



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

What personality scales measure

Citation for published version:

McCrae, RR & Mottus, R 2019, 'What personality scales measure: A new psychometrics and Its implications for theory and assessment', *Current Directions in Psychological Science*.
<https://doi.org/10.1177/0963721419849559>

Digital Object Identifier (DOI):

[10.1177/0963721419849559](https://doi.org/10.1177/0963721419849559)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Current Directions in Psychological Science

Publisher Rights Statement:

The final version of this paper has been published in Current Directions in Psychological Science, by SAGE Publications Ltd, All rights reserved. © R. McCrae & R. Mottus, 2019. It is available at:
<https://journals.sagepub.com/doi/10.1177/0963721419849559>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A New Psychometrics: What Personality Scales Measure, with Implications For Theory and Assessment

Robert R. McCrae¹
Gloucester, Massachusetts, USA

René Mõttus
Department of Psychology, University of Edinburgh, UK
Institute of Psychology, University of Tartu, Estonia

Accepted for publication to *Current Directions in Psychological Science* (23rd March 24, 2019)

Abstract

Classical psychometrics held that scores on a personality measure were determined by the trait assessed and random measurement error. A new view proposes a much richer and more complex model, including trait variance at multiple levels of a hierarchy of traits and systematic biases shaped by the implicit personality theory of the respondent. The model has implications for the optimal length and content of scales, and for the use of scales intended to correct for evaluative bias, and suggests that personality assessments ought to supplement self-reports with informant ratings. The model also has implications for the very nature of personality traits.

A typical personality scale consists of several items—statements or single adjectives. The respondent rates the extent to which each item describes a target—either oneself (in self-reports) or someone whom the respondent knows well, such as a spouse, sibling, or close friend (in informant reports). The sum of these item ratings is the target’s score on the scale, and ideally it is directly proportional to the amount of the trait the target actually possesses. We will refer to the veridical portion of the observed score as the *true score*.

Neither test developers nor respondents are perfect, so the scale score only approximates the true score. Items can be ambiguous or misread, and numerical ratings may be somewhat arbitrary; all of these introduce *random error* into the scale score. This is why people seldom receive exactly the same score when re-administered the test even a short time later. *Retest reliability* is a measure of how close retests are to the original scores.

In classical psychometrics (e.g., Lord & Novick, 1968), the term “true score” referred to the hypothetical average of an indefinitely large number of administrations of a scale to a respondent; the observed score on any one occasion was thus seen as a constant (“true”) score plus-or-minus some quantity of random error. But in fact in fact the constant component may include systematic errors or *method biases* which do not accurately reflect the trait, and thus do not contribute to the true score in our sense of that term.

Although there are many different kinds of method bias, by far the most attention has been given to evaluative biases (variously conceived as social desirability, impression management, defensiveness, faking good, or simply lying): the enduring tendency of the respondent to

exaggerate good qualities and minimize bad qualities. People vary in this tendency; some actually have negative scores and even exaggerate their weaknesses. If assessors could measure evaluative bias, they could subtract it out of the observed score and better approximate the true score. Enormous efforts have been made to do just that, but with curiously limited success (Piedmont, McCrae, Riemann, & Angleitner, 2000; Roth & Altmann, 2019).

Psychologists have relied primarily on self-reports to assess personality. With only one source of information, a model of personality scores consisting of true score, random error, and evaluative bias may seem plausible. But a body of research has also examined personality ratings from well-acquainted informants and shown that, like self-reports, these are stable, heritable, and related to a variety of life outcomes. One would be tempted to say that self-reports and informant ratings are interchangeable, except that they show far less than perfect agreement. For example, Costa and McCrae (1988) reported a six-year longitudinal study of three broad traits—Neuroticism, Extraversion, and Openness—using both self-reports and spouse ratings. Both assessments showed substantial stability, with retest correlations of about .80. But agreement between self-reports and spouse ratings was much lower, only about .55. Clearly, there must be something in self-reports that is stable over time but not shared by spouse ratings, and vice-versa. We will argue that this fact allows us to build a better model of personality scores.

One might guess that the traits being rated are actually somewhat different: Perhaps the self-report, which is based in part on private thoughts, feelings, and desires, reflects the “inner” personality, whereas the informant rating reflects the “outer.” If that were true, we would expect that two different informants would agree more strongly with each other than they agree with the target’s self-report, because both informants see the outer personality. But that is not the case; informants agree with each other only about as well as they do with self-reports (e. g., Costa & McCrae, 1992).

It appears that each respondent has a view of the target shaped in part by the trait, but also in part by systematic, enduring, and idiosyncratic biases—*source method biases*—that consistently over- or under-estimate trait levels. By examining properties of traits, we can tease apart the magnitude of these biases. For example, in the Costa & McCrae (1988) study, about 80% of the variance was stable over time, but only 55% was shared by different sources, suggesting that $80\% - 55\% = 25\%$ is stable source method bias. Discovery of the presence, structure, and magnitude of source method biases has ushered in a new view of what scales measure, with profound implications for how personality should be assessed.²

Nuances: A New Source of True Score

Most psychologists know the adage that the reliability of a scale sets an upper limit to its validity—but there are different forms of reliability. McCrae, Kurtz, Yamagata, and Terracciano (2011) compared internal consistency reliability (the extent to which all items in a scale assess the same thing) and retest reliability as predictors of the longitudinal stability, heritability, and cross-observer validity of the 30 facets of the Revised NEO Personality Inventory (NEO-PI-R; McCrae & Costa, 2010; see Table 1). More reliable facets did show higher validity by all three criteria—but only for retest reliability. Further, retest reliabilities almost invariably exceeded internal consistency reliabilities,³ suggesting that scales contain something reliable beyond their

items' shared variance.

In an attempt to understand these findings, McCrae (2015) analyzed different components of variance in facet scales and items, distinguishing between true score, method bias, and random error. It became clear that there were different kinds of true score variance: Some was shared by all the items in a scale, and some was specific to each item. For example, the NEO Angry Hostility facet has one item that concerns feelings of bitterness and resentment, and another about being hot-blooded and quick-tempered. Both are expressions of the same trait, but they are also subtly different in content; each represents a different *nuance* of Angry Hostility. Figure 1 illustrates the components of variance in a typical facet scale and shows how they contribute to various properties of the trait.⁴ This model explains why retest reliability is a better predictor of validity than internal consistency is: It contains all the true score variance ($T + s_{I-k}$), whereas internal consistency contains only the common portion (T).

Figure 1. Illustration of the variance attributable to random error (ϵ), method (M), common trait (T), and item specifics (s_{I-k}) in facet scale scores from an observer and from a self-report test and short-term retest. Dashed lines indicate distinct contributions from different items. Internal consistency (α) is attributable to method and trait variance shared by items. Cross-observer agreement (r_{CA}) and retest reliability (r_{tt}) reflect shared components of variance across observers and occasions, respectively. The bottom half of each panel represents true score; the top represents random and systematic error. Adapted by permission from R. R. McCrae, A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, 19(2), pp. 97-112. Copyright © 2015 by Robert R. McCrae. doi.org/10.1177/1088868314541857

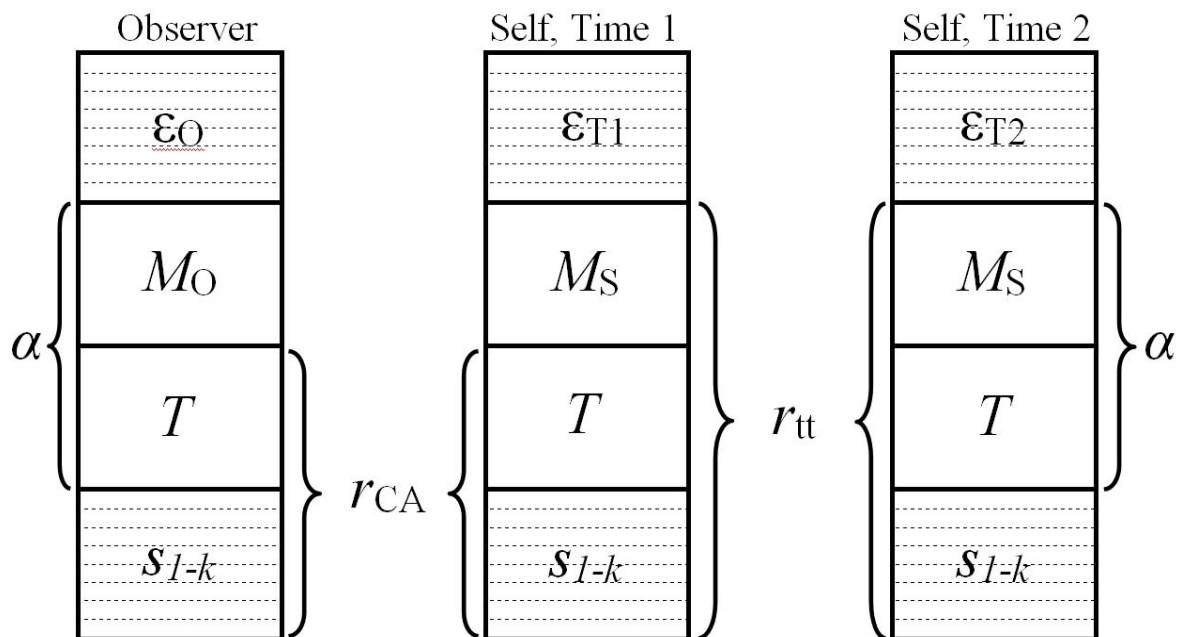


Table 1. *Domains, Definitions, and Facets of the NEO Inventories.*

Domain	Definition: <i>The tendency to . . .</i>	Facets
Neuroticism	Experience many forms of emotional distress, have unrealistic ideas and troublesome urges. Individuals low in Neuroticism are emotionally stable, do not get upset easily, and are not prone to depression.	Anxiety Angry Hostility Depression Self-Consciousness Impulsiveness Vulnerability
Extraversion	Prefer intense and frequent interpersonal interactions; be energized and optimistic. Individuals low in Extraversion are reserved and tend to prefer a few close friends to large groups of people.	Warmth Gregariousness Assertiveness Activity Excitement Seeking Positive Emotions
Openness to Experience	Seek out new experience and have a fluid style of thought. Individuals low in Openness are traditional, conservative, and prefer familiarity to novelty.	Fantasy Aesthetics Feelings Actions Ideas Values
Agreeableness	Regard others with sympathy and act unselfishly. Individuals low in Agreeableness are not concerned with other people and tend to be antagonistic and hostile.	Trust Straightforwardness Altruism Compliance Modesty Tender-Mindedness
Conscientiousness	Control one's behavior in the service of one's goals. Individuals low in Conscientiousness have a hard time keeping to a schedule, are disorganized, and can be unreliable.	Competence Order Dutifulness Achievement Striving Self-Discipline Deliberation

Note: Adapted from McCrae & Sutin, 2007. In the public domain.

It has been known for decades that personality traits form a hierarchy: The five broad domains at the top of the NEO-PI-R hierarchy are composed of narrower and more specific facets (see Table 1). McCrae's (2015) article hypothesized that there is at least one more lower level, the

nuances that express facets. But are they actually traits? If so, like higher-level traits, they should be stable over time and heritable, and different observers should agree on them. Mõttus, Kandler, Bleidorn, Riemann, and McCrae (2017) tested these hypotheses in a German sample of twins on whom self-report and informant rating data were available from two assessments five years apart. Not surprisingly, most of the items were stable, heritable, and consensually valid, because they are expressions of higher-order traits that are known to have those properties. But Mõttus and colleagues also analyzed items' nuance-specific variance after statistically removing all the higher-order variance, and a majority of these item residuals still showed the characteristic properties of traits. These findings were later replicated in samples from other countries (Mõttus et al., 2018). Nuances are real—if quite narrow—traits, and people appear to have hundreds of them.

Hierarchy in Source Method Biases

McCrae (2015) focused on components of variance within items and facets. A later examination of relations between facets and the broader domains led to further insights about source method biases (McCrae, 2018). Although a global evaluative bias might affect all traits, there are also biases specific to domains and to facets. If for some reason—say, a faulty first impression—a rater overestimates a target's Conscientiousness, the rater is likely to overestimate the target's standing on all its facets: Competence, Order, Dutifulness, and so on (see Table 1). The same rater might systematically underestimate the target's true level of Agreeableness and all its facets. Similar biases exist at the facet level, affecting all the items in a facet scale.

McCrae, Mõttus, Hřebíčková, Realo, and Allik (in press) tested these ideas in Estonian and Czech samples. They subtracted informant ratings from self-reports, essentially removing the true score component, leaving only biases and random error. When facet residuals (informant ratings minus self-reports) were factored, they showed a structure that mimicked the structure of traits seen in Table 1. When item residuals were factored, they reproduced the structure of facets within each domain. These results provided support for the hypothesis that biases have a hierarchical structure similar to that of true scores (McCrae, 2018).

The analysis of facet residuals might have shown just a single factor of evaluative bias, or two or three method factors, or no meaningful structure at all. Why should biases have the same structure as real traits? Probably because people at some level understand how traits go together; they know, for example, that if people are sociable they are also likely to be cheerful and active—they are extraverts. That people have such an implicit personality theory has been known for some time. The five-factor structure of facet-level traits emerges from ratings of strangers, in which there can be very little true score component (Passini & Norman, 1966);⁵ such structure must be attributed to the mind of the rater rather than characteristics of the target. McCrae and colleagues (in press) showed that people also have an implicit personality theory for nuance-level traits.

McCrae's (2018) model suggested that about 17% of the variance in the average facet scale was attributable to method bias at the domain level (influencing all facets in a domain), and another 22% was method bias at the facet level (influencing all items in a facet). McCrae and colleagues (in press) calculated the average domain level bias to be 18% in the Estonian sample

and 17% in the Czech sample, almost exactly in line with predictions. These consistent findings give grounds for some confidence in McCrae's (2018) model and lead to the sobering conclusion that facet level variance contains almost as much source method bias (39%)⁶ as it does the combined true score of facets and nuances (42%; the rest is random error).

Although most of the analyses discussed here have used the NEO Inventories, there is every reason to expect similar findings with other hierarchical instruments, such as Lee and Ashton's (2004)'s six-factor HEXACO model or Krueger, Derringer, Markon, Watson, and Skodol's (2012) Personality Inventory for *DSM-5* (Ashton, de Vries, & Lee, 2017; McCrae, 2018).

Theoretical Implications

It now appears that the variance in personality scales is a mixture of true score variance specific to domains, facets, and nuances; source method biases that mimic the hierarchical structure of the true score; and random error. Moreover, we can now quantify these components. This new psychometrics has both theoretical and practical implications.

- A century ago most psychologists thought the human mind was a blank slate, with personality traits shaped by enculturation, childhood experiences, and learning. We have known for the past 50 years that traits are in part heritable; now we know that this applies not only to broad temperament factors, but to specific traits even at the level of unique nuances. One person may be prone to express anger through loud outbursts, whereas another is prone to seethe with resentment, and this nuanced difference is almost as heritable as any other trait. Personality theories need to account for this.
- Personality traits are important because they predict a host of outcomes, from occupational interests to happiness. Today most researchers use broad domains as predictors, although some have advocated the use of more informative facets (Paunonen & Ashton, 2001). A program of research is needed to determine whether nuances provide yet more predictive power; initial evidence is that they do (Seeboth & Mõttus, 2018).
- The hierarchical model of trait structure has implications for how traits are conceptualized. Is Extraversion the sum (or, in the language of set theory, the *union*) of its component facets and nuances, or is it only the essential core all of its facets and nuances have in common (their *intersection*)? Union traits include both the variance shared by all components and that unique to each, whereas intersection traits exclude the unique variance; this can make a substantial difference. Personality scale scores usually sum up items and thus implicitly adopt the union model, but statistical modeling approaches such as factor analysis and item response theory almost always assume an intersection model. New statistical methods are needed to analyze union traits (Bentler, 2017).
- Nuances are natural units for, and lend credibility to, network and functionalist approaches that explore associations among numerous personality components to understand why they coalesce into broader traits such as Extraversion in the first place (Mõttus & Allerhand, 2018; Wood, Gardner, & Harms, 2015).

Implications for Research and Assessment

- Retest reliability appears to be more informative than internal consistency, and must be used to correct estimates of long-term stability (Costa, McCrae, & Löckenhoff, 2019). But researchers

usually report only internal consistency, because they can calculate it from a single administration and need not reassemble their sample a few days later to complete the scale again. Wood, Qiu, Lu, Lin, and Tov (2018) have proposed that retest reliability can also be assessed by a retest administered within the same session. If so, reporting retest reliability should become routine. Note that retest-reliability can also be calculated for single-item nuances, whereas internal consistency cannot.⁷

- To save respondents time, researchers sometimes measure broad traits with very short scales of only one or two items. Besides true score variance for the broad trait of interest, such mini-scales likely contain as much or more nuance-specific variance. If the mini-scales predict some outcome, one cannot determine whether it is because of the trait variance, the nuance variance, or both. Mini-scales should be used to measure broad traits only as a last resort, and with the acknowledgement that conclusions drawn about their correlates are tentative and must be confirmed using the full-length scale. (Mini-scales with narrow content are suitable for measuring nuances.)
- However, very long scales may be ill-advised as well. If scales contained only true score and random error, lengthening the scale would consistently improve its validity; the errors would cancel out, leaving relatively more true score (Lord & Novick, 1968). But scales also contain source method bias, which does not cancel out with increased scale length. The ratio of true score to method bias sets an upper limit to the validity of scales completed by a single source.
- Very similar items over-represent some nuances and thus bias the overall scale. Scale developers could maximize information and efficiency by including a single item for each distinct nuance that can be identified for a trait. Attention should therefore be paid to individual items' psychometric properties (e.g., retest reliability or sufficient variance).
- Efforts to measure evaluative bias and thus improve estimates of the true score have had limited success, in part because these measures almost invariably also contain true score variance; in fact, correction can sometimes reduce scale validity (Piedmont et al., 2000). But even if successful, eliminating evaluative bias would still leave domain- and facet-level biases. Single source assessments of personality are inherently limited.
- Perhaps the only practical way to improve assessment is by combining information from two or more sources. Researchers can increase the precision of their trait estimates by using multiple raters. Clinicians (e.g., Piedmont, 1998; Singer, 2005) who gather both self-reports and informant ratings of their clients—often from a spouse or partner—find a comparison of personality profiles especially revealing: Major disagreements on a trait may point to problems in the client's self-concept or to sources of conflict in the relationship. Although the use of multiple sources complicates personality assessment, it also enriches it, and should become standard practice.

Acknowledgement

Robert R. McCrae receives royalties from the NEO Inventories.

Endnotes

1. Address correspondence to Robert R. McCrae, 90 Magnolia Avenue, Gloucester, MA 01930. Email: RRMcCrae@gmail.com
2. There are sources of error in personality scale scores that are not considered in our model, such as response sets (or *transient errors*; Schmidt, Le, & Ilies, 2003), which are systematic biases on one occasion, but variable across occasions; and what Schmidt and colleagues call *specific factor errors*, such as a respondent's consistent misunderstanding of a word, which are random errors on one administration, but systematic errors across administrations. Empirical data suggest that our model works reasonably well, at least for the NEO Inventories as these are normally used. Research is needed on whether more fully elaborated models would work better with other instruments or circumstances.
3. This was not true in a study of the Big Five Inventory-2 (Soto & John, 2017), perhaps because it used an eight-week retest interval in a student sample, where true personality changes might reduce test-retest correlations.
4. This is a simplified model; for a more detailed model, see McCrae (2018).
5. However, as little acquaintance as viewing a photograph can lead to better-than-chance accuracy for some traits (Borkenau & Liebler, 1992).
6. Facet scales have a higher proportion of method bias than the domain scales they contribute to, because the bias specific to each facet tends to cancel out when facet scales are summed.
7. One-week retest data for the items of the NEO-PI-R are available in "Item retest correlations" at <https://osf.io/kcjqs/>.

References

- Ashton, M. C., de Vries, R. E., & Lee, K. (2017). Trait variance and response style variance in the scales of the Personality Inventory for *DSM-5* (PID-5). *Journal of Personality Assessment, 99*, 192–203.
- Bentler, P. M. (2017). Specificity-enhanced reliability coefficients. *Psychological Methods, 22*, 527–540.
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645–657.
- Costa, P. T., Jr., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54*, 853–863.
- Costa, P. T., Jr., & McCrae, R. R. (1992). Trait psychology comes of age. In T. B. Sonderegger (Ed.), *Nebraska Symposium on Motivation: Psychology and aging* (pp. 169–204). Lincoln, NE: University of Nebraska Press.
- Costa, P. T., Jr., McCrae, R. R., & Löckenhoff, C. E. (2019). Personality across the life span. *Annual Review of Psychology, 70*, 423–448.
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for *DSM-5*. *Psychological Medicine, 42*, 1879–1890.

- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*, 329-358.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*, 97-112.
- McCrae, R. R. (2018). Method biases in single-source personality assessments. *Psychological Assessment*, *30*, 1160-1173.
- McCrae, R. R., & Costa, P. T., Jr. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*, 28-50.
- McCrae, R. R., Mõttus, R., Hřebíčková, M., Realo, A., & Allik, J. (in press). Source method biases as Implicit Personality Theory at the domain and facet levels. *Journal of Personality*.
- McCrae, R. R., & Sutin, A. R. (2007). New frontiers for the Five-Factor Model: A preview of the literature. *Social and Personality Psychology Compass*, *1*, 423-440, 10.1111/j.1751-9004.2007.00021.x
- Mõttus, R., & Allerhand, M. (2018). Why do traits come together? The underlying trait and network approaches. In *SAGE handbook of personality and individual differences: Volume 1. The science of personality and individual differences* (pp. 130–151). Los Angeles, CA: Sage.
- Mõttus, R., Kandler, C., Bleidorn, M., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*, 474-490.
- Mõttus, R., Sinick, J., Terracciano, A., Hrebickova, M., Kandler, C., Ando, J., ... Jang, K. (2018). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability and utility of personality nuances. *Journal of Personality and Social Psychology*.
- Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology*, *4*, 44-49.
- Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of personality and social psychology*, *81*(3), 524.
- Piedmont, R. L. (1998). *The Revised NEO Personality Inventory: Clinical and research applications*. New York: Plenum Press.
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, *78*, 582-593.
- Roth, M., & Altmann, T. (2019). A multi-informant study of the influence of targets' and perceivers' social desirability on self-other agreement in ratings of the HEXACO personality dimensions. *Journal of Research in Personality*, *78*, 138-147.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, *8*, 206-224.
- Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European*

Journal of Personality, 32, 186–201.

- Singer, J. A. (2005). *Personality and psychotherapy: Treating the whole person*. New York: Guilford Press.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113, 117-143.
- Wood, D., Gardner, M. H., & Harms, P. D. (2015). How functionalist and process approaches can explain trait covariation. *Psychological Review*, 122, 84–111.
- Wood, D., Qiu, L. Lu, J., Lin, H., & Tov, W. (2018). Adjusting bilingual ratings by retest reliability improves estimation of translation quality. *Journal of Cross-Cultural Psychology*, 49, 1325-1339.

Recommended Readings

- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15, 28-50. **Reports the puzzling finding that stimulated later papers.**
- McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, 19, 97-112. **Introduces nuances and the distinction between traits as unions and intersections.**
- Möttus, R., Kandler, C., Bleidorn, M., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, 112, 474-490. **Empirical evidence that nuances are real traits, using a longitudinal, multi-observer German twin study.**
- McCrae, R. R. (2018). Method biases in single-source personality assessments. *Psychological Assessment*, 30, 1160-1173. **Proposes that method biases are hierarchical; online Supplemental materials (<http://dx.doi.org/10.1037/pas0000566.supp>) provide a tutorial on the (fairly elementary) statistical methods used to develop models of variance components.**
- McCrae, R. R., Möttus, R., Hřebíčková, M., Realo, A., & Allik, J. (in press). Source method biases as Implicit Personality Theory at the domain and facet levels. *Journal of Personality*. **Empirical evidence that method biases are hierarchical and structured by Implicit Personality Theory, using Estonian and Czech data.**