



## UvA-DARE (Digital Academic Repository)

### Distinguishing between demoralization and specific personality traits in clinical assessment with the NEO-PI-R

Noordhof, A.; Sellbom, M.; Eigenhuis, A.; Kamphuis, J.H.

**DOI**

[10.1037/pas0000067](https://doi.org/10.1037/pas0000067)

**Publication date**

2015

**Document Version**

Final published version

**Published in**

Psychological Assessment

**License**

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care>)

[Link to publication](#)

**Citation for published version (APA):**

Noordhof, A., Sellbom, M., Eigenhuis, A., & Kamphuis, J. H. (2015). Distinguishing between demoralization and specific personality traits in clinical assessment with the NEO-PI-R. *Psychological Assessment*, 27(2), 645-656. <https://doi.org/10.1037/pas0000067>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

# Distinguishing Between Demoralization and Specific Personality Traits in Clinical Assessment With the NEO-PI-R

Arjen Noordhof  
University of Amsterdam

Martin Sellbom  
Australian National University

Annemarie Eigenhuis  
University of Amsterdam

Jan H. Kamphuis  
University of Amsterdam and the Viersprong Institute for  
Studies on Personality Disorders

Demoralization, a nonspecific unpleasant state that is common in clinical practice, has been identified as a potential source of nonspecificity in the assessment of personality and psychopathology. The aim of this research was to distinguish between Demoralization and specific personality traits in a widely used measure of personality: the Neuroticism–Extraversion–Openness Personality Inventory-Revised (NEO-PI-R). NEO-PI-R and Minnesota Multiphasic Personality Inventory-2 questionnaires were completed by 278 patients of a specialized clinic for personality disorders in The Netherlands. Furthermore, a replication sample was used consisting of 405 patients from the same institution who completed NEO-PI-R questionnaires, as well. A measure of Demoralization was derived (NEOdem, a NEO-PI-R-based Demoralization scale) using factor analytic techniques. Results indicated that the Demoralization Scale scores were reliable and showed expected patterns of convergence and divergence with conceptually relevant Minnesota Multiphasic Personality Inventory-2-RF scales. When items contributing to Demoralization-related variance were removed from the NEO-PI-R scales, increased specificity was notable with regard to external correlates. These results provide supportive evidence for the validity and heuristic potential of distinguishing between Demoralization and specific personality traits within the NEO-PI-R.

*Keywords:* NEO-PI-R, Demoralization, assessment, MMPI-2-RF

*Supplemental materials:* <http://dx.doi.org/10.1037/pas0000067.supp>

Demoralization has been advocated as an important concept in psychotherapy and psychotherapy research (Clarke & Kissane, 2002). In his influential *Persuasion and Healing: A Comparative Study of Psychotherapy*, Jerome Frank (1961) argued that many or even most clients in clinical practice are in a state of demoralization and that nonspecific features of psychotherapy are effective due to their remoralizing potential. If true, one would expect demoralization to be an important factor in clinical assessment, as well. Indeed, demoralization has been related to the large general factor that is often found when analyzing self-report instruments and has been identified as a potential source of nonspecificity in assessment (Tellegen, 1985). This application of the concept of demoralization to assessment has its roots in empirical research and theorizing on the structure of self-reported affect (see Watson

& Tellegen, 1985). As recently reviewed by Tellegen, Ben-Porath, and Sellbom (2009), psychological assessment and intervention research has identified Demoralization as a ubiquitous, affect-laden dimension among patients presenting with diverse physical and psychological problems in a variety of settings. Moreover, recent research has confirmed demoralization as an important construct undergirding dimensions of psychopathology. For example, based on structural equation modeling, Sellbom, Ben-Porath, and Bagby (2008) presented evidence supporting the construct validity of demoralization in the context of the recently proposed distinction between distress and fear disorders (Watson, 2005).

The presence of items that are substantially influenced by demoralization may cause several measurement problems. First, scales that contain many demoralization items are hard to interpret: Do high scores reflect demoralization or a specific personality trait? Second, artificially high correlations may emerge between scales that both contain demoralization-related items. Third, nonspecific correlations with external criteria, including psychopathological outcomes, may result. These issues are of particular importance in the context of clinical practice. A high level of demoralization is a common affective state accompanying different kinds of psychopathology and a likely reason for seeking professional help. If demoralization were to influence the measurement of multiple personality scales simultaneously, this could result in an overgeneralized negative view of a client's personality

---

Arjen Noordhof, Department of Psychology, University of Amsterdam; Martin Sellbom, Department of Psychology, Australian National University; Annemarie Eigenhuis, Department of Psychology, University of Amsterdam; Jan H. Kamphuis, Department of Psychology, University of Amsterdam, and Viersprong Institute for Studies on Personality Disorders.

Correspondence concerning this article should be addressed to Arjen Noordhof, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018XA, Amsterdam, The Netherlands. E-mail: a.noordhof@uva.nl

functioning. This is not to say, of course, that one would not expect problematic personality traits in those seeking professional help, nor that prolonged experiences of demoralization would be unrelated to personality traits. However, if these experiences of demoralization are “coloring” the assessment of personality in a nonspecific negative way, the resulting view of clients’ personalities may lack specificity, underestimate their specific strengths and overestimate difficulties. For these reasons, it would be useful to distinguish the assessment of demoralization from the assessment of specific personality traits (e.g., negative affectivity, extraversion). If this were successfully done, one would expect improved specificity of scales, attenuated intercorrelations between (within instrument) subscales, and improved specificity of correlations between personality scales and psychopathological outcomes.

Demoralization is related but conceptually distinct from the notion of a general factor of personality (GFP). GFPs are factors derived from the intercorrelations between personality traits. There are several ways to derive a GFP, which may partly explain the inconclusive evidence for its importance and meaning (e.g., Hopwood, Wright, & Donnellan, 2011; Loehlin, 2012; Revelle & Wilt, 2013). GFPs have been interpreted as artifacts by some (Ashton, Lee, Goldberg, & De Vries, 2009) and as a broad overarching personality trait by others (Rushton, Bons, & Hur, 2008). Demoralization by contrast is not a broad overarching factor directly *derived from* intercorrelations between personality traits, but rather a specific *explanation of* such intercorrelations (see online supplement Appendix A for a more complete comparison of GFP and demoralization). It is hypothesized to be an affective phenomenon of specific importance for clinical assessment and is derived from a theory of self-reported affect (Tellegen, 1985; Watson & Tellegen, 1985; Watson, Wiese, Vaidya, & Tellegen, 1999). It has been derived from the circumplex model of multiple self-report questionnaires, which distinguishes a dimension of positive affect (PA) and a dimension of negative affect (NA). Many items from questionnaires on affective experience are not specifically associated with either PA or NA, but rather with a combination of low PA and high NA (or vice versa), resulting in a large first factor in (unrotated) factor analyses of such questionnaires. According to Tellegen and colleagues (2003), the unpleasant pole of this first factor can be used to define demoralization; it indicates the nonspecific unpleasant experiences that are common in clinical practice. Items with high loadings on this first factor *and* low loadings on more specific factors are good markers of demoralization. Conversely, items with low loadings on the first factor and high loadings on a specific factor are good indicators of more specific personality traits or aspects of psychopathology (i.e., NA and PA).

These principles have been successfully applied in the development of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) restructured clinical (RC) scales. More specifically, Tellegen et al. (2003) observed that one major factor that contributed to excessive intercorrelations among the original MMPI-2 clinical scales was a nonspecific emotional distress factor. Using a series of principal components analyses, they extracted demoralization markers from the Clinical Scales 2 and 7, measures of depression and anxiety, respectively. Subsequently, they used these markers to extract demoralization, in order to isolate potential core constructs, from each of the eight original clinical scales. These procedures resulted in a Demoralization (RCd) Scale and specific scales reflecting the distinct core constructs (RC1–RC9)

of the clinical scales. Findings indicated that (a) intercorrelations between RC scales were substantially lower than those between the original clinical scales, and (b) correlations with psychopathological outcomes showed more specificity as indicated by improved divergence (i.e., lower correlations with outcomes that are not specifically related to a construct) and in some cases improved convergence (i.e., higher correlations with outcomes that are specifically related; Tellegen et al., 2003; see also Tellegen, Ben-Porath, & Sellbom, 2009, for a review). Moreover, of clinical relevance, it has been shown that the RC scales have great utility in interpreting MMPI-2 profiles in clinical practice, especially for those clients who show clinically significant elevations on many clinical scales (Kamphuis, Arbisi, Ben-Porath, & McNulty, 2008; Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006; Wygant & Fleming, 2008).

In the current study we investigated to what extent demoralization may influence scores on the Neuroticism–Extraversion–Openness Personality Inventory-Revised (NEO-PI-R). Demoralization might be especially influential in the MMPI-2, as it was constructed by empirical keying and allowed for item overlap across scales, both of which likely increase demoralization-related variance. Yet, we conceive of demoralization not as a methodological artifact (e.g., also in the MMPI-2, RCd is not a validity scale) specific to one instrument or construction method, but rather as a real affective phenomenon of important influence on item responses in self-report instruments. In a factor-analytically derived structure of (normal) personality traits, demoralization will probably be less influential than for the MMPI-2, but a substantial influence of demoralization can still be expected. The NEO-PI-R is based on explorative factor analyses, resulting in substantial cross loadings across scales. Simple structure is not to be expected for complex phenomena. Yet, a substantial number of items from multiple domains and facets of the NEO-PI-R explicitly refer to nonspecific unpleasant affective experience, which is the defining feature of demoralization. Accordingly, facets of these multiple domains of the NEO-PI-R (e.g., Positive Emotions, Depression, Competence) are strongly correlated with measures of demoralization (Sellbom, Ben-Porath, and Bagby, 2008). Particularly, items belonging to the domain of neuroticism appear related to demoralization, which is confirmed by finding a high correlation between Neuroticism and Demoralization ( $r = .62$ ; Sellbom et al., 2008). Furthermore, the NEO-PI-R is a self-report instrument that is often used in clinical practice to assess normal personality traits in order to draw inferences regarding maladaptive personality functioning or the influence of personality on current psychopathology (Widiger, Trull, Clarkin, Sanderson, & Costa, 1994). These assessments often take place during a period of marked demoralization. It seems a priori plausible that when completing an instrument with many items that contain demoralization-related variance, a client’s demoralized condition will be of substantial influence on his or her item responses. This assertion is consistent with, but is certainly not proven by, findings indicating important state effects of current psychopathological condition on NEO-PI-R scores (Karsten et al., 2012).

For these reasons, we aimed to investigate whether distinguishing between demoralization and specific personality traits within the NEO-PI-R is possible and useful by testing three hypotheses. First, we hypothesized that a substantial number of NEO-PI-R items from multiple scales are influenced by demoralization. Sec-

ond, we hypothesized that these items can be used to construct a reliable index of demoralization. Thus, we aimed to create a specific scale to measure the hypothesized phenomenon of demoralization. Such a scale subsequently allows to specifically test hypotheses regarding the influence of Demoralization. Third, we hypothesized that removing demoralization-related items from their original scales results in improved specificity in trait measurement. In other words, we aimed to identify “core constructs” for each of the NEO-PI-R domains and expected that these would show (a) lower intercorrelations among each other, (b) lower correlations with indices of nonspecific unpleasant affect, and (c) improved specificity with regard to indices of specific affect (NA or PA) and other psychopathological outcomes.

## Method

### Participants and Procedures

Data were collected from two samples from the Viersprong, a clinic for specialized treatment of personality disorders in The Netherlands. The first (“development”) sample was used for the construction and validation of revised NEO-PI-R indices. The second (“replication”) sample was used to replicate the development process. The development sample consisted of patients, who had been referred to a 7-week extensive diagnostic assessment ( $N = 201$ ) or to participate in one of several specific treatment programs involving schema-focused therapy ( $N = 197$ ) between 2005 and 2010. For 120 participants, it was not possible to obtain the relevant data (MMPI-2 and NEO-PI-R) for this study, so these were excluded. In the remaining sample ( $n = 278$ ), 122 (43.9%) were male, 156 (56.1%) female. The mean age of the respondents was 30.7 years ( $SD = 9.2$ ; range 18 to 60). No information on ethnicity was obtained, but a large majority of patients referred to this clinic are White. Diagnostic information based on the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)* Axis I Disorders (SCID-I) and Axis II Disorders (SCID-II) was obtained from the files of the participants, but was not available for 49 participants. Of the remaining 229 patients, 203 (88.6%) had received a *DSM-IV* Axis I diagnosis, while 116 (50.7%) had been diagnosed with a *DSM-IV* Axis II personality disorder (diagnostic details are given in Table 1).

The replication sample consisted of 405 consecutive prospective patients who had undergone an intake procedure at the Viersprong in 2009 or 2010, but who were referred to other treatment programs at the Viersprong, external treatment options, or to no treatment at all. Hence, there was no overlap between the two samples. From these participants, we were able to obtain NEO-PI-R information from electronic files, but unfortunately, no diagnostic information was available from these files. In this sample, 122 (43.9%) were male, 156 (56.1%) female. The mean age of the respondents was 34.2 years ( $SD = 10.8$ ; range 15 to 65; for 9 respondents, age was missing in the files).

Participants in both samples completed the standard intake assessment procedure at the Viersprong, which consists of two or three unstructured clinical interviews, a SCID-I and a SCID-II structured clinical interview, and a few self-report questionnaires, among which the NEO-PI-R. After the standard intake procedure, most patients are either directly referred to a specific treatment

Table 1  
*SCID-I- and SCID-II-Based Diagnostic Information for the Development Sample (N = 229)*

<i>DSM-IV</i> diagnosis	<i>N</i>	%
Axis I (SCID-I)	153	66.8
Mood disorder	93	40.6
Anxiety disorder	96	41.9
Substance-related disorder	30	13.1
Somatoform disorder	12	5.2
Eating disorder	18	7.9
Other Axis I	8	3.5
Axis II (SCID-II)	116	50.7
Paranoid PD	4	1.7
Schizoid PD	0	0
Schizotypal PD	0	0
Antisocial PD	2	0.9
Borderline PD	29	12.7
Histrionic PD	0	0
Narcissistic PD	9	3.9
Avoidant PD	58	25.3
Obsessive-compulsive PD	18	7.9

*Note.* *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*; SCID-I = Structured Clinical Interview for *DSM-IV* Axis I Disorders; SCID-II = Structured Clinical Interview for *DSM-IV* Axis II Disorders; PD = personality disorder.

program at the Viersprong or referred elsewhere. However, for some patients, a more extensive assessment is requested before final treatment selection. These patients take part in a 7-week intensive assessment program, which includes an assessment of the MMPI-2. Furthermore, some treatment options at the Viersprong include a standard extra assessment before treatment commences, which includes an assessment of the MMPI-2, as well. Thus, the MMPI-2 was always assessed after the NEO-PI-R, but times between these assessments varied substantially. The exact dates of the MMPI-2 administration were not retrievable, but given the logistic flow of the Viersprong assessment process, most likely all assessments took place between 3 weeks and 3 months after the NEO-PI-R administration.

### Materials

**NEO-PI-R.** All participants completed the Dutch version of the NEO-PI-R (Hoekstra, Ormel, & De Fruyt, 1996). This self-report instrument contains 240 items, which are rated on a 5-point Likert scale. These items measure 30 rationally constructed facets (eight items per facet), which each capture one aspect of a five-factor model domain (i.e., Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness; six facets per domain). For these five domain scale scores, previous work reported Cronbach’s alphas between .86 and .92 (Hoekstra et al., 1996) and for the development sample and replication sample very similar values were found ( $\alpha$  between .89 and .91).

**MMPI-2-RF.** The MMPI-2-RF is a 338-item self-report questionnaire with dichotomous response format (true/false). In this study, MMPI-2-RF scales were computed on the basis of the Dutch version of the 567-item MMPI-2 (Derksen, De Mey, Sloore, & Hellenbosch, 1997; which contains all MMPI-2-RF items). We used the RC scales (Tellegen et al., 2003), which have been extensively validated (see, e.g., Tellegen, Ben-Porath, & Sellbom,

2009, for a review) and for which we additionally calculated internal consistencies for the scale scores in the development sample. Apart from RCd (24 items;  $\alpha = .87$ ), these scales measure eight specific domains of psychopathology: Somatic Complaints (RC1; 27 items;  $\alpha = .80$ ), Low Positive Emotions (RC2; 17 items;  $\alpha = .74$ ), Cynicism (RC3; 15 items;  $\alpha = .79$ ), Antisocial Behavior (RC4; 22 items;  $\alpha = .74$ ), Ideas of Persecution (RC6; 17 items;  $\alpha = .72$ ), Dysfunctional Negative Emotions (RC7; 24 items;  $\alpha = .84$ ), Aberrant Experiences (RC8; 18 items;  $\alpha = .76$ ), and Hypomanic Activation (RC9; 28 items;  $\alpha = .80$ ). Furthermore, we used three higher-order scales, capturing the broad domains of Emotional/Internalizing Dysfunction (EID; 41 items;  $\alpha = .89$ ), Behavioral/Externalizing Dysfunction (BXD; 23 items;  $\alpha = .77$ ), and Thought Dysfunction (THD; 26 items;  $\alpha = .77$ ), which converge with dimensional higher-order models of psychopathology (e.g., Kotov et al., 2011). For more extensive information on the psychometric properties of these scales, we refer to the MMPI-2-RF technical manual (Tellegen & Ben-Porath, 2008/2011).

### Statistical Analyses

Analyses were done using SPSS, version 20.0.0 or *Mplus* version 7 (Muthén & Muthén, 2012). Furthermore, we used a Web-based computator (Hoerger, 2013) to compute Steiger's  $Z$  test (based on the  $-Z_1^*$ , Formula 14, in Steiger, 1980) and the psych package in R version 2.15.0 (Revelle, 2013; R Project, 2012) to compute McDonald's omega ( $\omega_h$ ; McDonald, 1999).

The methods employed in this study broadly correspond to the procedures used in developing the MMPI-2 RC scales (Tellegen et al., 2003), but differ in one important way. As emphasized in the introduction, we conceive of demoralization as a real phenomenon of general influence in many self-report questionnaires. Thus, we believe it can conceptually be applied to a broad range of self-report instruments. Yet, the specific statistical methods to do so have to be adapted to the specific instrument studied. The crucial difference in this case is that the NEO-PI-R, in contrast with the MMPI-2, is based on a facet-based factor analytic structure. In the current study we take this often replicated complex and heterogeneous structure as a given. Therefore, we will often assume the existence of six facets within a domain. Also, we accept substantial heterogeneity will remain within each broad domain after our restructuring efforts and do not aim for a more complete restructuring process as was done for the MMPI-2.

First, we constructed a NEO-PI-R Demoralization Index (NEOdem) within the development sample, following a three-step procedure. In Step 1, we used exploratory bifactor analysis to select a starting set of good demoralization markers. An exploratory bifactor model (Jennrich & Bentler, 2011) is an exploratory factor analysis for which a specific rotation method is applied by which the first general factor (G) is not rotated (i.e., represents the first unrotated factor), while all subsequent factors are rotated according to a regular rotation method (in this case GEOMIN, which is the default option in *Mplus*). As described in the introduction, demoralization was conceptualized as nonspecific unpleasant affect, which can be distinguished from specific affects. Therefore, we selected four facets (N1: Anxiety, N3: Depression, E1: Warmth, E6: Positive Emotions) that conceptually contain indicators of positive, negative and nonspecific unpleasant affect. It was expected that the 32 items of these scales would show a large

general factor (G) of nonspecific unpleasant affect as well as four factors related to more specific features of each of the selected facets. Items with a high loading on G ( $>.5$ ) and lower loadings on the specific factors were selected as Demoralization markers.

In Step 2, we selected candidate Demoralization items from the full NEO-PI-R item pool. To this end, we used exploratory structural equation modeling (ESEM) with target rotations. Detailed descriptions of this technique can be found elsewhere (e.g., Asparouhov & Muthén, 2009). In general terms, the technique can be understood as an exploratory approach to factor analysis. However, factor rotations are based on a set of informed a priori expectations about loadings (i.e., target rotations) rather than a fully exploratory rotational approach. Our rationale for using this technique is that the exact measurement structure of broadband personality measures is generally not sufficiently clear to allow for strong confirmatory approaches (Marsh et al., 2010), but at the same time, we did have specific a priori expectations about factorial structure. For each of the five NEO-PI-R domains, we fitted an exploratory structural equation model based on the 48 items of that domain together with the items that we selected in Step 1 as core markers of demoralization. We assumed that an eight-factor structure would be adequate for each of these analyses: one factor capturing demoralization-related variance, one factor capturing domain-related variance (e.g., extraversion), and six factors capturing the facet-specific variance. In other words, for each item, we expected loadings on Demoralization and its corresponding domain and facet. Target loadings were chosen based on these expectations.

In Step 3, we made a final selection of items to be included in the NEOdem. To ensure reliability, the items should mainly reflect one general Demoralization factor. As the ESEM analyses were conducted for each domain separately, these analyses did not guarantee a homogeneous NEOdem. Therefore, we used exploratory bifactor modeling on all candidate Demoralization items, expecting a large general factor and a specific factor for each five-factor model domain from which items were taken. Those items that loaded substantially higher on the specific factors than on the general factor were excluded from the final index. A second reason for exclusion of an item was that statistical selection of items does not fully guarantee that the content of the items matches the theoretical concept of demoralization. As such, the authors discussed informally whether the item content diverged from the concept of Demoralization, as described in the introduction.

Subsequently, revised NEO-PI-R scales were constructed by removing those items that were deemed to better reflect Demoralization than the specific trait. The results of the target-rotated ESEMs (see Step 2) were used to select items for removal. We investigated reliabilities of these revised scales and tested whether intercorrelations among these scales were attenuated in comparison with the original scales. Furthermore, zero-order correlations with MMPI-2-RF scales were used to examine the convergent and discriminant validity associated with our newly constructed measures. Steiger's  $Z$  test for the difference between two correlations was used to test whether these zero-order correlations significantly differed between the original and revised scales.

To test for robustness of our results, this development procedure was duplicated in the replication sample using the exact same steps. In the Results section, we report at each step to what extent results were replicated. Complete results of these replication anal-

yses can be obtained from the authors upon request. Finally, we did a set of post hoc explorative analyses in order to compare demoralization with GFPs. For these analyses, GFPs were derived using three methods that have often been used in previous studies (see Loehlin, 2012). Due to space constraints, we only present a brief summary of these analyses in the present paper (full results are provided as online supplementary material Appendix A).

For all significance tests, the level of significance was set at  $\alpha = .01$  in order to decrease chance findings. In both the bifactor models and exploratory structural equation models, we primarily relied upon the root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), and standardized mean square residual (SRMR) to evaluate model fit. Rules of thumb for the interpretation of these indices are RMSEA values below .06 and SRMR values below .08 indicate adequate fit (Hu & Bentler, 1999). CFI values above .90 are often considered acceptable (Byrne, 2011; Kline, 2011; Little, 2013), but according to Hu and Bentler (1999), a cutoff at .95 is preferable.

Reliability was examined by estimating Cronbach's alpha and McDonald's omega (McDonald, 1999). McDonald's omega has been advocated as more informative than Cronbach's alpha (Revelle & Zinbarg, 2009), especially for heterogeneous scales, and is an indicator of general factor saturation, which can be interpreted as the precision with which scale scores estimate a latent variable common to all test items. The estimations of  $\omega_h$  are based on exploratory factor analyses with a Schmid-Leiman transformation (e.g., Revelle & Wilt, 2013).<sup>1</sup>

## Results

### Construction of a Demoralization Index

**Step 1: Core markers of demoralization.** First, we aimed to select a small set of Demoralization markers to use in the analyses of the complete NEO-PI-R item pool (Step 2). The bifactor model showed a mediocre fit (RMSEA = .063, CFI = .88, SRMR = .044). We selected as core markers of demoralization items with a high loading on the general factor (loading on G > .5) and lower loadings on the four specific factors (loadings are reported in Table 2). Upon inspection of the content of these 12 items, the authors uniformly agreed that they indeed represented good indicators of Demoralization, as defined in the introduction.

Robustness of these analyses was shown in the replication sample, in which a very similar bifactor model fitted the data (RMSEA = .056, CFI = .91, SRMR = .036) and resulted in the selection of similar, albeit not exactly the same, items (nine out of 12 items were replicated; three were not due to small differences in loadings).

**Step 2: Selecting potential demoralization items.** As described in the introduction, it was hypothesized that a subset of the 240 NEO-PI-R items would be better considered as indicators of demoralization than as indicators of the factor or facet to which they were originally assigned. To select those items, we used the technique of ESEM with target rotations as described earlier. Fit indices for these analyses are reported in Table 3. RMSEA and SRMR indicated adequate model fit (RMSEA < .06; SRMR < .05). Although CFIs were not satisfactory (CFI < .90), we nevertheless judged model fit to be acceptable for using these models for the purpose of selecting items.<sup>2</sup> Subsequently, we used the loading

Table 2  
*Standardized Loadings in the Bifactor Analysis on Four NEO-PI-R Facets*

Item	Fg	Fs1	Fs2	Fs3	Fs4	Marker
1	.27	-.06	.03	-.06	.19	
31	.45	.51	-.18	-.03	.04	
61	.45	.43	.14	.05	.02	
91	.43	.33	.00	.04	.03	
121	.44	.13	.14	.02	-.02	
151	.61	.40	.21	.00	-.05	X
181	.38	.44	-.06	.00	.08	
211	.51	.40	.04	.10	.01	X
11	.66	.11	.28	-.11	.01	X
41	.55	-.29	-.03	-.02	.37	X
71	.51	.05	.40	-.08	-.02	X
101	.18	.08	.09	.12	.09	
131	.51	.00	.00	.06	.29	X
161	.67	-.10	-.04	-.02	.17	X
191	.65	.03	.34	.06	-.01	X
221	.55	.07	.10	-.06	.04	X
2	-.23	-.18	.16	.01	.20	
32	-.43	.01	.77	-.09	-.09	
62	-.20	-.12	.21	.08	.27	
92	-.05	-.09	.19	.05	.13	
122	-.36	-.10	.64	.07	.08	
152	-.53	.05	.54	-.03	.02	
182	.01	-.02	.25	.31	.13	
212	-.05	-.01	.32	.10	.25	
27	-.24	-.09	.06	.71	-.01	
57	-.19	.04	-.09	.79	-.04	
87	-.66	.07	-.03	-.08	.44	X
117	-.40	.07	-.02	.55	.02	
147	-.74	.08	-.02	-.01	.47	X
177	-.66	.06	.11	-.02	.47	X
207	-.29	.06	.09	.35	.04	
237	-.30	.12	.29	.13	.20	

*Note.* NEO-PI-R = Neuroticism–Extraversion–Openness Personality Inventory-Revised; Fg = general factor, Fs1–Fs4 = specific factors; Marker = items that were selected as core markers of Demoralization.

patterns of the analyses to select items that were strongly loading on Demoralization (>.40 as a conventional criterion of meaningful loading) and less strongly on the other factors. Such a loading pattern suggests that the item is more indicative of demoralization than of its original domain or facet. Twenty-seven items were selected this way, all from the domains of Extraversion (seven), Neuroticism (16), and Conscientiousness (four).

<sup>1</sup>  $\omega_h$  has to be estimated from an exploratory factor model with a specified number of factors. For the NEOdem Index, we chose a three-factor solution, because items were sampled from three domains. For NEO-PI-R domains, we chose six-factor solutions, because of its facet structure. If these models would show inadequate model fit, we also estimated  $\omega_h$  from the first model for which RMSEA was below .05 and checked whether this resulted in a lower  $\omega_h$ , which was never the case.

<sup>2</sup> The rationale behind these analyses is that we follow the original NEO-PI-R structure of one domain with six facets. This resulted in CFIs below .90, which may indicate inadequate model fit. However, when we did similar analyses based on the original model of only one domain and six facets (i.e. without demoralization and with the 48 items originally assigned to that domain), CFIs were also suboptimal and almost equal to our results. Therefore, we have concluded that the suboptimal model fit did probably *not* result from the inclusion of a demoralization factor and, hence, did not invalidate our procedure for selecting items, per se.

Table 3  
Fit Indices for the Exploratory Structural Equation Models With Target Rotation

Domain	RMSEA		CFI		SRMR		No. of Items		
	Con	Rep	Con	Rep	Con	Rep	Con	Rep	Both
Neuroticism	.057	.046	.84	.91	.044	.033	16	23	16
Extraversion	.056	.045	.85	.90	.040	.032	7	7	7
Openness	.053	.044	.80	.86	.046	.037	0	1	0
Agreeableness	.052	.036	.85	.92	.043	.031	0	1	0
Conscientiousness	.047	.038	.88	.92	.038	.030	4	4	3

Note. RMSEA = root mean square error of approximation; CFI = Comparative Fit Index; SRMR = standardized mean square residual; Con = construction sample; Rep = replication sample; Both = number of items that were selected in both the construction and replication sample.

Robustness of these analyses was tested in the replication sample. The exploratory structural equation models showed slightly improved fit indices (see Table 3). All but one of the items that were selected in the development sample were also selected in the replication sample. However, in the replication sample more items (36 instead of 27 items) were selected, mainly due to somewhat higher loadings of Neuroticism items on the Demoralization factor.

**Step 3: Selecting the final set of items for NEOdem.** We excluded nine items from the 27 candidate items for the NEOdem Index selected in Step 2 in view of the following considerations. The bifactor model with three specific factors (i.e., one for each domain from which items were included) showed mediocre to adequate model fit (RMSEA = .069, CFI = .89, SRMR = .045). For six of the 27 candidate items, the loadings on the general factor were substantially lower (a difference in loadings larger than .20) than on the specific factors; these items were removed from the final item set. Furthermore, the authors were in uniform agreement that three items (#12, 25, and 146) did not qualify from a content validity perspective. Therefore, the final NEOdem Index contains 18 items (see Table 4).

Robustness of these analyses was tested in the replication sample. In this sample, a similar bifactor model (RMSEA = .056, CFI = .91, SRMR = .038) was fitted on the 36 items that were selected in the replication on Step 2. Of the 36 items, only three were removed based on statistical criteria. We did not remove items on the basis of item content in this replication sample, so the NEOdem Index in the replication sample contained 33 items. Of the 18 items that were selected in the development sample, 17 were also selected in the replication sample.

### Removing Demoralization-Related Variance

The central idea underlying this study was that the interpretation of NEO-PI-R test results can be improved by distinguishing between demoralization-related variance and trait-specific variance. To test this hypothesis, we did not only create the NEOdem, as described earlier, but also aimed to remove demoralization-related variance from the original NEO-PI-R scales. Hence, from each domain scale (48 items), we removed items contributing to demoralization-related variance by selecting them on the basis of their loading patterns in the ESEM analysis with target rotations (Step 2). We removed items for which the loading on the Demoralization factor was substantial ( $>.30$ ).

For Openness to Experience none of the items was removed, indicating that this scale is not substantially influenced by Demor-

alization. For Neuroticism, 30 items were removed, resulting in an 18-item revised scale (N\_R). For Extraversion, 19 items were removed, resulting in a 29-item scale (E\_R). For Agreeableness, seven items were removed, resulting in a 41-item scale (A\_R). For Conscientiousness, 12 items were removed, resulting in a 36-item scale (C\_R).

The robustness of these analyses was tested in the replication sample by constructing revised indices applying exactly the same criteria (see Table 5). For most indices, the differences were minor and due to small differences in loadings and the inherent arbitrariness of using a cutoff at a loading of .30. For Neuroticism, the differences were somewhat larger, owing to the fact that Neuroticism items tended to show somewhat higher loading on Demoralization in the replication sample. This resulted in a smaller (13-item) N\_R index in the replication sample.

Table 4  
Items Selected for the NEOdem Index

Item	Facet	Description of content	NEOdem
151	Nanx	Worried	+ <sup>a</sup>
211	Nanx	Fearful thoughts	+
11	Ndep	Lonely or sad	+
41	Ndep	Worthless	+
71	Ndep	Sad or depressed	+
131	Ndep	Self-blame	+
191	Ndep	Sad and hopeless	+
221	Ndep	Easily disheartened	+
87	Epem	Not cheerful optimist	+
147	Epem	Not cheerful and happy	+
26	Nvul	Helpless	+
56	Nvul	Able to solve problems myself	-
86	Nvul	Feel like breaking down	+
236	Nvul	Emotionally stable	-
155	Ccom	Not successful in anything	+
130	Cord	Not able to get things under control	+
115	Cdis	Difficulty to do what needs to be done	+
205	Cdis	Don't want to know all things that need to be done	+

Note. NEOdem = Demoralization scale based on the Neuroticism–Extraversion–Openness Personality Inventory–Revised; Nanx = Neuroticism–Anxiety; Ndep = Neuroticism–Depression; Epem = Extraversion–Positive Emotionality; Nvul = Neuroticism–Vulnerability; Ccom = Conscientiousness–Competence; Cord = Conscientiousness–Order; Cdis = Conscientiousness–Discipline.

<sup>a</sup> To obtain NEOdem scores items with a + should be summed, while those with a - should be reverse coded.

Table 5  
Number of Items and Reliabilities for Reconstructed  
NEO-PI-R Scales

Scale	No. of Items			$\alpha$	$\omega_h$
	Con	Rep	Both		
NEOdem	18	33 <sup>a</sup>	17	.89	.71
N_R	18	13	11	.77 (.91 <sup>b</sup> )	.43 (.56)
E_R	29	25	24	.79 (.91)	.48 (.60)
O <sup>c</sup>	48	47	47		
A_R	41	39	37	.85 (.89)	.53 (.51)
C_R	36	32	32	.89 (.91)	.56 (.58)

Note. NEO-PI-R = Neuroticism–Extraversion–Openness Personality Inventory-Revised; Con = construction sample; Rep = replication sample; Both = number of items that were selected in both the construction and replication sample; NEOdem = NEO-PI-R-based Demoralization scale; N\_R = Neuroticism Revised; E\_R = Extraversion Revised; O = Openness; A\_R = Agreeableness Revised; C\_R = Conscientiousness Revised. <sup>a</sup> These are all statistically selected items (step 3 of the analysis). For the replication sample, we did not remove items on the basis of item content. <sup>b</sup> Reliability coefficients for the original scale scores are reported in parentheses. <sup>c</sup> For O, no reconstructed scales were created.

### Reliability

The 18-item NEOdem scale scores (based on the results in the development sample) were associated with good internal consistency ( $\alpha = .89$ ) and a satisfactory general factor saturation ( $\omega_h = .71$ ). The revised indices also showed adequate to good internal consistencies ( $\alpha$  between .76 and .89; see Table 5), but lower general factor saturations ( $\omega_h$  between .43 and .56). Relatively low general factor saturations are also found for the original scales ( $\omega_h$  between .51 and .60; see Table 5) and indicate that substantial heterogeneity exists within the broad five-factor model domains, as can be expected from its facet structure.

### External Correlates

First, we computed correlations between the NEOdem Index and conceptually relevant MMPI-2-RF scales. These are reported in Table 6. As evident from this table, correlation patterns for the NEOdem Index were highly consistent with a priori expectations: (a) large correlations with scales reflecting nonspecific unpleasant affect; (b) lower, but yet substantial, correlations with other indices of affect; and (c) weak correlations with nonaffective indices.

Second, we tested whether the original NEO factors were correlated with measures of nonspecific affect and whether these correlations would attenuate after removing Demoralization-related items. As evident in Table 6, all NEO-PI-R domain scales, except Openness to Experience, were correlated with MMPI-2-RF RCd and with MMPI-2-RF EID. After removing the items that were most strongly related to demoralization, these correlations were significantly attenuated (Steiger's  $Z$  test, all  $p < .01$ ) for both RCd and EID.

Third, we tested whether intercorrelations between NEO factors would attenuate after removing demoralization-related variance (see Table 7). This was the case for the correlation between Neuroticism and Extraversion, which showed a moderately negative correlation in the original scales and a nonsignificant (positive) correlation after removing demoralization. However, the correlation between Neuroticism and Conscientiousness did not attenuate and the correlation between Neuroticism and Agreeableness increased, indicating that the specific variances associated with these domains are related for reasons beyond demoralization.

Fourth, we tested the hypothesis that removing demoralization variance would result in increased affective specificity for revised NEO-PI-R domain scales. We expected to find that N\_R and E\_R would show more divergent correlational patterns relative to the original Neuroticism and Extraversion domain scales with regard to specific affective indicators of the MMPI-2-RF. The original

Table 6  
Pearson Correlations of (Revised) NEO-PI-R Scales With MMPI-2/MMPI-2-RF Scales

MMPI-2-RF scales	NEOdem	Neuroticism		Extraversion		Agreeableness		Conscientiousness	
		Orig <sup>1</sup>	Con	Orig	Con	Orig	Con	Orig	Con
EID	.65	.70	.43 <sup>†</sup>	-.45	-.25 <sup>†</sup>	-.15	-.08 <sup>†</sup>	-.23	-.11 <sup>†</sup>
RCd	.66	.64	.40 <sup>†,a</sup>	-.30	-.12 <sup>†</sup>	-.19	-.12 <sup>†</sup>	-.32	-.20 <sup>†</sup>
RC2	.42	.43	.17 <sup>†</sup>	-.56	-.41 <sup>†</sup>	.03	.08 <sup>†</sup>	-.09	.01 <sup>†</sup>
RC7	.43	.61	.45 <sup>†</sup>	-.30	-.16 <sup>†</sup>	-.26	-.19 <sup>†</sup>	-.14	-.07 <sup>†</sup>
BXD	.12	.19	.35 <sup>†</sup>	.25	.33 <sup>†</sup>	-.38	-.38	-.36	-.38
THD	.12	.26	.30	-.02	-.02	-.16	-.12 <sup>†</sup>	-.05	-.05
RC1	.27	.39	.27 <sup>†</sup>	-.13	-.04 <sup>†</sup>	-.08	-.02 <sup>†</sup>	.01	.05
RC3	.18	.23	.31	-.06	-.04	-.45	-.39 <sup>†</sup>	-.22	-.19
RC4	.18	.25	.36 <sup>†</sup>	.10	.19 <sup>†</sup>	-.30	-.28	-.40	-.39
RC6	.17	.34	.39	-.06	-.04	-.24	-.19	-.11	-.09
RC8	.14	.27	.30	-.03	.00	-.16	-.13	-.09	-.08
RC9	.07	.13	.34 <sup>†</sup>	.39	.42	-.50	-.51	-.20	-.21

Note. Critical values for significant correlations ( $p < .01$ ) are  $r < -.16$  and  $r > .16$ ;  $N = 278$ . NEO-PI-R = Neuroticism–Extraversion–Openness Personality Inventory-Revised; MMPI-2 = Minnesota Multiphasic Personality Inventory-2; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2–Response Format; NEOdem = NEO-PI-R-based Demoralization scale; Orig = original scales in construction sample; Con = revised scales in construction sample; EID = Emotional/Internalizing Dysfunction, RCd = Demoralization, RC1–RC9 = Revised Clinical Scales, BXD = Behavioral/Externalizing Dysfunction, THD = Thought Dysfunction. <sup>a</sup> Correlation was lower ( $r = .34$ , 99% confidence interval [.22, .45]) when using indices created in the replication sample.

<sup>†</sup> A significant difference ( $p < .01$ ) with the original scale, based on Steiger's  $Z$  test.

Table 7  
Pearson Correlations and 99% Confidence Intervals Between Original and Restructured  
NEO-PI-R Scales

Domain	NEOdem		Neuroticism		Extraversion		Agreeableness	
	Orig	Con	Orig	Con	Orig	Con	Orig	Con
Neuroticism	.81	.43						
Extraversion	-.42	-.17	-.34	.14 <sup>a</sup>				
Agreeableness	-.14	-.05	-.20	-.34	.03	-.06		
Conscientiousness	-.40	-.19	-.29	-.26	.05	-.15	.16	.15

Note. Critical values for significant correlations ( $p < .01$ ) are  $r < -.16$  and  $r > .16$ ;  $N = 278$ . NEO-PI-R = Neuroticism–Extraversion–Openness Personality Inventory–Revised; NEOdem = NEO-PI-R-based Demoralization scale; Orig = original scales in construction sample; Con = revised scales in construction sample.

<sup>a</sup> Correlation was higher and significant when using indices created in the replication sample:  $r = .20$  ( $p < .01$ ).

Neuroticism scale correlated with both Low Positive Emotions (RC2) and Dysfunctional Negative Emotions (RC7). After removing demoralization, a significant decrease ( $Z = -6.255$ ,  $p < .01$ ; see Table 6) of the correlation between Neuroticism and RC2 was observed. The correlation with RC7 also decreased significantly ( $Z = 4.392$ ,  $p < .01$ ), but remained substantial. The original Extraversion scale correlated negatively with both RC2 and RC7; the removal of demoralization led to a significantly ( $Z = -5.422$ ,  $p < .01$ ) attenuated correlation with RC7. The correlation with RC2 also attenuated ( $Z = -5.674$ ,  $p < .01$ ), but remained substantial. In sum, after removing demoralization, Neuroticism and Extraversion became more divergent with regard to their respective associations with RC7 and RC2. However, contrary to what we initially expected, both indices showed decreased convergence with these specific affective indices (i.e., attenuated correlations).

Fifth, we used the correlation results (see Table 6) to explore what variance was left in each domain scale after removing demoralization. It appeared that N\_R was correlated with measures of both internalizing pathology (EID, RC7, RC1), externalizing pathology (BXD, RC4, RC9) and thought problems (THD, RC6, RC8). E\_R showed more specific correlation patterns, with low positive emotions (RC2) and hypomanic activation (RC9), but also with externalizing dysfunction (BXD). C\_R, in contrast to the original scale, appeared as a clearly nonaffective personality scale correlating specifically and negatively with indicators of externalizing pathology (BXD, RC4, RC9). The correlates of A\_R did not appear substantially different from those of the original scale.

To test for robustness, we also computed zero-order correlations between the MMPI-2-RF and the indices as derived from the replication analyses. As described above, these indices showed some differences with those developed in the development sample (see Table 5). In Tables 6 and 7, we reported those instances where the results were somewhat different (difference in  $r > .05$ ); these differences were minimal.

### Comparing Demoralization With General Factors of Personality

We expected GFPs to be significantly associated with Demoralization, but to be broader and more heterogeneous constructs. Higher heterogeneity of GFP relative to Demoralization was supported by finding a much lower value for McDonald's omega ( $\omega_h = .34$  vs.  $\omega_h = .71$  for NEOdem). Furthermore, the analyses

showed that two ways of deriving GFP, both of which are often used in the literature, resulted in a GFP that was substantially correlated to NEOdem. However, both were even more strongly correlated to a sum score of all NEO-PI-R items. These findings support the notion that GFP is a broad and heterogeneous factor encompassing many aspects of personality as measured by the instrument from which it is derived. Within the NEO-PI-R these GFPs seem significantly associated with, but not isomorphic to, Demoralization. However, a third way of deriving GFP (i.e., iterations of principle component analyses on subfactors until only one general factor with eigenvalue  $>1$  results, see Hopwood, Wright, & Donnellan, 2011) resulted in a GFP that was very specifically and very strongly associated with Extraversion ( $r = .91$ ) but only moderately associated with Demoralization. This finding illustrates the important influence that the specific method for constructing GFP has on the results obtained, or stated differently, how different methods, all of which have been used in the literature, can lead to substantively different versions of GFP.

### Discussion

The major aim of this investigation centered on differentiating demoralization and more specific personality traits within NEO-PI-R measurement of the five-factor model domains. Our findings are consistent with the hypotheses that (a) several NEO-PI-R domains include substantial demoralization variance, (b) NEO-PI-R items can be used to construct a reliable index of demoralization (NEOdem), and (c) removing demoralization-related items results in increased specificity in trait measurement. Taken together, we believe these results provide supportive evidence for the validity and utility of differentiating demoralization from specific personality traits in the five-factor model domains.

The conclusion that multiple NEO-PI-R domains include demoralization variance was supported by the finding (Step 2 of the analysis) that a large number of items loaded substantially on a factor that was "anchored" on core markers of demoralization. Furthermore, the set of items with primary loadings on Demoralization was large enough to construct a reliable index of demoralization (i.e., NEOdem). Correlations with MMPI-2-RF RC scales supported the construct validity of these indices. More specifically, they exhibited convergence with RC Demoralization and divergence from most of the specific RC scales, except for RC2 and RC7, as expected. This substantial association with both Low

Positive Emotions (RC2) and Dysfunctional Negative Emotions (RC7) is consistent with interpreting Demoralization as the unpleasantness vector in Tellegen, Watson, and colleagues' (e.g., Tellegen, Watson, & Clark, 1999; Watson & Tellegen, 1985) mood model.

The second part of our analysis resulted in specific indices for neuroticism, extraversion, agreeableness and conscientiousness upon removal of demoralization-related content. Each of these indices showed attenuated correlations with *nonspecific* unpleasant affect (RCd), and improved divergence with regard to *specific* affective problems; more specifically, the revised Neuroticism scale correlated specifically with Dysfunctional Negative Emotions, but not Low Positive Emotions, whereas the Extraversion scale showed the opposite pattern. Furthermore, an attenuated correlation was found between Extraversion and Neuroticism, showing that demoralization may indeed partly account for the negative association between these two domains. However, unexpectedly, attenuated correlations were not found between the other scales. This result shows that demoralization may be a partial but cannot be a complete explanation for finding a GFP in factor analytic explorations of the NEO-PI-R scales. This is in line with the findings from the additional analyses (online Appendix A), which show that demoralization is related to, but not equivalent to, broader and more heterogeneous GFPs. What these analyses also illustrate, as has been shown in several earlier studies (e.g., Hopwood, Wright, & Donnellan, 2011; Loehlin, 2012; Revelle & Wilt, 2013), is that the substantive meaning of GFP is, in large part, a function of the specific method used to construct it.

To our knowledge, this is the first study to show that the strategy of distinguishing demoralization from specific traits can be applied to a different personality inventory than the MMPI-2. An important yield of this strategy is that it allows for the exploration of remaining variance in a domain after items contributing to demoralization-related variance have been removed. With regard to Neuroticism, the original scale correlated .81 with NEOdem, with the implication being that it may be difficult to differentiate between Neuroticism and Demoralization. However, the results also show that the NEO-PI-R Neuroticism domain is *not* equivalent to Demoralization; indeed, the revised Neuroticism scale only exhibited a moderate correlation with NEOdem and was associated with a different pattern of extratest correlates relative to NEOdem. The substantial correlations of this scale with many MMPI-2-RF RC scales indicate that the scale is associated with dysfunctional negative emotions, externalizing behaviors, and thought disorder. Inspection of item content allowed us to infer that it consists mainly of items from the Angry Hostility and Impulsiveness facets; as such, this revised measure appears more specific to alienation and aggression components of negative affectivity (e.g., Tellegen & Waller, 2008) that would be reflected in a range of psychopathological conditions (e.g., Sellbom & Ben-Porath, 2005) in a way that is different from nonspecific unpleasant affect. With regard to Extraversion, the revised scale was negatively correlated with Low Positive Emotions and positively with externalizing proclivities, indicating that it is a domain of positive emotionality, energy and sensation seeking. This is consistent with research on positive affectivity being linked to reward responsivity/approach motivation (Depue & Collins, 1999), which has implications for well-being, social engagement, and impulsive behavior. With re-

gard to Conscientiousness and Agreeableness, the results do not imply major changes in scale interpretation.

We also think that the development of the NEOdem Index (which can be computed using the information in Table 4) is itself potentially useful. Some readers may doubt the wisdom of creating a new scale in an existing instrument of 30 facets and five domains. Is this variance not already captured well enough within the hierarchical structure of the original instrument? We think not. It is our hypothesis that interindividual differences in demoralization produce covariance between items and facets from multiple domains. This phenomenon is, in our perspective, factor-analytically masked by the assumption that each item belongs to one facet and each facet to one domain, which may be a too strict hierarchical assumption. In the current research, we showed some, but certainly no conclusive, evidence for this position. In future research, the NEOdem scale can be used to test whether this hypothetical construct is indeed useful in improving our understanding of NEO-PI-R-based results. For example, it can be hypothesized that a substantial part of the association between measures of personality and psychopathology (e.g., Ormel et al., 2013) can be understood as reflecting interindividual differences in demoralization. The NEOdem Index can be readily used as an instrument to test this idea, for example by introducing it as a covariate in regression analyses. Such analyses may prove useful in distinguishing ways in which personality measures predict negative outcomes. For example, the current study suggests that there are at least two ways in which Neuroticism is associated with psychopathology: as an indicator of nonspecific unpleasant affect, but also as a measure of negative affectivity centered on stress reactivity, alienation, and aggression. Furthermore, it can be hypothesized that the effects of both psychopathology (e.g., Karsten et al., 2012) and treatment (e.g., Quilty, Meusel, & Bagby, 2008; Tang et al., 2009) on measures of personality are mainly due to their effects on demoralization, which would be in line with the theory that demoralization is a common feature of psychopathological conditions and that remoralization is a common feature of treatments (Frank, 1961). The NEOdem Index could be used to test this by introducing it as a mediating variable in the relation between change in psychopathological condition or treatment and change in personality measures. Moreover, the NEO-PI-R is likely not the only instrument with demoralization-related items. Indeed, the strategy applied in the current paper, which was derived from but not equal to the procedures used to restructure the MMPI-2, may prove useful for other instruments as well in order to meet the challenge of *not* measuring demoralization (Tellegen, 1985). Specifically, the hypothesis of demoralization would be most relevant for self-report instruments that (a) tap a presumably related measurement domain, (b) contain many items directly related to unpleasant affective experiences, and (c) are routinely used in clinical practice. For example, a dimensional approach to personality disorders was recently included as an optional alternative system in *DSM-5* (American Psychiatric Association, 2013); insofar as self-report instruments are used to assess these dimensions, the results might be influenced by demoralization, as well.

Finally, the results indicate that, while removal of demoralization resulted in greater specificity of scales with regard to internalizing problems (EID, RCd, RC2, RC7), it did not (and should not) result in more specificity with regard to externalizing problems (BXD, RC4, RC9). On the contrary, Neuroticism and Extra-

version both showed *increased* correlations with externalizing outcomes. These increased correlations are consistent with interpreting Demoralization as a source of *affective* nonspecificity and furthermore indicate that it may not be the only source of nonspecificity in NEO-PI-R domains. A hypothesis about the common theme underlying the revised scales is that they all involve regulation of emotions (e.g., stress reactivity, aggression) and impulses (e.g., approach motivation). This hypothesis might be further tested by identifying a set of core markers of this common feature and using a similar strategy as the one followed in the current study.

This project is associated with several limitations that warrant discussion. First, the results in the replication sample were not completely equivalent. Several items showed higher loadings on Demoralization than in the original sample, which may be due to more homogeneity in the development sample, and thus, reduced variability. The development sample consisted exclusively of patients assigned to a selected set of treatments and hence can be expected to be more homogeneous than the replication sample, which included all patients seen at intake.<sup>3</sup> However, this does not invalidate the overall conclusions of this paper, as in both samples, all three hypotheses were clearly supported. The final NEOdem Index consists of items that were identified in both samples, and thus should be viewed as robust markers for this construct. A second limitation is that the CFI fit indices were somewhat lower than what is typically viewed as acceptable for the ESEM-analyses in the development sample (Step 2 of the analyses); this limitation was probably related to problems in fitting the original NEO-PI-R structure, as explained in Footnote 1, rather than being specific to our analytical approach. Furthermore, several authors have argued that the CFI may underestimate model fit in the case of factor analyses with a large amount of item data, as this index is based on the proportion of improvement relative to a null model in which interitem correlations are fixed to zero, and item level correlations tend to be small, to a degree where there is not much room for improvement (Marsh, Hau, & Grayson, 2005). This becomes particularly important when the potential for range restriction also exists, which is likely more the case for the development sample rather than replication sample. A third limitation concerns the method of constructing the revised domain scales (i.e., N\_R, E\_R, A\_R, C\_R). The item-removal procedure is adequate in order to test the hypothesis that such removal results in increased affective specificity. However, the analyses are preliminary and as such insufficient for a complete restructuring of the NEO-PI-R. Furthermore, it would be preferable to undertake such restructuring efforts using the more recent NEO-PI-3, which has only very recently been translated into Dutch. It is very likely that demoralization is also of importance in that instrument, but of course, it should be investigated whether the procedures used for this project can be successfully applied for the NEO-PI-3, as well. Hence, until further replication and extended validation has been conducted, we can only recommend using and further developing these scales in research, but not in clinical practice. A further reason to not adopt these procedures too readily in clinical practice is that additional validation is needed, preferably including non-self-report external correlates.

These cautionary remarks notwithstanding, our findings have some implications for clinical practice. When clients are experiencing periods of severe demoralization, high scores on Neuroti-

cism and low scores on Extraversion or Conscientiousness may reflect nonspecific unpleasant affect rather than specific personality trait information. Hence, a thorough assessment of Demoralization is important for interpreting scores on self-report measures of personality traits in clinical practice as well as the development of methods for score interpretation in demoralized patients. The point is certainly *not* that Demoralization can be regarded simply as a state effect unaffected by stable interindividual differences (RCd, for instance, shows quite meaningful rank-order stability; Tellegen et al., 2003). Instead, the point worth making is that high levels of demoralization are only informative about relatively frequent and/or intense experiences of nonspecific unpleasant affect. To illustrate the clinical utility of this concept, Finn and Kamphuis (2006) cite an excerpt from a feedback session with a patient who had produced a generally elevated MMPI-2 Clinical Scale profile that was best explained by high demoralization rather than by core psychopathology:

Finn: As you can see, the first thing that stands out is that you have a lot of high . . . very high elevations on the problem scales.

Cl: I'm pretty screwed up, aren't I?

Finn: Well, I'm not reading it that way, myself. What I see and want to check with you is that you're . . . your main problem seems to be that you are extremely miserable. This page (shows Restructured Scales, RCs) breaks these other scores down, and you can see you have one really high score—on a scale that measures psychological misery. Everything else going on with you seems to pale in comparison with that. Am I right? The test says you are feeling really badly lately and that is overshadowing everything and affecting you in lots of ways.

As the current study shows, NEO-PI-R scores on Neuroticism and Extraversion clearly contain more specific information than merely demoralization. However, without an explicit distinction between demoralization and core constructs it is quite a challenge for clinicians to extract that specific information when assessing severely demoralized clients.

<sup>3</sup> In hindsight, it might have been preferable to start with the more heterogeneous sample. However, at the start of this project, we only had the data from the developmental sample available. At a later stage, we were able to obtain the second dataset and used that specifically for the purpose of replication.

## References

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: Author.
- Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher order factors of personality: Do they exist? *Personality and Social Psychology Review*, *13*, 79–91. <http://dx.doi.org/10.1177/1088868309338467>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. <http://dx.doi.org/10.1080/10705510903008204>
- Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.

- Clarke, D. M., & Kissane, D. W. (2002). Demoralization: Its phenomenology and importance. *Australian and New Zealand Journal of Psychiatry*, *36*, 733–742. <http://dx.doi.org/10.1046/j.1440-1614.2002.01086.x>
- Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and Extraversion. *Behavioral and Brain Sciences*, *22*, 491–517. <http://dx.doi.org/10.1017/S0140525X99002046>
- Derksen, J. J. L., De Mey, H. R. A., Sloore, H., & Hellenbosch, G. (1997). *MMPI-2: Handleiding bij Afname, Scoring en Interpretatie* [MMPI-2: Manual for Administration, Scoring and Interpretation] (3rd ed.). Nijmegen, The Netherlands: Pen Test Publisher.
- Finn, S. E., & Kamphuis, J. H. (2006). The MMPI-2 Restructured Clinical (RC) Scales and restraints to innovation, or “what have they done to my song?” *Journal of Personality Assessment*, *87*, 202–210. [http://dx.doi.org/10.1207/s15327752jpa8702\\_10](http://dx.doi.org/10.1207/s15327752jpa8702_10)
- Frank, J. D. (1961). *Persuasion and healing: A comparative study of psychotherapy*. Oxford, England: Johns Hopkins University Press.
- Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *NEO Persoonlijkheidsvragenlijst NEO-PI-R en NEO-FFI. Handleiding*. Lisse, The Netherlands: Swets and Zeitlinger.
- Hoerger, M. (2013).  $Z_H$ : An updated version of Steiger's Z and Web-based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from [http://www.psychmike.com/dependent\\_correlations.php](http://www.psychmike.com/dependent_correlations.php)
- Hopwood, C. J., Wright, A. G. C., & Donnellan, M. B. (2011). Evaluating the evidence for the general factor of personality across multiple inventories. *Journal of Research in Personality*, *45*, 468–478. <http://dx.doi.org/10.1016/j.jrp.2011.06.002>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*, 537–549. <http://dx.doi.org/10.1007/s11336-011-9218-4>
- Kamphuis, J. H., Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2008). Detecting comorbid Axis-II status among inpatients using the MMPI-2 Restructured Clinical Scales. *European Journal of Psychological Assessment*, *24*, 157–164. <http://dx.doi.org/10.1027/1015-5759.24.3.157>
- Karsten, J., Penninx, B. W., Riese, H., Ormel, J., Nolen, W. A., & Hartman, C. A. (2012). The state effect of depressive and anxiety disorders on Big Five personality traits. *Journal of Psychiatric Research*, *46*, 644–650. <http://dx.doi.org/10.1016/j.jpsychires.2012.01.024>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New dimensions in the quantitative classification of mental illness. *Archives of General Psychiatry*, *68*, 1003–1011. <http://dx.doi.org/10.1001/archgenpsychiatry.2011.107>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.
- Loehlin, J. C. (2012). How general across inventories is a general factor of personality? *Journal of Research in Personality*, *46*, 258–263. <http://dx.doi.org/10.1016/j.jrp.2012.02.003>
- Marsh, H. W., Hau, K., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Erlbaum Publishers.
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. <http://dx.doi.org/10.1037/a0019227>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: L. Erlbaum Associates.
- Muthén, L., & Muthén, B. (2012). *MPlus user's guide version 7.0*. Los Angeles, CA: Authors.
- Ormel, J., Jeronimus, B. F., Kotov, R., Riese, H., Bos, E. H., Hankin, B., . . . Oldehinkel, A. J. (2013). Neuroticism and common mental disorders: Meaning and utility of a complex relationship. *Clinical Psychology Review*, *33*, 686–697. <http://dx.doi.org/10.1016/j.cpr.2013.04.003>
- Quilty, L. C., Meusel, L. A., & Bagby, R. M. (2008). Neuroticism as a mediator of treatment response to SSRIs in major depressive disorder. *Journal of Affective Disorders*, *111*, 67–73. <http://dx.doi.org/10.1016/j.jad.2008.02.006>
- R Project. (2012). The R Project for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Revelle, W. (2013). Using R and the psych package to find  $\omega$ . Retrieved from <http://personality-project.org/r/psych/HowTo/omega.tutorial/omega.html#x1-150005.1>
- Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, *47*, 493–504. <http://dx.doi.org/10.1016/j.jrp.2013.04.012>
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145–154. <http://dx.doi.org/10.1007/s11336-008-9102-z>
- Rushton, J. P., Bons, T. A., & Hur, Y. M. (2008). The genetics and evolution of the general factor of personality. *Journal of Research in Personality*, *42*, 1173–1185. <http://dx.doi.org/10.1016/j.jrp.2008.03.002>
- Sellbom, M., & Ben-Porath, Y. S. (2005). Mapping the MMPI-2 Restructured Clinical scales onto normal personality traits: Evidence of construct validity. *Journal of Personality Assessment*, *85*, 179–187. [http://dx.doi.org/10.1207/s15327752jpa8502\\_10](http://dx.doi.org/10.1207/s15327752jpa8502_10)
- Sellbom, M., Ben-Porath, Y. S., & Bagby, R. M. (2008). Personality and psychopathology: Mapping the MMPI-2 Restructured Clinical (RC) scales onto the five factor model of personality. *Journal of Personality Disorders*, *22*, 291–312. <http://dx.doi.org/10.1521/pedi.2008.22.3.291>
- Sellbom, M., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., & Graham, J. R. (2006). Elevation differences between MMPI-2 clinical and restructured clinical (RC) scales: Frequency, origins, and interpretative implications. *Assessment*, *13*, 430–441. <http://dx.doi.org/10.1177/1073191106293349>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251. <http://dx.doi.org/10.1037/0033-2909.87.2.245>
- Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., Shelton, R., & Schalet, B. (2009). Personality change during depression treatment: A placebo-controlled trial. *Archives of General Psychiatry*, *66*, 1322–1330. <http://dx.doi.org/10.1001/archgenpsychiatry.2009.166>
- Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum, Inc.
- Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Technical manual*. Minneapolis, MN: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 Restructured Clinical (RC) scales: Development, validation, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., & Sellbom, M. (2009). Construct validity of the MMPI-2 restructured clinical (RC) scales: Reply to Rouse, Greene, Butcher, Nichols, and Williams. *Journal of Personality Assessment*, *91*, 211–221. <http://dx.doi.org/10.1080/00223890902794192>
- Tellegen, A., & Waller, N. (2008). Exploring personality through test construction: Development of the multidimensional personality ques-

- tionnaire. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *The Sage handbook of personality theory and assessment: Vol. 2 — Personality measurement and testing*. London, UK: Sage Publications Ltd.
- Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297–303. <http://dx.doi.org/10.1111/1467-9280.00157>
- Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology, 114*, 522–536. <http://dx.doi.org/10.1037/0021-843X.114.4.522>
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219–235. <http://dx.doi.org/10.1037/0033-2909.98.2.219>
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820–838. <http://dx.doi.org/10.1037/0022-3514.76.5.820>
- Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa, P. T. (1994). A description of DSM–III–R and DSM–IV personality disorders with the five factor model of personality. In P. T. Costa, Jr., & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 41–56). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/10140-003>
- Wygant, D. B., & Fleming, K. P. (2008). Clinical utility of the MMPI-2 Restructured Clinical (RC) scales in therapeutic assessment: A case study. *Journal of Personality Assessment, 90*, 110–118. <http://dx.doi.org/10.1080/00223890701845112>

Received January 17, 2014

Revision received October 29, 2014

Accepted October 30, 2014 ■