3. Limitations

Although specific limitations were presented at the end of each study, several general overarching limitations can be mentioned regarding some of the strategies and choices made in the current dissertation.

A first important drawback pertains to the exclusive reliance on self-reports, given the significant limitations associated with the use of subject reports for assessing personality and personality pathology. Participants may not be entirely honest in the report of undesirable traits and behaviors, and their reports may be distorted by their clinical and/or emotional state at the time of assessment (Stuart, Simons, Thase, & Pilkonis, 1992). Especially older adults may be influenced by the stigma attached to socially undesirable behaviors (Abrams & Bromberg, 2007). In addition, a valid personality assessment requires an adequate self-insight in one's own behaviors and their impact on others in social interactions (Klein, 2003), whereas maintaining a stable sense of self and managing interpersonal relationships are the core problems for people with maladaptive personality features. Unfortunately, the sole use of self-reports in personality research is to date still the rule rather than the exception, despite the knowledge that basing personality assessments on a combination of patient's and informant's reports would certainly benefit a valid assessment process (Klonsky, Oltmanns, & Turkheimer, 2002). As discussed in Chapter 6, informant reports may be a meaningful complement or a useful alternative, especially when working with older adults suffering from degenerative diseases such as dementia or Alzheimer's disease. Further research is needed however to further explore the comparative validity of these two data sources, because

which source has the greatest validity and whether they provide unique information remain open empirical questions (Klein, 2003).

A second limitation relates to the recruitment procedures applied in the current studies. College students were asked to recruit older adults in return for course credit (Chapter 2, 3, and 4), or elderly participants were voluntarily recruited from leisure clubs and senior meetings, whether or not preceded by email contact (Chapter 1). In both cases, it is plausible that friendly, cooperative people are overrepresented in the current samples, because of some kind of self-selection bias. It is not inconceivable for example, that people who chose to participate in the study possess some personality characteristics (e.g., open-mindedness, helpfulness) that substantially differ from those who did not participate. Also, this procedure implies that only healthy, well-functioning individuals were included, because students may intuitively approach someone who has sufficient (cognitive and physical) abilities to fill in all the required questionnaires. These pitfalls should be kept in mind when drawing conclusions from the current findings, because the use of convenience samples question their generalizability towards older adults, as they are probably not representative of this age group in general. On the other hand, numerous personality disorder studies make use of samples of convenience, including undergraduate students or relatively accessible groups of (young) patients. Their use is therefore justifiable to some extent, especially in relatively new and emerging fields such as personality assessment in older adults (Oltmanns & Balsis, 2011). Also the different lower bounds of age 50 (Chapter 3), 60 (Chapter 3 and 4) and 65 (Chapter 2) may limit the generalizability of our findings. We decided to use the general term "older adults" throughout this dissertation, but the wide diversity and heterogeneity of this group should not be underestimated. Clearly there might be significant differences between a 60- and an 90-year-old, for example in terms of physical and cognitive functioning. Thus, as a continuation of the current initial studies, it might be interesting for further research to strive for more representative (clinical) samples (e.g., in terms of age,

gender, educational level,…), and to take the heterogeneity of this group into account by using smaller age-ranges.

7.4. Directions for further research

The development of reliable and valid tools for personality pathology assessment in older adults is certainly an interesting and important avenue for further research. However, as discussed in Chapter 6, investigators primarily need to address the fundamental issue of age-specificity versus age-neutrality, since both approaches have their advantages and disadvantages. Further research is also needed that focuses on the practical applicability and psychometric properties of existing personality measures like the HAP (Barendse et al., *in press*) and the GPS (Van Alphen et al., 2006), and on the value of age-specific measures for personality pathology assessment purposes more generally. An important topic in this regard is to explore how the "aging context" differs from a "young context" in terms of physiological, social, and occupational aspects, and how these contexts influence the presentation of personality disorders across the lifespan. Much would be learned from large-scale longitudinal studies that run into old age in which experts follow personality disordered patients and provide detailed descriptions of their behaviors and personality features. The acquired knowledge regarding the course and manifestation of personality disorders across the lifespan could in turn inform investigators in developing an age-sensitive personality disorder nosology, with associated assessment instruments. Another challenge for further research lies in the development of an age-neutral measurement system that works equivalently well across all age groups. In the process of identifying items that contain no age-related measurement bias, researchers will face the difficult task of searching for the core aspects that capture personality disorders. For instance, the item "Avoids occupational activities" may contain bias because it lacks face validity within a retired population. The general concept of social avoidance however might be an essential feature of this particular personality

disorder, so the challenge would be to create a neutral item that captures the phenomenon equally well in both a younger and an older sample (Oltmanns & Balsis, 2011, p. 14). Initial steps towards the detection of potentially biased items in the NEO-PI-R and the PID-5 were undertaken in the current dissertation, however more sophisticated research methods are needed, for example to learn more about the underlying causes of DIF. Within an Item Response Theory (IRT) framework, the possible influence of confounding variables such as cohort-effects, gender, and educational level needs to be further investigated.

In sum, more studies are needed to develop adequate measures that allow us to fully understand the prevalence, course, and influence of personality in later life (Tackett et al., 2009). The current transition period between DSM-IV and DSM-5 offers an interesting opportunity to study and incorporate developmental issues in the conceptualization of personality disorders. The publication of the fifth edition of the DSM in May 2013 is undeniably an intriguing event within the mental health field and plays a crucial role in determining directions for further research. As of this writing however, it is not clear how the actual personality disorder classification will look like in DSM-5, and final recommendations are still under construction. The DSM-5 proposal for the diagnosis of a personality disorder is currently based on two fundamental criteria: impaired personality functioning and the presence of pathological traits. Our focus was on the latter, and a discussion of the former falls beyond the scope of this dissertation. However, exploring the general effects of aging on personality (dys)functioning might be an interesting topic for further research as well, along with the development of age-appropriate tools to measure it.

7.5. Practical implications

Some major issues concerning the assessment of personality and personality pathology in older adults were extensively discussed in

Chapter 6. In this section we focus on the use of personality questionnaires and make an effort to answer a pertinent question that might occupy clinicians in the field after reading this dissertation: How should I measure personality (pathology) in older adults, and which measures should I use?

In one of the leading papers concerning the assessment of personality disorders in older adults, Zweig (2008) emphasized how assessing personality disorders in older adults poses unique diagnostic difficulties to psychologists, and provided several practical strategies. His recommendations sounded as follows: "(a) Utilize measures that are psychometrically suitable to older adults (i.e., validated in normative samples of older persons (…)); (b) exercise caution in applying measures or methods developed for younger adults populations, and tailor assessments to older adults' specific contexts" (Zweig, 2008, p. 303); (c) evaluate cognitive change, health status and medications, and functional impairment as part of a comprehensive assessment; and (d) maximize collaboration with interdisciplinary professionals and other informants as part of the assessment process (APA, 2004; APA Working Group on the Older Adult, 1998)". These recommendations broadly coincide with the practical guidelines recently proposed by the Dutch-Belgian expert board on personality and older adults (*Expertpanel Persoonlijkheid & Ouderen*) (Van Alphen, Barendse, Tummers, & Rossi, 2010). They advise to combine the Longitudinal, Expert, and All Data (LEAD) standard (Spitzer, 1983), with a stepwise, multidimensional approach for personality assessment in older adults (for a more detailed description of this approach we refer to Van Alphen et al., 2010; for a case-based illustration we refer to Van den Broeck, Barendse, Van Alphen, Thissen, & Rossi, 2012). In short, the proposed stepwise diagnostic procedure includes the screening for personality pathology in a first phase, followed by a global (or more elaborated) personality assessment in a second (or third) phase if necessary.

Based on the findings presented in this dissertation, we can recommend administering the NEO-PI-R in a first screening phase. Also the GPS (Van Alphen et al., 2006) and a general symptom checklist as the SCL-90 (Derogatis, 2003) can be easily administered in this first stage. By computing the proposed FFM PD counts and comparing a patient's score to the normative data presented in Chapter 3, clinicians may have a first impression of their patient's personality profile and whether or not a more elaborated personality assessment is needed. We must recognize however, that the administration of a 240-item NEO-PI-R questionnaire can be demanding for older people with physical or cognitive constraints. From this view, the study of Mooi and colleagues (2011) regarding the construction of a short version of the NEO-PI-R for older adults is worth mentioning. They asked experts to evaluate the NEO-PI-R items in terms of irrelevance of content, and vocabulary and formulation complexity. In doing so, 120 items were removed, leading to a shorter version of 120 items, the NEO-PI-R-SF. Given its time-saving qualities and feasibility, it might be interesting to further explore its usage in the computation of concise FFM PD counts for use with older adults.

When an in-depth personality assessment is warranted, we recommend the use of a semi-structured diagnostic interview, because these are often considered the gold standard in personality disorder assessment (Lawton, Shields, & Oltmanns, 2011). However, given the limitations associated with the poor face validity of some diagnostic criteria and the lack of norms for older adults, caution is warranted and interpretations must be made in light of the specific context, as is also discussed in Chapter 6. Other instruments that might be considered in this stage are, for instance, the MMPI-2 (if feasible), and the HAP. Of all commonly used personality pathology measures in Flanders and the Netherlands (e.g., ADP-IV, MCMI-III, VKP), these are the only ones that have been validated in clinical normative samples of older adults (Van Alphen, Barendse, Tummers, & Rossi, 2010). We conclude here by stating that, unfortunately, one has to play with the cards that were dealt, and that

this is particularly true for personality disorder assessment in older adults. We restricted ourselves in this section to the discussion of personality questionnaires (and semi-structured interviews), but we emphasize the importance of integrating various sources of information (e.g., biographical material, behavioral observations, clinical interviews), and underscore the value of informant reports in the assessment process (cfr. Chapter 6).

7.6. Conclusion

Central theme of this dissertation is the great need for a valid and useful measure for personality pathology assessment in older adults. To develop and present such a tool would have been a dreamed outcome of this PhD, but unfortunately that proved unfeasible within the given time span. Nevertheless we are convinced that the initial findings of the empirical studies in this dissertation are valuable, and we hope they contribute and inspire further research in the field. Overall, the findings of the current NEO-PI-R studies are interesting not only from a clinical perspective, but also for research purposes. On one hand, these findings justify the comparability of NEO-PI-R (facet-level) profiles across age, interesting for those investigators who wish to study personality traits cross-sectionally across younger and older individuals. On the other hand, they provide those who want to study the course of personality longitudinally with a valid assessment tool on which they can rely without worrying about possible age-associated measurement artifacts (Tackett et al., 2009). Regarding the PID-5 studies, we hope they may contribute to the further refinement of a maladaptive personality trait model and corresponding assessment instrument for DSM-5, with special attention for the validity and applicability in older adults.

7.7. References

Abrams, R. C., & Bromberg, C. E. (2007). Personality disorders in the elderly. *Psychiatric Annals, 37(2),* 123-127.

American Psychiatric Association. (2012). *Rationale for the proposed changes to the personality disorders classification in DSM-5*. Retrieved from http://www.dsm5.org.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed., text revision. Author, Washington, DC: 2000.

American Psychological Association Working Group on the Older Adults. (1998). What practitioners should know about working with older adults. *Professional Psychology: Research and Practice, 29,* 413-427.

American Psychological Association. (2004). Guidelines for psychological practice with older adults. *American Psychologist, 59,* 236-260.

Balsis, S. (2009). Measuring personality disorders in later life: hybrid criteria. Manuscript submitted for publication.

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging, 22(1),* 171-185.

Barendse, H. P. J., Thissen, A. J. C., Rossi, G., Oei, T. I., & Van Alphen, S. P. J. Psychometric properties of an informant personality questionnaire (the HAP) in a sample of older adults in the Netherlands and Belgium. *Aging and Mental Health, in press.*

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: perennial issues and an emerging reconceptualization. *Annual Review of Psychology,* 58, 227–57.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58,* 227-257.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI). Professional manual. Psychological Assessment Resources; Odessa, FL.

Derogatis, L. R. Bewerkers: W.A. Arrindell & J.H.M. Ettema.(2003) SCL-90 Symptom Checklist. Amsterdam, Pearson.

Klein, D. N. (2003). Patients' versus informants' reports of personality disorders in predicting 7 ½ year outcome in outpatients with depressive disorder. *Psychological Assessment, 15(2),* 216-222.

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9*, 300-311.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42,* 1879-1890.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2011). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42(9),* 1879-1890.

Kushner, S. C., Quilty, L. C., Tackett, J. L., & Bagby, M. (2011). The hierarchical structure of the dimensional assessment of personality pathology (DAPP-BQ). *Journal of Personality Disorders, 25(4),* 504-516.

Kushner, S. C., Quilty, L. C., Tackett, J. L., & Bagby, M. (2011). The hierarchical structure of the dimensional assessment of personality pathology (DAPP-BQ). *Journal of Personality Disorders, 25(4),* 504-516.

Lawton, E. M., Shields, A. J., & Oltmanns, T. F. (2011). Five-factor model personality disorder prototypes in a community sample: Self- and informant-reports predicting interview-based DSM diagnoses. Personality Disorders: Theory, Research and Treatment, 2(4), 279-292.

Markon, K. E., Krueger, R. F., Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139-157.

Markon, K. E., Krueger, R. F.,Watson, D. (2005). Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *Journal of Personality and Social Psychology, 88,* 139–57.

Mooi, B., Comijs, H. C., De Fruyt, F., De Ritter, D., Hoekstra, H., & Beekman, A. T. F. (2011). A NEO-PI-R short form for older adults. *International Journal of Methods in Psychiatric Research, 20(3),* 135-144.

Morey, L. C. (1991). Personality Assessment Inventory. Odessa, FL: Psychological Assessment Resources, Inc.

Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: Questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology, 7,* 321-349.

Oltmanns, T. F., & Balsis, S. Assessment of personality disorders in older adults. In Lichtenberg, P. A. (ed.). *Handbook of Assessment in Clinical Gerontoloy*. Academic/ Elsevier, 2010, p. 101-122.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43(4),* 295-312.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43(4),* 295-312.

Pincus, A. L. (2011). Some comments on nomology, diagnostic process, and narcissistic personality disorder in the DSM-5 proposal for personality and personality disorders. *Personality Disorders, 2,* 41-53.

Samuel, D. B. (2011). Assessing personality in the DSM-5: The utility of bipolar constructs. *Journal of Personality Assessment, 93(4),* 390-397.

Samuel, D. B. (2011). Assessing Personality in the DSM–5: The Utility of Bipolar Constructs, *Journal of Personality Assessment, 93(4),* 390-397.

Stuart, S., Simons, A., Thase, M. E., & Pilkonis, P. (1992). Are personality assessments valid in acute major depression? *Journal of Affective Disorders, 24,* 281-289.

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: Developmental

considerations for the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. *Development and Psychopathology, 21,* 687-713.

Van Alphen, S. P. J., Barendse, H., Tummers, J., & Rossi, G. (2010). Diagnostiek. In Van Alphen, S. P. J. (Ed.), *Persoonlijkheidsstoornissen bij ouderen: diagnostiek, behandeling en gedragsadvisering* (pp. 49-99). Hogrefe Uitgevers: Amsterdam.

Van Alphen, S. P. J., Derksen, J. J. L., Sadavoy, J., & Rosowsky, E. (2012). Editorial: Features and challenges of personality disorders in late life. *Aging & Mental Health, 16(7),* 805-810.

Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., Hoijtink, H. J. A., & Derksen, J. J. L. (2006). A preliminary study of the diagnostic accuracy of the Gerontological Personality Disorders Scale (GPS). *International Journal of Geriatric Psychiatry, 21,* 862-868.

Van den Broeck, J., Barendse, H. P. J., van Alphen, S. P. J., Thissen, T. & Rossi, G. (2012). Testdiagnostiek van persoonlijkheid en persoonlijkheidsstoornissen bij ouderen: een praktijkgerichte benadering. *Tijdschrift voor Klinische Psychologie, 42(1),* 24-32.

Widiger, T. A. (2011a). A shaky future for personality disorders. *Personality Disorders, 2,* 54-67.

Widiger, T. A. (2011b). The DSM-5 dimensional model of personality disorder: rationale and empirical support. *Journal of Personality Disorders, 25,* 222-234.

Widiger, T. A., & Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: A proposal for DSM-V. *Annual Review of Clinical Psychology, 5,* 197-220.

Widiger, T.A., & Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: A proposal for DSM-V. *Annual Review of Clinical Psychology, 5,* 197-220.

Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology.* Advance online publication. doi: 10.1037/a0027669.

Zweig, R. A. (2008). Personality disorder in older adults: assessment challenges and strategies. *Professional Psychology, Research and Practice, 39 (3), 298-305*.