Interpreting neuroticism scores across the adult life course

Ormel, Johan; Riese, Harriette; Rosmalen, Judith G. M.

Published in:
Clinical Psychology Review

DOI:
10.1016/j.cpr.2011.10.004

Publication date:
2012

Clinical Psychology Review 32 (2012) 71–79

Contents lists available at SciVerse ScienceDirect

Clinical Psychology Review

Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect?

Johan Ormel [a,*], Harriëtte Riese [a,b], Judith G.M. Rosmalen [a]

[a] University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center for Psychiatric Epidemiology, Graduate School of Behavioral and Cognitive Neurosciences (BCN), Graduate School of Health Sciences (SHARE), The Netherlands
[b] University Medical Center Groningen, Department of Epidemiology, Unit of Genetic Epidemiology and Bioinformatics, The Netherlands

ARTICLE INFO

Article history:
Received 26 May 2011
Revised 27 September 2011
Accepted 24 October 2011
Available online 9 November 2011

Keywords:
Neuroticism
Psychopathology
Set point
Negative affect
Dynamic stability

ABSTRACT

Neuroticism (N) scores predict psychopathology. Therefore, it is important to know how to best interpret N-scores. This paper reviews prior interpretations, the item content of N-measures and relevant empirical studies. We propose that N-scores reflect person-specific negative affect set points. We distinguish three possible set point models. (1) The immutable set point model in which N-scores fluctuate with short-term perturbations in reaction to positive and negative experiences but always return to their person-specific set point. (2) The experience-dependent model in which an individual's set point can change during any life stage when prompted by far-reaching experiences. (3) The mixed model, a combination of the first two models, which separates the variation in neuroticism into stable and changing components. The changing component is experience-dependent. Current evidence provides little support for the immutable model. Rather, the evidence, though inconclusive, suggests that the experience-dependent or mixed model may help to explain between- and within-subject differences in N-scores across the life span. In particular, the observation that the differential consistency of N-scores tends to drop over time, but has not been shown to approach zero, is consistent with the mixed model. We discuss implications of the models and how to distinguish them empirically.

Contents

1. Introduction

High neuroticism (N) scores are strongly associated with the development, presence, and persistence of many forms of psychopathology (see for reviews Kotov, Gamez, Schmidt, & Watson, 2010; Lahey, 2009; Ormel, Rosmalen, & Farmer, 2004; Watson & Clark, 1984). N-scores also explain substantial covariance among current and lifetime mental disorders, especially internalizing disorders. This suggests that high N-scores are a core feature of internalizing

* Corresponding author at: University Center for Psychiatry, University Medical Center Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands. Tel.: + 31 50 3612078; fax: + 31 50 3619722.
E-mail addresses: j.ormel@med.umcg.nl (J. Ormel), h.riese@med.umcg.nl (H. Riese), j.g.m.rosmalen@med.umcg.nl (J.G.M. Rosmalen).

0272-7358/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cpr.2011.10.004

psychopathology (Clark, 2005). N-scores are also associated with subjective physical health. Various large studies of representative samples even found that low N-scores predict longevity after controlling for age, sex, education, smoking, alcohol consumption, physical activity, and baseline health (see for review Lahey, 2009). The economic burden of neuroticism to the society as a whole is enormous and exceeds those of common mental disorders (Cuijpers et al., 2010).

The significance of the prospective associations between N-scores and psychopathology is not straightforward. At least three factors hamper a straight forward causal interpretation of the prospective association. First, there is considerable overlap in item content between measures of neuroticism and measures of common forms of psychopathology. Second, accumulating evidence suggests substantial overlap in the etiology of neuroticism and psychopathology, in particular social anxiety and depression (Hettema, Neale, Myers, Prescott, & Kendler, 2006; Ormel, Riese, Bos, et al., 2011). Third, N-scores are elevated during an episode of mental disorder, particularly for internalizing disorders (Akiskal, Hirschfeld, & Yerevanian, 1983; Costa, Bagby, Herbst, & McCrae, 2005; Kendler, Neale, Kessler, Heath, & Eaves, 1993; Ormel, Oldehinkel, & Vollebergh, 2004). Although most evidence suggests that this so-called state effect is temporary, the effect's implications are unclear and strongly debated. Some claim it causes a major distortion of personality assessment (Akiskal et al., 1983) while others believe it to be a reliable and valid reflection of current personality (Costa et al., 2005).

The primary objective of this paper is to identify the best way to interpret N-scores. This may advance understanding of the association between N-scores and psychopathology. We review previous interpretations and critically examine the item content of N-measures. We also consider research on differential change and consistency of N-scores over time and the determinants of individual differences in intraindividual change such as life events. Research that targets individual differences in personality change is important because it addresses both why individual personality trajectories change and why mean-level changes in population occur (Roberts & Mroczek, 2008). Recently, several studies using growth modeling analytic techniques have shown substantial individual deviations from the normative pattern of mean-level change in neuroticism (see for overview Lüdtke, Roberts, Trautwein, & Nagy, 2011; Mroczek & Little, 2006).

Though still limited, the evidence suggests that neuroticism both predicts life experiences (selection effect) and responds to them (causation effect) (Lüdtke et al., 2011; Vaidya, Gray, Haig, & Watson, 2002) as N-scores seem to fluctuate over time around a person-characteristic set point of negative affect in response to life events. It also suggests that this set point may change in reaction to far-reaching experiences. Therefore, we will evaluate three variants of this dynamic equilibrium model of N-scores. (1) The immutable set point model in which N-scores fluctuate with short-term perturbations in reaction to positive and negative experiences but always return to their person-specific set point. (2) The experience-dependent model in which an individual's set point can change during any life stage when prompted by far-reaching experiences. (3) The mixed model, a combination of the first two models, which separates the variation in neuroticism into stable and changing components, with the changing component being experience-dependent. Finally, we describe how future research could evaluate and distinguish between the models to determine the most plausible interpretations of N-scores.

2. The concept of neuroticism versus neuroticism scores

Despite the central importance of neuroticism in personality theory, there is no consensus on its definition (Buss & Plomin, 1984; Cattell, Eber, & Tatsuoka, 1980; Digman, 1997; Eysenck & Eysenck, 1975; Mathews, Fox, Yiend, & Calder, 2003; McCrae & Costa, 1997; Tellegen & Waller, 1997; Widiger, 2009; Zuckerman, 1991; Zuckerman, 2003). Eysenck and others define neuroticism as a temperamental trait of emotionality: a tendency to arouse quickly when stimulated and to inhibit slowly when aroused (Eysenck & Eysenck, 1985). Costa and McCrae (1992b) define neuroticism as a major domain of personality that contrasts adjustment or emotional stability with maladjustment or negative emotionality. Neuroticism has also been defined as: inability to control urges; inefficiently coping with stress; preferring a particular threat management strategy; a disposition to complain; and tending to have unrealistic ideas, to appraise situations as stressful, and to experience aversive emotional states (John, Robins, & Pervin, 2008; Mathews et al., 2003; Pervin & John, 1999; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984). While some of these definitions closely match the item content of neuroticism questionnaires, others suggest possible underlying biological or psychological mechanisms that may generate high N-scores.

Researches have created many labels to characterize high N-scores. The term neuroticism has been favored by Eysenck, Costa and McCrae, and others, while the terms emotionality, emotional instability, and negative affectivity have been introduced by, respectively (Goldberg, 1993; Lee & Ashton, 2004; Watson, Clark, & Tellegen, 1988). These labels are often used interchangeably in the literature because of evidence-based consensus that individuals scoring high on N-measures exhibit emotional instability and negative affectivity (Shankman & Klein, 2003). Although many consider these alternative labels as similar to one another, their measures do reflect significant differences (Pervin & John, 1999). The differences are most marked in the level of the facet scales included in the measures. For instance, the neuroticism trait of the California Psychological Inventory (CPI) Big Five (Soto, John, Gosling, & Potter, 2008) distinguishes four facets (Anxiety, Irritability, Depression, and Rumination–Compulsiveness); the NEO Personality Inventory—Revised (NEO-PI-R) (Costa & McCrae, 1992a) six facets (Anxiety, Angry Hostility, Depression, Self-Consciousness, Vulnerability, and Impulsiveness); the Eysenck Personality Profiler three facets (Anxiety, Inferiority, Unhappiness); the Big Five Aspects Scale (DeYoung, Quilty, & Peterson, 2007) two aspects (Withdrawal, Volatility); and the Multidimensional Personality Questionnaire (MPQ) (Tellegen & Waller, 1997) three facets (Alienation, Stress Reaction, Aggression). While Anxiety–Withdrawal, Depression–Unhappiness, Vulnerability–Stress Reaction are typically considered major facets of neuroticism, there is less agreement whether Angry Hostility–Aggression, Impulsivity, Inferiority, and Dependency belong to the neuroticism domain (Widiger, 2009).

Both the broad trait of neuroticism and its lower-order facets rest almost entirely on factor analyses of words used to describe behaviors in daily life, the lexical approach (Matthews, Deary, & Whiteman, 2003; Pervin & John, 1999). Self-report questionnaires and verbal reports are the most common methods of personality assessment (John et al., 2008). Objective behavioral tests have not been validated and are rarely used (Mathews et al., 2003; Pervin & John, 1999). The item content of neuroticism questionnaires and measures of common mental disorders are strikingly similar. Many N-items are similar to the symptom measures of anxiety disorders, mood disorders, impulse control problems, psychological distress, and the negative affect and cognitions that often accompany other mental disorders. Table 1 lists a large number of items from the 48-item N-domain of the NEO-PI-R and similar items from two popular symptom measures, the 90-item Symptom Checklist (SCL) and the 60-item General Health Questionnaire (GHQ). However, there are two major differences between N-items and symptom measures: N-items lack a specific time frame and severity descriptors. Instead, neuroticism inventories use non-specific descriptors of frequency, intensity, and duration. This non-specificity of items facilitates the disclosure of negative

Table 1
Item overlap between the neuroticism domain of the NEO-PI-R and two popular symptom measures, the 90-item Symptom Checklist (SCL) and the 60-item General Health Questionnaire (GHQ).

Neuroticism facets NEO-PI-R[a]		SCL 90 (past 2 weeks)	GHQ (past 30 days)
N1 anxiety			
1	I am not a worrier.	Worrying too much about things.	Worrying unduly?
31	I am easily frightened.	Suddenly scared for no reason.	Scared and panicky …no good reason?
61	I rarely feel fearful or anxious.	Feeling fearful.	
91	I often feel tense and jittery.	Nervousness or shakiness inside.	Easily upset over things? Feeling nervous and strung up?
151	I often worry about things that might go wrong.		Afraid something awful is going to happen?
181	I have fewer fears than most people.		
211	Frightening thoughts sometimes come into my head.	The idea that something is wrong with your mind.	Unwelcome thoughts keep coming in?
N2 irritability			
36	I'm an even-tempered person.	Temper outbursts that you could not control.	Edgy and bad tempered?
66	I am known as hot-blooded and quick-tempered.	Shouting or throwing things.	Losing temper?
96	I am not considered a touchy or temperamental person.		Others regard you as touchy?
156	I takes a lot to get me mad.	Getting into frequent arguments.	Little annoyances, upset and angry?
216	Even minor annoyances can be frustrating to me.	Feeling easily annoyed or irritated.	
N3 depression			
11	I rarely feel lonely or blue.	Feeling blue.	
41	Sometimes I feel completely worthless.	Feeling or worthlessness.	Felt that life isn't worth living?
71	I am seldom sad or depressed.		Feeling unhappy and depressed.
101	I have sometimes experienced a deep sense of guilt or sinfulness.	The idea that you should be punished for your sins. Feelings of guilt.	
131	I tend to blame myself when anything goes wrong.	Blaming yourself for things.	Blaming yourself things going wrong?
161	I have a low opinion of myself.	Feelings inferior to others.	
221	Too often, when things go wrong, I get discouraged and feel like giving up.	Worried about sloppiness or carelessness.	
N4 consciousness			
46	I seldom feel self-conscious when I'm around people.	Feeling very self-conscious with others.	
76	At times I have been so ashamed I just wanted to hide.		Feeling that you're a burden?
136	I often feel inferior to others.	Feeling inferior to others.	Thinking of yourself as worthless?
196	If I have said or done the wrong thing to someone, I can hardly bear to face them again.	Worried about sloppiness or carelessness.	Afraid….. express foolish mistakes?
N5 impulsiveness			
171	I sometimes eat myself sick.	Overeating.	
201	Sometimes I do things on impulse that I later regret.	Having urges to beat, injure or harm someone. Having urges to break or smash things.	
231	I am always able to keep my feelings under control.	Temper outbursts that you could not control.	
N6 vulnerability			
26	I often feel helpless and want someone else to solve my problems.		Couldn't overcome difficulties?
56	I feel I am capable of coping with most of my problems.		Able to face your problems?
116	I keep a cool head in emergencies.	Your mind going blank.	
146	It's often hard for me to make up my mind.	Difficulty making decisions.	
206	When everything seems to be going wrong, I can still make good decisions.		Capable of making decisions?
236	I'm pretty stable emotionally.	Your feelings being easily hurt.	Feeling easily hurt?

Note. NEO-PI-R NEO = Personality Inventory-Revised (Costa & McCrae, 1992a); SCL-90 = Symptom Check List (Derogatis, Lipman, & Covi, 1973); GHQ = General Health Questionnaire (Goldberg & Williams, 1988).
[a] Number refers to the NEO-PI-R item number.

affect. In sum, we may conclude that neuroticism inventories measure a variety of negative emotions and cognitions that feature heavily in the measures and diagnostic criteria of many mental disorders, in particular internalizing disorders but also externalizing disorders although to a lesser extent. In addition, neuroticism questionnaires measure the propensity to express these emotions.

The difficulties in defining the construct of neuroticism are largely due to a lack of consistent evidence on the neurobiological bases and determinants of N-scores. In particular, psychophysiological research regarding central non-specific activation, autonomic peripheral activation, and hypothalamo-pituitary-adrenal axis activation (cortisol) has yielded inconsistent results (Claridge & Davis, 2001; Ormel, Riese, Bastiaansen, et al., 2011; Zuckerman, 2003). Recent neurobiological (Depue, 2009), neurogenetic (Canli, 2008) and evolutionary-psychological approaches (Matthews, 2004) have produced promising findings, but these approaches are still in their infancy and their findings need replication and further elaboration. Hence, it seems wise to base the interpretation of N-scores on the definable characteristics of N-measures, in particular item content, stability and change of N-scores across the adult life course, and the determinants of change in N-scores. The latter two topics are reviewed in the next two sections.

3. Differential consistency and change in N-scores

The literature typically distinguishes two types of consistency (or stability): group mean level consistency (absolute stability) versus rank-order (differential) consistency. In other words, the stability of the group as a whole versus the retention of an individual's relative placement within that group. Differential change is fueling individual differences in intra-individual change (Mroczek & Spiro, 2003). The largest changes in personality after childhood occur in late adolescence and early adulthood (Roberts, Walton, & Viechtbauer, 2006). In that period the population mean of neuroticism drops substantially. After age 30, the group mean level of neuroticism in the population

tends to remain stable. This suggests that age-related shifts in neuroticism are limited during adulthood (Roberts et al., 2006).

Meta-analytic estimates of mean population test-retest correlation coefficients show clearly that differential consistency increases with age, from around .30 during infancy to .45 during adolescence and .60 during young and middle adulthood (until age 50) (Roberts & DelVecchio, 2000). Differential consistency over protracted intervals is substantial, especially after age 30—but importantly also steadily drops over longer time intervals across the entire life-span. The negative correlation between time interval and differential consistency is modest but highly significant (−.36). Starting at age 20, Roberts and DelVecchio estimated the average trait consistency over a 1-year period at .55; over 5 years .52; over 10 years .49; over 20 years .41, and over 40 years .25. These estimates were largely based on studies that assessed neuroticism twice. Although attrition, which was typically substantial, did not distort the differential consistency estimates, many studies used samples drawn from privileged and educated populations. This may have biased consistency estimates upward (Roberts & DelVecchio, 2000).

Two recent studies assessed neuroticism at least four times during a period of about two decades, reporting somewhat lower rank-order stabilities (Ormel & Rijsdijk, 2000; Wray, Birley, Sullivan, Visscher, & Martin, 2007). Wray and colleagues found test–retest correlations of around 0.60 for intervals of about 10 years and from 0.57 to 0.45 for intervals ranging from 13 to 22 years. These are similar to the test–retest correlations of about 0.58 for 8- to 10-year intervals and 0.51 to 0.45 for 14- to 16-year intervals reported by Ormel and Rijsdijk. Collectively, the longitudinal studies suggest substantial continuity of neuroticism during adulthood, with a gradual decline in consistency reaching an asymptote of about 0.40. This pattern of longitudinal correlations suggests the workings of two dynamic processes: a trait process producing differential consistency and an autoregressive process producing differential change.

To explain the pattern of longitudinal correlations, Ormel and Rijsdijk (2000) compared three longitudinal models using 5-wave neuroticism data covering 16 years of adult life from a random population sample of 296 adults. (1) The trait (or common factor) model asserts that neuroticism is correlated over time only because of underlying traits which are immutable in adulthood. The trait model predicts that test–retest correlations are independent of the length of interval (assuming stable trait expression over time). (2) The autoregressive (or simplex) model posits ongoing cumulative differential change. That is, neuroticism changes, on average, slowly but continuously. The auto-regression model predicts that test–retest correlations decrease gradually over increasing time, ultimately declining towards a correlation of zero. Both of these models express extreme positions. (3) Fig. 1 presents the mixed model for a 5-wave assessment of neuroticism. This mixed model disentangles the variation in neuroticism into one immutable or trait component (T, in Fig. 1) and a second component (S_i) that changes over time according to a first order auto-regressive process. This autoregressive process includes inertia-driven carry-over effects and effects from environmental exposures that lead to experiences that change the individual's N-score for shorter or longer periods of time. The mixed model predicts that correlations will decrease with time, but the constant influence of the immutable component will ensure that they never approach zero.

Though it provides little support for the trait model, the data was found to fit both the auto-regression and mixed models (Ormel & Rijsdijk, 2000). The auto-regression model estimated the 16-year auto-regression at .60 (95%CI .50–.69), which gives an annual auto-regression of .967. This estimate is similar to the annual auto-regression of 0.98 reported earlier (Conley, 1984). Thus, the auto-regression model suggests a slow but sustained rate of change in N-scores. In the mixed model, the trait component accounted for 38% of the variance in neuroticism (95% CI 0%–55%). Note that the CI includes zero variance. Extrapolation to 30 years yields a 30-year differential consistency of .37 according to the auto-regression model and of .45 according to the mixed model. The strong auto-regressive effect in longitudinal evaluations of N-scores

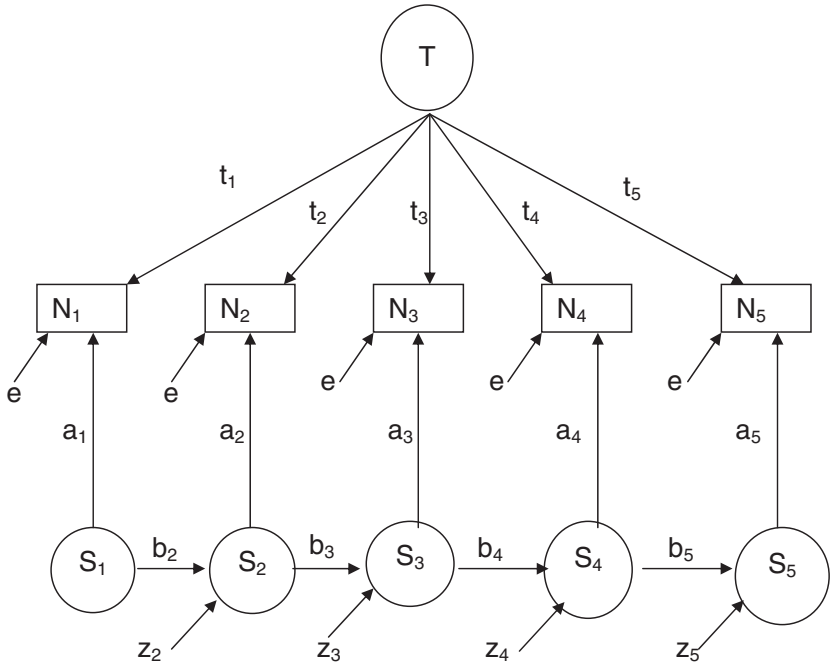

Fig. 1. Mixed model for 5-wave assessment of neuroticism. Note: T = latent trait factor; Ni = observed neuroticism scores at 5 waves; Si = latent changing component; ti = regression coefficient (factor loading indicating the stable variance in the observed N-scores); ai = regression coefficient (indicating the changing variance in the observed N-scores). e = measurement error variance; zi = influence of unobserved determinants of change in N-scores; bi = auto-regression coefficient (if intervals are equal, these will typically be equal as well). If the bottom part of the model (Si)) is removed, the trait model emerges. If the top part of the model (T) is removed, the autoregressive model emerges.
Adapted from Ormel & Rijsdijk, 2000).

accounts for the large confidence interval in the trait variance estimate of the mixed model (95% CI 0%–55%). This suggests the need for a large sample size in order to accurately disentangle the contributions of trait effect versus auto-regression (carry-over effects) to differential consistency.

The existence of both differential consistency and change is consistent with results from studies of twins. Twin studies generally report heritability variance estimates for neuroticism in the 40% to 50% range (Eaves et al., 1999; Flint, 2004; Heath, Neale, Kessler, Eaves, & Kendler, 1992; Jardine, Martin, & Henderson, 1984; Loehlin, 1992; Tambs, Sundet, Eaves, Solaas, & Berg, 1991; Viken, Rose, Kaprio, & Koskenvuo, 1994). Furthermore, the correlations over time of nonmeasured genetic influences are very high, suggesting that neuroticism's genetic determinants remain largely unchanged over time. Conversely, the correlations over time of nonmeasured environmental determinants (latent environmental variables) are much smaller suggesting that they are more time-specific and subject to change (McGue, Bacon, & Lykken, 1993; Wray et al., 2007). For instance, in the Australian–Dutch collaborative study (Wray et al., 2007), the correlations between nonmeasured genetic influences over periods of up to 22 years were all extremely high (mean 0.90) and hardly decreased over time. In contrast, the correlations between nonmeasured environmental influences over 22 years dropped from 0.53 to 0.24, suggesting that environmental influences are more time-specific and are replaced by other environmental events. Both these findings are consistent with the mixed model.

4. Determinants of change in N-scores

Which experiences during adulthood cause sustained change in N-scores and explain differential change? It is unlikely that aging itself is involved, since mean levels of neuroticism in the population remain relatively stable after age 30. The substantial decrease in mean level during late adolescence and early adulthood and the much smaller decrease during later adulthood is typically attributed to intrinsic maturational processes (McCrae et al., 2000; Roberts & Mroczek, 2008) including the effects of age-graded social roles (Roberts, Wood, & Caspi, 2008). The low rate of differential change in N-scores during adulthood suggests that experiences that bring about sustained change are infrequent.

Exposure to persistent uncontrollable stress, in particular during adolescence and young adulthood, seems to contribute to an individual's N-scores later in life (Roberts, Wood, & Smith, 2005; Rutter, 2006; van Os & Jones, 1999). In contrast, controllable stress paired with adequate social support tends to reduce N-scores (Andrews, Page, & Neilson, 1993; Rutter, 2006). Another category of change agents include success and failure in the major social roles of marriage and work that seem dynamically related to personality (Caspi & Roberts, 1999; Robins, Fraley, Roberts, & Trzesniewski, 2001; Vaidya et al., 2002). High N-scores predict feelings of relationship insecurity, while the formation of a romantic relationship leads to a decrease in N-scores (Lucas, Clark, Georgelis, & Diener, 2004; Neyer & Asendorpf, 2001). Marriage, remarriage, increasing marital satisfaction, and satisfying and engaging employment are associated with decreases in N-scores and associated measures (Roberts & Chapman, 2000; Scollon & Diener, 2006). In contrast, conflict, poor relationship quality, divorce (for men) and chronic or repeated unemployment lead to increases in N-scores (Costa, Herbst, McCrae, & Siegler, 2000; Lucas et al., 2004; Lüdtke, Trautwein, & Husemann, 2009; Robins, Caspi, & Moffitt, 2000; Scollon & Diener, 2006). This evidence supports the neosocioanalytic model of personality development (Roberts et al., 2008). This model posits that adult personality development is driven by the experiences that go along with age-graded social roles. The expectations and contingencies associated with age-graded roles in work, family, community, and society bring along rewards and punishments that prompt people to become less neurotic (and more conscientious and agreeable) (Roberts et al., 2008).

The evidence on the dynamic relationship between social role experiences and personality also supports a transactional model with its selection and causation effects. Selection refers to the influence of pre-event personality on people's experiences. Most experiences do not occur randomly. Rather, they are brought about by the individual, at least to some extent, because people's personalities select them into particular situations or because their personally-related behaviors evoke certain reactions from others (Headey & Wearing, 1989; Ormel & Wohlfarth, 1991; Scarr & McCartney, 1983). Causation effects refer to the causal influence of experiences on personality and are also called socialization effects in the developmental literature (Vaidya et al., 2002).

Another important theoretical explanation for both mean-level change and individual differences in intra-individual change in N is life-span developmental theory which stresses the role of nonnormative life events (Baltes & Brim, 1979). Such nonnormative experiences are infrequent and outside the range of typical social role experiences and hypothesized to have the potential to alter the trajectory of personality development (Costa et al., 2000; Löckenhoff, Terracciano, Patriciu, Eaton, & Costa, 2009; Lüdtke et al., 2011; Middeldorp, Cath, Beem, Willemsen, & Boomsma, 2008; Mroczek & Spiro, 2003). In particular, experiences that affect central aspects of one's identity such as loss of a job, divorce, personal illness or injury, and other major changes in work, marital, and health status, have been found to predict neuroticism change. Here too, the still limited evidence suggests transactional effects in that neuroticism predicts (selection effect) and responds to (causation effect) life experiences as well. For instance, Ludtke and colleagues found that baseline N-scores predict negative life events (but not positive life events) whereas change in neuroticism is linked to the experience of negative and positive life events as well, with negative events being linked to an increase in N-scores and positive events with a decrease in N-scores. Finally, there is some evidence that effective treatment of depression with SSRI's (Tang et al., 2009) and CBT (Jorm, 1989) can reduce N-scores in excess and independent of the effect on depression.

It is worth of note that nonexperimental real-world research can never demonstrate that truly causal processes are driving the selection and causation associations; these associations could be due to the effect of unobserved third variables, including genetic confounding. Genetic confounding refers to the possibility that genetic characteristics have causal effects on the environmental exposure as well as on the development of neuroticism, thereby producing non-causal statistical association among these variables. The monozygotic (MZ) co-twin control design is a strong design to reduce the potential for genetic confounding and increase causal plausibility. The few studies using MZ twins have found that neuroticism scores increase after exposure to negative life events, particularly when the twin experience personal illness or injury (Middeldorp et al., 2008) .

5. Interpretation of N-scores

How should we interpret N-scores? As described earlier, the content of N-items overlaps with the symptoms of common mental disorders, in particular those of internalizing psychopathology. In addition, N-items use vague descriptors of frequency, intensity, and duration. They do not define the time frames about which they ask, increasing the likelihood that respondents disclose complaints about how they feel and function. These distinctive features suggest an interpretation of N-scores as a person's self-perceived level of negative affect over a prolonged period. Psychosocial events, and measurement error, may temporarily alter the estimate of negative affect. However, decay of the effect of events and homeostatic

normalizing forces will tend to neutralize these deviations and return negative affect levels back to a person-specific set point.

The set point theory of subjective well-being is relevant here. Its central tenet is that adult individuals have differing but stable levels of subjective well- and ill-being (Brickman & Campbell, 1971; Headey & Wearing, 1991; Lykken & Tellegen, 1996; Ormel & Schaufeli, 1991). In accordance with the set point theory, an individual's N-score is in a dynamic equilibrium, fluctuating over time around a person-characteristic set point in reaction to positive and negative experiences. The dynamic equilibrium model is compatible with evidence showing substantially higher N-scores during episodes of exposure to stressful life events and mental illness in comparison to pre- and post-episode periods (Akiskal et al., 1983; Costa et al., 2005; Kendler et al., 1993; Ormel, Oldehinkel, & Vollebergh, 2004; Ormel, Rosmalen, & Farmer, 2004).

We distinguish two fundamental conceptualizations of set points. The *immutable set point* posits that individual set points of neuroticism do not change. Many personality psychologists espouse this view regarding personality traits as internally driven and independent of environmental influences. This dynamic-equilibrium model of an immutable person-characteristic set point of neuroticism is compatible with the trait model discussed above. It implies that homeostatic forces keep an individual's set point constant over time, bringing it back to its original level after particular experiences have pushed it away. However, though the notion of an immutable set point fits the evidence of neuroticism's high differential consistency, but it is inconsistent with the robust evidence showing a slow but steady decline in differential consistency over time.

In contrast, the *experience-dependent set point* assumes that set points can change, throughout life, when prompted by far-reaching experiences that become biologically, cognitively, or environmentally embedded. Biological embedding holds that environmental signals cause sustained changes in regulatory neurophysiological systems (Weaver et al., 2004; Zhang & Meaney, 2010). Recently many studies have focused on the molecular mechanisms mediating the sustained changes including epigenetic processes such as DNA methylation and chromatin remodelling. Such changes often occur early in life and can be quite permanent. From a biological point of view, the experience-dependent model presupposes lifelong environmentally driven neural plasticity, based on epigenetic processes (Zhang & Meaney, 2010). Cognitive embedding occurs when far-reaching experiences lead to persistent alterations in beliefs about the self and others and the appraisal of and coping with stressful events. Environmental embedding occurs when far-reaching experiences cause sustained changes in the social environment that in turn cause long-term changes in exposure to determinants of negative affect. Examples of such experiences are a good or a bad marriage, entering a good career or becoming chronically unemployed, or lasting major health changes.

The *mixed set point* model combines these two basic concepts. In the mixed set point model, N-scores partly reflect an immutable set point and partly a changing experience-dependent set point that undergoes shifts beyond those of measurement error and temporary variation. The mixed set point model assumes that consistency correlations will fall with time yet never reach zero because of the persistent influence of immutable characteristics.

6. How to distinguish the models

The experience-dependent and mixed set point models are difficult to distinguish empirically, especially when the rate of change in N-scores is very low. Because N-changing life events are rare and typically have only small effects, differential change will be very slow. This would be true even assuming an entirely experience-dependent model with no immutable factor. Thus, when N-scores change slowly, most N-score variance will appear immutable. Notwithstanding these difficulties, the models have three differing empirical implications. First, the mixed set point model predicts that the drop in differential consistency will level off with increased time. In contrast, the experience-dependent model predicts an ongoing drop even after intervals of 10–20 years. Second, the strength of the association between major life events and N-scores at successive follow-ups will drop faster in the mixed model than in the experience-dependent model. This is because the mixed model assumes that part of the N-score variance is 'untouched' by life events. Third, if the experience-dependent model is correct, statistical auto-regression of longitudinal N-scores during adulthood should fit better than a mixed model with a large immutable variance component.

It is not easy to test these predictions, though the statistical tools for the analysis of change and its determinants have become available (Little, Bovaird, & Slegers, 2006; Ormel & Rijsdijk, 2000; Roberts et al., 2008). To compare the models, longitudinal data sets are required with at least three assessments of neuroticism using the same instrument over protracted periods of time in relatively large population-based samples (Bollen & Curran, 2006; Kenny & Zautra, 1995; Ormel & Rijsdijk, 2000). The total length of the study should be preferably 10 years or longer, in particular when the participants are 30 years or older at the start of the study. Ideal studies would cover adolescence, young adulthood and midlife. It is best if between-wave intervals are longer than a year given the very small drop in differential consistency per year. Studies with more than three waves are highly desirable as 3-wave models require rather strong assumptions to identify the mixed model (Ormel & Rijsdijk, 2000). Models with more waves allow testing such assumptions. Variation in interval length is not a problem as long as the longer intervals can be expressed as a function of the shortest interval. Sample size is important for statistical power and preferably be over 1000 (Ormel & Rijsdijk, 2000). It is best if the sample is population-based with a narrow age range as this helps ensure that the sample is in steady state, so that mean and variance of N-scores will not change over time from factors other than those to which the entire cohort is exposed such as aging, birth cohort, and period effects. Samples of individuals that are recruited at 'non-random time points in their life' such as when in treatment or experiencing a major event are problematic because their baselines N-scores are 'biased'.

Not many studies meet these requirements. Table 2 describes a number of long-term longitudinal studies of neuroticism that involved at least three assessment waves. The overview is not meant to be exhaustive but exemplary to illustrate what has been done and the strengths and weaknesses of the databases for studying the models. Studies that target N-scores as dependent variables are uncommon and recent but are relatively easy to find because they mention that neuroticism was assessed at multiple waves. In contrast, studies which used neuroticism only as an independent variable – although assessed multiple times – are difficult to track down because such studies rarely indicate that neuroticism was assessed at multiple occasions. These studies often do not mention the multiple assessment of neuroticism at all or "hide" it in the Methods section. For each study in Table 2 we indicate the characteristics and strengths and weaknesses for testing the models. Attrition is a widespread problem and unfortunately, is rarely random. In general, the longer a study is, with more waves, the smaller the sample size and the larger the attrition. Nonetheless some studies, in particular when pooled, may be usable to test the models.

7. Concluding comments: Towards a better interpretation of neuroticism

Neuroticism has long been considered to be the high-order personality trait most relevant to psychopathology. However, historically, the supporting research has been based solely on factor analyses of

Table 2
Characteristics of Long-term longitudinal studies of neuroticism with at least three assessment waves in adulthood.

Reference	Waves of N[a]	Total length[b]	Interval length[b]	*N* at start	Attrition[c]	Age (yr) (*M*; *SD*)	Strength[d]	Weakness[d]
Lüdtke et al. (2009)	3	4	2	4544	60%	19; 1	N	Waves; attrition; length
Scollon and Diener (2006)	5	9	2	1130	35%[c]	37; 13	N; >3 waves	Age range; administration[e]
Neyer and Asendorpf (2001)	3	8	4	637	47%	24; 4		Waves; attrition
Steunenberg, Twisk, Beekman, Deeg, and Kerkhof (2005)	3	6	3	2117	42%	55–85	N	Waves; attrition; age range
Roberts and Chapman (2000)	4	30	6–16	142	27%	21, 1	>3 waves; length	N
Mroczek and Spiro (2003)	5	12	1–3	1663	26%[c]	43–91	N; >3 waves; length	Age range; initially healthy[f]
Jones, Livson, and Peskin (2003)	4	42	13	279	22–69%	33–35	Length	N; attrition
Kendler, Gardner, and Prescott (2002)	3	7	1–4	2354	~26%	36; 8	N; attrition	Waves; administration[e]
Ormel, Oldehinkel, and Vollebergh (2004)	3	3	1–2	7067	32%	18–64	N; attrition	Waves; length; age range
Ormel and Rijsdijk (2000)	5	16	1–8	383	23%	34; 12	>3 waves; length	N; age range
Wray et al. (2007)	4	22	3–10	4040	>20%[c]	33; ??	>3 waves; length	Many were not followed up [c].
Middeldorp et al. (2006)	3	6	2–4	7969	> 58%	18; 2	N	Attrition; waves
Terracciano, McCrae, Brant, and Costa (2005)	5	9	M = 2.8	1944	65%	57;17	>3 waves	Attrition; age range

Note. To be included a study had to meet the following criteria: population-based, 3+ waves, total length 3+ years. The overview is not meant to be exhaustive. Its purpose is to illustrate what has been done and the strengths and weaknesses of the studies for the analysis of trait-state models.

[a] Number of times neuroticism was measured.

[b] Time span in years between first and last measurements of neuroticism; Interval length in years as well.

[c] Attrition was defined contingent on the number of waves. For 3- and 4-wave studies, attrition refers to the percentage of participants who missed at least 1 administration; for studies with 5 or more waves, attrition refers to the percentage of participants who missed at least 2 administrations. Total attrition in 5+ wave studies is typically much higher. In the Scollon and Diener (2006), Mroczek and Spiro (2003), and Ormel and Rijsdijk (2000) studies, respectively, only 33%, 26%, and 58% provided all 5 measurements. In the Wray study, exact attrition rates cannot be computed because many were not eligible for follow-up(s). Less than 1000 individuals were assessed at three or more occasions. In the Terracciano et al., 2005 study the frequency of assessments ranged from 1 to 11. Attrition refers to the 1260 who participated twice or less; 684 participated at 3 or more occasions.

[d] As strengths we consider 4+ waves, a total length of 10+ years, a sample size of 1000+, attrition <35%, and a limited age range. As weaknesses, we consider 3 waves, a total length of <5 years, a sample size of <500, attrition >40%, and a large age range.

[e] Administration refers to a change in how the neuroticism measure was administrated during the study, typically from paper and pencil to (phone) interview or vice versa.

[f] Subjects with physical or mental health problems were excluded.

adjectives used to describe behavior. This methodological limitation has resulted in uncertainty about the interpretation of N-scores. This uncertainty is increased by the lack of consensus on the optimal conceptualization and facet structure of neuroticism, its largely unknown underlying biological mechanisms, and the strong overlap between measures of neuroticism and symptoms of common mental disorders. Even so, because N-scores remain popular in psychopathological research, it is crucial to determine what N-scores mean.

The mixed model described in this paper posits that N-scores reflect a person's level of negative affect during a particular period. That level of negative affect fluctuates over time around a person-specific set point in response to positive and negative experiences. Thus, an individual's negative affect set point predicts risk of psychopathology. This should hold true whether the negative affect set point is immutable over the life span, or whether it changes in response to life experiences. The uncertainty on the interpretation of N-scores does not imply that N-scores are without value. Rather, the measure of person-specific negative set points provides a powerful and easy-to-measure marker of risk for mental disorder (Ormel, Rosmalen, & Farmer, 2004; Watson & Walker, 1996). However, for N-scores and the risk they index to become etiologically and pathogenetically informative, their neurobiological basis and specific genetic and environmental determinants need to be established.

The immutable, experience-dependent and mixed set point models provide a promising framework for future research to better understand N-scores. Current evidence, though inconclusive, provides little support for a fully immutable model. Rather, it suggests that an experience dependent or mixed model may help to explain the course of N-scores across the life span, including individual differences in intraindividual change (Mroczek & Spiro, 2003). In particular, the observation that the differential consistency of N-scores drops over time, but has not been shown to approach zero, is most consistent with the mixed model. The immutable component may reflect the possibility that people with low N-scores are not susceptible to sustained changes in negative affect whereas people with relatively high N-scores are more likely to experience-dependent changes. There is some evidence that if people with high N-scores are exposed to either particularly favorable or adverse life events, their set-points of positive and negative affect may change (Belsky & Pluess, 2009; Boyce & Ellis, 2005; Huppert, Baylis, & Keverne, 2005). By beginning to understand how genetic and environmental factors affect neuroticism, N-scores may transcend their current utility as a risk-index of mental disorder to become an effective tool in the treatment of psychopathology.

Conflict of interest

Drs Ormel, Riese, and Rosmalen report no conflict of interest.

Acknowledgement

Dr Ormel is financially supported by the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; NWO investment grant 175.010.2003.005); and by the European Science Foundation (EuroSTRESS project FP-006). These funding bodies did not have any role in the study design, data collection, analyses and interpretation.

References

Akiskal, H. S., Hirschfeld, R. M., & Yerevanian, B. I. (1983). The relationship of personality to affective disorders. *Archives of General Psychiatry*, *40*, 801–810.

Andrews, G., Page, A. C., & Neilson, M. (1993). Sending your teenagers away. controlled stress decreases neurotic vulnerability. *Archives of General Psychiatry*, *50*, 585–589.

Baltes, P. B., & Brim, O. G. (Eds.). (1979). *Life-span development and behavior*. London: Academic Press.

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*, 885–908.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* New York: Wiley.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology*, *17*, 271–301.

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation-level theory: A symposium* (pp. 287–302). New York: Academic Press.

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits.* Hillsdale, NJ: Lawrence Erlbaum Ass Inc.

Canli, T. (2008). Toward a neurogenetic theory of neuroticism. *Annals of the New York Academy of Sciences*, *1129*, 153–174.

Caspi, A., & Roberts, B. W. (1999). In L. A. Pervin, & O. P. John (Eds.), *Personality change and continuity across the life course. Handbook of personality theory and research*, *vol. 2*. (pp. 300–326) New York: Guilford Press.

Cattell, R., Eber, H., & Tatsuoka, M. (1980). *Handbook for the sixteen personality questionnaire (16PF).* Champaign, Il: Institute for Personality and Ability Testing.

Claridge, G., & Davis, C. (2001). What's the use of neuroticism? *Personality and Individual Differences*, *31*, 383–400.

Clark, L. A. (2005). Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*, *114*, 505–521.

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality and self-opinion. *Personality and Individual Differences*, *5*, 11–25.

Costa, P., Bagby, R., Herbst, J., & McCrae, R. (2005). Personality self-reports are concurrently reliable and valid during acute depressive episodes. *Journal of Affective Disorders*, *89*, 45–55.

Costa, P., Herbst, J., McCrae, R., & Siegler, I. (2000). Personality at midlife: Stability, intrinsic maturation, and response to life events. *Assessment*, *7*, 365–378.

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

Costa, P. T., Jr., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*, 343–359.

Cuijpers, P., Smit, F., Penninx, B. W. J. H., de Graaf, R., ten Have, M., & Beekman, A. T. F. (2010). Economic costs of neuroticism. A population-based study. *Archives of General Psychiatry*, *67*, 1086–1093.

Depue, R. A. (2009). Genetic, environmental, and epigenetic factors in the development of personality disturbance. *Development and Psychopathology*, *21*, 1031–1063.

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale-preliminary report. *Psychopharmacology Bulletin*, *9*, 13–28.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, *93*, 880–896.

Digman, J. M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology*, *73*, 1246–1256.

Eaves, L., Heath, A., Martin, N., Maes, H., Neale, M., Kendler, K. S., et al. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Research*, *2*, 62–80.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck personality questionnaire.* San Diego: EdITS.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences. A natural science approach.* New York: Plenum Press.

Flint, J. (2004). The genetic basis of neuroticism. *Neuroscience and Biobehavioral Reviews*, *28*, 307–316.

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

Goldberg, D., & Williams, P. (1988). *A user's guide to the general health questionnaire.* Nelson Publishing.

Headey, B., & Wearing, A. (1989). Personality, life events, and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology*, *57*, 731–739.

Headey, B., & Wearing, A. (1991). Subjective well-being: A stock and flows framework. In F. Strack, M. Argyle, & N. Schwartz (Eds.), *Subjective well-being* (pp. 49–77). Oxford: Pergamon Press.

Heath, A. C., Neale, M. C., Kessler, R. C., Eaves, L. J., & Kendler, K. S. (1992). Evidence for genetic influences on personality from self-reports and informant ratings. *Journal of Personality and Social Psychology*, *63*, 85–96.

Hettema, J., Neale, M., Myers, J., Prescott, C., & Kendler, K. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders RID B-1418-2008. *The American Journal of Psychiatry*, *163*, 857–864.

Huppert, F., Baylis, N., & Keverne, B. (2005). In F. Huppert, N. Baylis, & B. Keverne (Eds.), *The science of well-being.* New York: Oxford University Press.

Jardine, R., Martin, N. G., & Henderson, A. S. (1984). Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genetic Epidemiology*, *1*, 89–107.

John, O. P., Robins, R., & Pervin, L. A. (2008). In O. P. John, R. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*. New York: Guilford.

Jones, C. J., Livson, N., & Peskin, B. (2003). Longitudinal hierarchical linear modeling analyses of California psychological inventory data from age 33 to 75: An examination of stability and change in adult personality. *Journal of Personality Assessment*, *80*, 294.

Jorm, A. F. (1989). Modifiability of trait anxiety and neuroticism — A meta-analysis of the literature. *The Australian and New Zealand Journal of Psychiatry*, *23*, 21–29.

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *The American Journal of Psychiatry*, *159*, 1133–1145.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A twin study of recent life events and difficulties. *Archives of General Psychiatry*, *50*, 789–796.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*, 52–59.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*, 768–821.

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, *64*, 241–256.

Lee, K., & Ashton, M. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*, 329–358.

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. K. Mroczek, & T. D. Little (Eds.), *Handbook of personality development* (pp. 181–212). Mahwah, NJ: Lawrence Erlbaum.

Löckenhoff, C. E., Terracciano, A., Patriciu, N. S., Eaton, W. W., & Costa, P. T., Jr. (2009). Self-reported extremely adverse life events and longitudinal changes in five-factor model personality traits in an urban sample. *Journal of Traumatic Stress*, *22*, 53–59.

Loehlin, J. C. (1992). *Genes and environment in personality development.* Thousand Oaks: Sage Publications.

Lucas, R. E., Clark, A. E., Georgelis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science*, *15*, 8–13.

Lüdtke, O., Roberts, B. W., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: Life paths, life events, and personality trait change at the transition to university life. *Journal of Personality and Social Psychology*, *101*, 620–637.

Lüdtke, O., Trautwein, U., & Husemann, N. (2009). Goal and personality trait development in a transitional period: Assessing change and stability in personality development. *Personality and Social Psychology Bulletin*, *35*, 428–441.

Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science*, *7*, 186–189.

Mathews, A., Fox, E., Yiend, J., & Calder, A. (2003). The face of fear: Effects of eye gaze and emotion on visual attention. *Visual Cognition*, *10*, 823–835.

Matthews, G. (2004). Neuroticism from the top down: Psychophysiology and negative emotionality. In R. M. Stelmack (Ed.), *On the psychobiology of personality* (pp. 249–266). : Elsevier Ltd.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits.* New York, NY: Cambridge University Press.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*, 509–516.

McCrae, R., Costa, P., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M., et al. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, *78*, 173–186.

McGue, M., Bacon, S., & Lykken, D. T. (1993). Personality stability and change in early adulthood: A behavioral genetic analysis. *Developmental Psychology*, *29*, 96–109.

Middeldorp, C. M., Cath, D. C., Beem, A. L., Willemsen, G., & Boomsma, D. I. (2008). Life events, anxious depression and personality: A prospective and genetic study. *Psychological Medicine*, *38*, 1557–1565.

Middeldorp, C. M., Cath, D. C., van den Berg, M., Beem, A. L., van Dijck, R., & Boomsma, D. I. (2006). The association of personality with anxious and depressive psychopathology. In T. Canli (Ed.), *Biology of personality and individual differences* (pp. 251–271). : The Guilford Press.

Mroczek, D. K., & Little, T. D. (2006). Theory and research in personality development at the beginning of the 21st century. In D. K. Mroczek, & T. D. Little (Eds.), *Handbook of personality development* (pp. 3–10). Mahwah, NJ: Lawrence Erlbaum.

Mroczek, D., & Spiro, A. (2003). Modeling intraindividual change in personality traits: Findings from the normative aging study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *58*, P153–P165.

Neyer, F. J., & Asendorpf, J. B. (2001). Personality-relationship transaction in young adulthood. *Journal of Personality and Social Psychology*, *81*, 1190–1204.

Ormel, J., Oldehinkel, A. J., & Vollebergh, W. (2004). Vulnerability before, during, and after a major depressive episode: A 3-wave population-based study. *Archives of General Psychiatry*, *61*, 990–996.

Ormel, J., & Rijsdijk, F. V. (2000). Continuing change in neuroticism during adulthood-structural modelling of a 16-year, 5-wave community study. *Personality and Individual Differences*, *28*, 461–478.

Ormel, J., Rosmalen, J., & Farmer, A. (2004). Neuroticism: A non-informative marker of vulnerability to psychopathology. *Social Psychiatry and Psychiatric Epidemiology*, *39*, 906–912.

Ormel, J., & Schaufeli, W. B. (1991). Stability and change in psychological distress and their relationship with self-esteem and locus of control: A dynamic equilibrium model. *Journal of Personality and Social Psychology*, *60*, 288–299.

Ormel, J., & Wohlfarth, T. D. (1991). How neuroticism, long-term difficulties, and life situation change influence psychological distress: A longitudinal model. *Journal of Personality and Social Psychology*, *60*, 744–755.

Ormel, J., Riese, H., Bastiaansen, J., Bos, E. H., Servaas, M., Ellenbogen, M., Aleman, A. (2011). The biological and psychological basis of neuroticism: Current status and future directions. Unpublished manuscript.

Ormel, J., Riese, H., Bos, E. H., Jeronimus, B. F., Hankin B., R., J.G.M., & Oldehinkel, A. J. (2011). Neuroticism in psychopathology research: Does it have a future? Unpublished manuscript.

Pervin, L. A., & John, O. P. (1999). *Handbook of personality: Theory and research.* New York, NY: Guilford Press.

Roberts, B. W., & Chapman, C. N. (2000). Change in dispositional well-being and its relation to role quality: A 30-year longitudinal study. *Journal of Research in Personality*, *34*, 26–41.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25.

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, *17*, 31–35.

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*, 1–25.

Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 375–398). New York: Guilford Press.

Roberts, B. W., Wood, D., & Smith, J. (2005). Evaluating five factor theory and social investment perspectives on personality trait development. *Journal of Research in Personality*, *39*, 166–184.

Robins, R. W., Caspi, A., & Moffitt, T. E. (2000). Two personalities, one relationship: Both partners' personality traits shape the quality of their relationship. *Journal of Personality and Social Psychology*, *79*, 251–259.

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, *69*, 617–640.

Rutter, M. (2006). *Genes and behavior. Nature—nurture interplay explained.* Oxford UK: Blackwell Publishing.

Scarr, S., & McCartney, K. (1983). How people make their own environments — a theory of genotype–environment effects. *Child Development*, *54*, 424–435.

Scollon, C. N., & Diener, E. (2006). Love, work, and changes in extraversion and neuroticism over time. *Journal of Personality and Social Psychology*, *91*, 1152–1165.

Shankman, S. A., & Klein, D. N. (2003). The relation between depression and anxiety: An evaluation of the tripartite, approach–withdrawal and valence–arousal models. *Clinical Psychology Review*, *23*, 605–637.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, *94*, 718–737.

Steunenberg, B., Twisk, J. W., Beekman, A. T., Deeg, D. J., & Kerkhof, A. J. (2005). Stability and change of neuroticism in aging. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *60*, 27–33.

Tambs, K., Sundet, J. M., Eaves, L., Solaas, M. H., & Berg, K. (1991). Pedigree analysis of Eysenck personality questionnaire (EPQ) scores in monozygotic (MZ) twin families. *Behavior Genetics*, *21*, 369–382.

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.

Tellegen, A., & Waller, N. G. (1997). In S. Briggs, & J. Cheek (Eds.), *Exploring personality through test construction: Development of the multidimensional personality questionnaire. Personality measures: Development and evaluation*, *vol 1*. (pp. 171–248) Greenwich, CT: JAI Press.

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

Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology*, *83*, 1469–1484.

van Os, J., & Jones, P. B. (1999). Early risk factors and adult person–environment relationships in affective disorder. *Psychological Medicine*, *29*, 1055–1067.

Viken, R. J., Rose, R. J., Kaprio, J., & Koskenvuo, M. (1994). A developmental genetic analysis of adult personality: Extraversion and neuroticism from 18 to 59 years of age. *Journal of Personality and Social Psychology*, *66*, 722–730.

Watson, D., & Clark, L. A. (1984). Negative affectivity — the disposition to experience aversive emotional states. *Psychological Bulletin*, *96*, 465–490.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Watson, D., & Walker, L. M. (1996). The long-term stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology*, *70*, 567–577.

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., et al. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*, 847–854.

Widiger, T. A. (2009). Neuroticism. In M. R. Leary, & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 129–146). New York; London: The Guilford Press.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and DSM-III. *Archives of General Psychiatry*, *41*, 1005–1012.

Wray, N. R., Birley, A. J., Sullivan, P. F., Visscher, P. M., & Martin, N. G. (2007). Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Research and Human Genetics*, *10*, 695–702.

Zhang, T. Y., & Meaney, M. J. (2010). Epigenetics and the environmental regulation of the genome and its function. *Annual Review of Psychology*, *61*, 439–466.

Zuckerman, M. (1991). *Psychobiology of personality.* New York, NY: Cambridge University Press.

Zuckerman, M. (2003). In T. Millon, & M. Lerner (Eds.), *Biological bases of personality. Handbook of psychology: Personality and social psychology*, *vol. 5*. (pp. 85–116) New York: John Wiley & Sons, Inc.