Differential Item Functioning: An Examination of the NEO-FFI by Sex in Older Adults

Marissa A. Gogniat1, Violeta J. Rodriguez1,2, Maria Granros1, Kharine R. Jean1, Talia L. Robinson1, and L. Stephen Miller1

Abstract
Understanding sex differences when assessing personality in older adults is important for researchers and clinicians. The current study utilized differential item functioning (DIF) to compare male and female older adults’ responses on the NEO-FFI to detect potential sources of measurement bias. Participants included 244 older adults (98 males, 146 females, mean age = 73). DIF by sex was determined using ordinal logistic regression and item response theory. Non-uniform DIF was present in item 31, and uniform DIF was present in item 26 in the Neuroticism scale. In the Extraversion scale, non-uniform DIF was present in items 32 and 37. In the Openness scale, non-uniform DIF was present in items 23 and 48; uniform DIF was present in items 53 and 58. Following Monte Carlo simulations to prevent overidentification, non-uniform DIF was present in item 31 in the Neuroticism scale and item 32 in the Extraversion scale. Results suggest that the NEO-FFI is a minimally biased measurement tool based on sex.

Keywords
personality, aging, Big Five, differential item functioning

Introduction
From a developmental lifespan perspective, understanding personality in older adulthood is an important but typically overlooked component. Research utilizing this perspective indicates that there are age differences and age-related changes in personality across the lifespan, which encompasses older adulthood (Allemand et al., 2007). Consistently replicated findings suggest that personality changes follow certain patterns in older adulthood (i.e., adults typically 65+ years old; Wagner & Mueller, 2020). Personality is commonly categorized into five factors (i.e., Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness; Goldberg, 1993). Research on aging and personality suggests that Neuroticism, Openness, and Extraversion tend to decline with advancing age, while Agreeableness tends to increase (Donnellan & Lucas, 2008; Mroczek & Spiro, 2003). Conscientiousness is thought to have a curvilinear relationship with age, such that it increases from birth to middle age and then decreases from middle age to older adulthood likely due to changes in cognition associated with aging (Luchetti et al., 2015; Terracciano et al., 2005). While cognitive changes across the lifespan may help account for some changes in personality, it is also necessary to recognize that environmental factors across the lifespan such as social demands and life experiences may also influence personality changes (Specht et al., 2011). This is important to consider when examining and conceptualizing personality in older adulthood. Frequently, older adults are not strongly represented in the creation of personality measures and may be underserved by current tools that assess personality (Oltmanns & Balsis, 2011; Zweig, 2008). The NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989) is a widely used measure that defines personality based on five factors (Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness). Although the NEO-FFI has been shown to be invariant in female and male adults (Gustavsson et al., 2008), recent research on the German version of the NEO-PI-R showed differential item functioning (DIF) for adult men and women, especially in the Neuroticism, Agreeableness, and Conscientiousness factors (Wetzel et al., 2013). In addition, researchers found evidence of DIF on the NEO-PI-R between younger adults and older adults in the Extraversion and Openness domains (Van den Broeck et al., 2012). However, to the authors’ knowledge, items on the
Sex Differences in Personality in Older Adulthood

Sex differences in personality exist across the lifespan, but they are studied less frequently in older adulthood. Research indicates that there are sex differences in susceptibility to neurodegenerative diseases, with women being more susceptible to diseases such as dementia and Alzheimer’s Disease than men of equivalent ages (Attarian et al., 2015). There is also some evidence of possible sex differences in neuropsychological functioning and brain volume in healthy adults (Munro et al., 2012). In addition, women are more likely to experience emotional disorders such as depression, while men are more likely to experience externalizing disorders, and this pattern continues into older adulthood (Carayanni et al., 2012; Hicks et al., 2007). Given these findings, there is a need to evaluate whether sex differences may influence how older men and women interpret, respond to, and view personality measures, especially when considering differences in gender socialization and societal expectations across the lifespan. Additionally, it is worth considering how sociocultural factors may influence how older men and women interpret statements used to assess personality. Sociocultural and socioeconomic factors including gender socialization, quality of education, and educational attainment may exert differential influences on older adult men and women, impacting how they interpret and respond to questions assessing personality.

Some important measures in psychology, including the Center for Epidemiological Studies-Depression and the Childhood Trauma Questionnaire, have been found to be non-invariant by gender (Stommel et al., 1993; Thombs et al., 2007). For these reasons previously mentioned, it would be important to examine whether older adults interpret questions on the NEO-FFI differently depending upon sex.

Current Study

Personality has been shown to influence functional independence and cognitive outcomes among older adults (Chapman et al., 2011; Luchetti et al., 2015). Personality measures are thus an important component of aging research and clinical practice. The development of personality across the lifespan is also essential to the study of interpersonal relationships, psychopathology, and health outcomes, as personality has been repeatedly shown to influence medication adherence, namely through increased conscientiousness (Christensen & Smith, 1995; Hazrati-Meimaneh et al., 2020). Sex differences in personality trajectories are of interest to aging researchers, as these differences may also affect functional independence and cognitive outcomes. However, these comparisons in personality trajectories assume measurement invariance, which has been sparsely evaluated using older populations and could lead to inaccurate sex comparisons. In addition to inaccurate sex comparisons, the presence of DIF by sex challenges the validity of the scale because it would indicate that construct-irrelevant characteristics (e.g., sex) are being captured in the scores. Given the limited research on assessment of personality in older adult populations, particularly in how these assessments may be influenced by sex, we sought to contribute to the literature by assessing DIF on the NEO-FFI. Differential item functioning refers to differences in the way a scale item functions across groups (e.g., race, sex, and class) while conditioning on the degree of the construct measured by the scale. Two forms of DIF exist: uniform DIF and non-uniform DIF. Uniform DIF refers to the functioning of a scale item as being consistently different at all levels of the construct being measured. For instance, one group may consistently endorse one item more than the other group, even if their “total scores” are the same on that particular scale. Non-uniform DIF is observed when there is an interaction between the total score on an item and group, such that at certain levels of the state or trait, one group is more likely to endorse the item, while at other levels of the state or trait, one group has a different probability of endorsing the item. By assessing DIF in the NEO-FFI, we are able to determine whether the NEO-FFI contains items biased by sex. We hypothesized that there are items on the NEO-FFI that may exhibit DIF by sex because items in any given subscale may reflect elements of gender socialization, differential susceptibility to emotional disorders, or educational quality and attainment.

Method

Participants

Participants included 244 older adults (98 males, 146 females, mean age = 73.47) from the surrounding community of a southeastern college town recruited from six independent studies assessing cognition and aging that took place approximately over the past 10 years. Participants were recruited through community flyers, newspaper advertising, and at community events. For the purposes of the present study, only data from baseline assessments were used because the NEO-FFI was only completed at baseline time points. All subjects participated in a baseline session that included neuropsychological testing and self-report questionnaires. All measures were presented by trained examiners who presented standard instructional sets. Participants had an average of 16.38 years of education, indicating that most were college educated. Most participants were subject to magnetic resonance imaging (MRI) throughout the course of their participation in the studies, and therefore all participants were only eligible if they were compatible with the MRI environment (i.e., no metal implants, no recent surgeries, etc.). In addition, participants were eligible if they had no
self-reported neurological (e.g., Alzheimer’s and Parkinson’s) or psychiatric disorders, were right-handed, and were Native English speakers. All participants were community-dwelling and were excluded from participation in studies if cognitive decline greater than mild cognitive impairment was suspected. The study protocols were approved by the Institutional Review Board at UGA and informed consent was obtained from all participants. The tenets of the Declaration of Helsinki were closely followed by all study personnel. Participants’ demographic characteristics are described in detail in Table 1.

**Measures**

**Demographic information.** Participants completed a demographic questionnaire which asked about their race, sex, age, and education.

**Personality.** All participants were administered the NEO-Five-Factor Inventory Form S (Adult; NEO-FFI; Costa & McCrae, 1989), a widely used self-report personality inventory that consists of questions that have been shown to load onto a five-factor model of personality: Neuroticism (12 items), Extraversion (12 items), Openness (12 items), Agreeableness (12 items), and Conscientiousness (12 items). The NEO-FFI consists of 60 questions using a 5-point scale ranging from “Strongly Disagree” (1) to “Strongly Agree.” (5).

**Statistical Analysis**

Univariate analyses (i.e., frequencies, means, and standard deviations) were used to describe the sample. The internal consistency of the NEO-FFI was evaluated using Mcdonald’s ω coefficient. Because the minimum cell size count was set to five to facilitate model estimation and convergence (Choi et al., 2011), the Strongly Disagree and Disagree categories, as well as the Agree and Strongly Agree categories were combined into a single category, resulting in three possible categories (1) Strongly Disagree and Disagree, (2) Neutral, and (3) Agree and Strongly Agree per recommendations by the developers of the package (Choi et al., 2011). Items 4, 17, 20, 21, 34, 35, 40, and 49 did not meet the minimum criteria of five respondents per category and were excluded from analyses (Choi et al., 2011).

Differential item functioning by sex was tested in R (Version 3.5.1; http://www.R-project.org) using the lordif package (Choi et al., 2011), which utilizes ordinal logistic regression as well as item response theory approaches to test DIF under the graded response model, as noted by model comparisons of the models below:

Model 0: \[ \logit P(u_i \geq k) = \alpha k \]

Model 1: \[ \logit P(u_i \geq k) = \alpha k + \beta_1 \times \text{NEO-FFI} \]

Model 2: \[ \logit P(u_i \geq k) = \alpha k + \beta_1 \times \text{NEO-FFI} + \beta_2 \times \text{Sex} \]

Model 3: \[ \logit P(u_i \geq k) = \alpha k + \beta_1 \times \text{NEO-FFI} + \beta_2 \times \text{Sex} + \beta_3 \times \text{NEO-FFI} \times \text{Sex} \]

As previously noted, uniform DIF, which indicates consistent item performance across all score groups of the NEO-FFI, is denoted by a significant difference between Models 1 and 2 at \( \alpha = .01 \) using −2 likelihood ratio chi-square tests. Non-uniform DIF, a more problematic form of DIF, reveals a probability of responding that is not constant across score groups of the NEO-FFI; non-uniform DIF is denoted by a statistically significant difference between Models 2 and 3 at \( \alpha = .01 \).

While likelihood ratio tests provide reasonable Type I error control in detecting DIF (Kim & Cohen, 1998), an additional step was undertaken to control Type 1 errors by using Monte Carlo simulations (Choi et al., 2011). For this reason and to account for potential Type 1 errors due to the number of −2 likelihood ratio tests, empirical thresholds to detect DIF were calculated using Monte Carlo simulations in DIF-free samples (\( \alpha = .01 \); 1,000 replications). The highest identified empirical threshold in the Monte Carlo simulations was used to detect uniform DIF and non-uniform DIF (Choi et al., 2011). Item and test characteristic curves were plotted for items and subscales that exhibited DIF using the lordif package (Choi et al., 2011).

**Results**

**Demographic Characteristics of Participants**

Participants were, on average, 73.47 years old (SD = 6.95). Nearly two-thirds (60%) of participants were female. Almost all (97%) of participants identified as White European, 2% as African American, and 1% combined as Hispanic, Multi-racial, and Pakistani. As stated earlier, average years of educational attainment was 16.38 (SD = 2.79). These demographic characteristics can be viewed in Table 1.

**Reliability**

For each of the factors, reliability was \( \omega = .85 \) for Neuroticism, \( \omega = .74 \) for Extraversion, \( \omega = .74 \) for Openness, \( \omega = .71 \) for Agreeableness, and \( \omega = .78 \) for Conscientiousness.

| Variable | % or M (SD) |
|----------|-------------|
| N        | 244         |
| Age (years) | 73.47 (6.95) |
| Sex (% female) | 59.8 |
| Race (% Caucasian) | 96.7 |
| Education (years) | 16.38 (2.79) |
Differential Item Functioning by Sex: Neuroticism

Two items (items 26 and 31) of the NEO-FFI Neuroticism scale were flagged for DIF by sex at $\alpha = .01$. The model comparison for Models 2 versus 3 showed that non-uniform DIF was present in item 31 at $p < .001$, with more men endorsing the item at low levels of theta (Neuroticism), but more women endorsing the item at high levels of theta (Neuroticism). The model comparison for Models 1 versus 2 showed that uniform DIF was present in item 26 at $p < .01$. The item true score functions by sex for items 26 and 31 and the test characteristic curves for all items and for DIF items only in the Neuroticism scale are presented in Figure 1; these suggested that the magnitude of DIF was greater at higher levels of Neuroticism.

Monte Carlo simulations were conducted to determine the highest threshold of DIF detection in the Neuroticism scale. The highest Monte Carlo simulation-generated empirical threshold derived from DIF-free samples was McFadden $R^2 = .0415$ for uniform DIF, and $R^2 = .0217$ for non-uniform DIF. According to these thresholds, only item 31 exhibited non-uniform DIF ($R^2 = .0348$).

Differential Item Functioning by Sex: Extraversion

Two items (32 and 37) of the NEO-FFI Extraversion scale were flagged for DIF by sex at $\alpha = .01$. The model comparison for Models 2 versus 3 showed that non-uniform DIF was present in item 32 and 37 at $p < .001$. For item 32, fewer women endorsed the item at low levels of theta (Extraversion), but more women endorsed the item at high levels of theta (Extraversion). For item 37, women endorsed the item more frequently at low levels of theta (Extraversion), but more men endorsed the item at high levels of theta (Extraversion). The item true score functions by sex for items 32 and 37 are presented in Figure 2. The test characteristic curves for all items and for DIF items only in the Extraversion scale are also presented in Figure 2; these suggested that the magnitude of DIF was greater at lower levels of Extraversion.

Monte Carlo simulations were also used to determine the highest threshold of DIF detection in the Extraversion scale. The highest Monte Carlo simulation-generated empirical threshold derived from DIF-free samples was McFadden $R^2 = .0515$ for uniform DIF, and $R^2 = .0300$ for non-uniform DIF. Using these thresholds, no items were flagged for DIF in the Openness scale.

Differential Item Functioning by Sex: Agreeableness and Conscientiousness

No items were identified for DIF in the NEO-FFI Agreeableness and Conscientiousness subscales via $-2$ likelihood ratio tests or Monte Carlo simulations. The test characteristic curves for these scales are depicted in Figure 4.

Discussion

Given the limited research on assessing personality in older adults, it was important to examine whether the NEO-FFI, a widely used personality measure, contains items that could be interpreted and answered differently between female older adults and male older adults. Our study determined that items that compose the Neuroticism, Extraversion, and Openness scales displayed non-uniform DIF, while items in the Agreeableness and Conscientiousness scales did not. Following Monte Carlo simulations, non-uniform DIF was only present in item 31 in the Neuroticism scale and item 32 in the Extraversion scale. While uniform DIF can be corrected with mathematical adjustments, non-uniform DIF can typically only be corrected when those items are removed from the scale (Pallant & Tennant, 2007).

In this study, we found two items from the Neuroticism scale that displayed DIF, but only one item (Item 31: “Rarely feel fearful/anxious”) that showed non-uniform DIF by sex in the Neuroticism subscale following Monte Carlo simulation, indicating that older adult women were more likely to disagree with this statement at higher levels of Neuroticism and older adult men were more likely to agree with the statement at higher levels of Neuroticism. At lower levels of Neuroticism, an opposite pattern emerged, with older women...
endorsing this item more frequently than older men. It is possible that higher prevalence rates of emotional disorders (Carayanni et al., 2012; Hicks et al., 2007) in older adult women may have contributed to the non-uniform DIF seen in this item, in addition to sociocultural variables that may have influenced responses to this question (i.e., increased anxiety as a byproduct of the prevalence of violence against women throughout the lifespan; Tjaden, 2000). Widely used psychological measures, including the Center for Epidemiological Studies-Depression and the Childhood Trauma Questionnaire have been found to be non-invariant by sex (Stommel et al., 1993; Thombs et al., 2007). This lack of invariance is hypothesized to be influenced by socioeconomic differences by gender, which may, in turn, influence how items are generated and responded to by men and women. While two items under the Extraversion scale were flagged for DIF, only one item (Item 32: “Often bursting with energy”) displayed non-uniform DIF by sex following Monte Carlo simulation. Older adult women were more likely to disagree with these statements at lower levels of Extraversion, while older adult men were more likely to agree with these statements at lower levels of Extraversion. It may be that these statements are

Figure 1. Category characteristic curves and test characteristic curves for Neuroticism for all items and items flagged for DIF prior to Monte Carlo simulations. Item 26 depicts uniform DIF while item 31 depicts non-uniform DIF. Only item 31 remained flagged for DIF following Monte Carlo simulations.
also capturing the idea that older adult women are more likely to experience emotional disorders such as depression than are older adult men, and as such, older women may display decreased levels of positive affect and energy, which may be exhibited as low Extraversion (Carayanni et al., 2012). The Openness scale had four items flagged for DIF, but these items did not survive Monte Carlo simulations indicating that this scale does not likely contain items that are biased by sex. While the Agreeableness and Conscientiousness scales did not have any items with DIF, it is important to note that six out of the eight items that were removed for not meeting the minimum criteria of five respondents per category fell under these two factors. Although this is an important limitation, it should also be noted that the low endorsement of these items suggests that these items may not be characteristic of older adults or that older adults may be unwilling to endorse these statements. Therefore, these items may be irrelevant to the measurement of personality constructs among older adults, and this should be explored and replicated in future research.

**Limitations and Future Directions**

The items exhibiting DIF may not be representative of Neuroticism, Extraversion, Openness, Agreeableness, or Conscientiousness and may actually represent a different,
sex-related construct. It is possible that individual and group level conclusions based on these scores may be biased. Therefore, this study also contributes to a growing literature on gender bias on widely used psychological measures (i.e., Stommel et al., 1993; Thombs et al., 2007). For example, older adult women’s individual scores may be misinterpreted when compared to population means, (i.e., older adult women may be labeled as “Neurotic” based on measurement bias alone, rather than actual scores on Neuroticism) while older adult men’s scores on this same measure would not be misinterpreted in this way. Future research should utilize a larger, more representative sample to ensure that results generalize to older adults from underrepresented backgrounds. Given that this sample was predominantly white and highly educated, it would be important to examine this same question in a more diverse sample. Future studies would also benefit from expanding upon this study with other commonly utilized personality measures. It is also important to again note that eight items did not meet the minimum criteria of five respondents per category and six of these came from two subscales (Agreeableness and Conscientiousness). This indicates that older adults answered these questions very similarly (i.e., there is minimal variability in responses); thus, these items may add little additional information and their usefulness in this measure for older adults may be limited. The removal of these items primarily from just two subscales lessened the number of items in the Agreeableness and Conscientiousness scales that could be used to test for DIF, and thus is a limitation of this study. It may be important to consider incorporating additional items with greater response variability in the Agreeableness and Conscientiousness subscales.

To the authors’ knowledge, this was the first study to examine differential item functioning of the NEO-FFI by sex in older adults, making an important contribution to the limited research on examining the assessment of personality in older adulthood. It is important to note that only two items displayed non-uniform DIF when Monte Carlo simulations were used. Given the relatively modest sample size for this study, it is not surprising that using an additional method to control Type 1 error beyond likelihood ratio tests resulted in

![Figure 3](image-url)  
**Figure 3.** Category characteristic curves for items identified for DIF in the Openness subscale. Items 23 and 48 depict non-uniform DIF, while items 43 and 58 depict uniform DIF. No items were flagged for DIF following Monte Carlo simulations.
fewer items identified with DIF by sex (Choi et al., 2011). Because of sample size limitations, it is important for future research to replicate these findings and to examine DIF by sex in larger samples of older adults. The results from this study have implications for the use and interpretation of the NEO-FFI in older adults in both a clinical setting and when

Figure 4. Test characteristic curves for Openness, Agreeableness, and Conscientiousness for all items and items flagged for DIF prior to Monte Carlo simulations.
used in research studies, suggesting that this measure is minimally biased by sex.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: VJR’s work on this manuscript was funded by a Ford Foundation Fellowship to Violeta J. Rodriguez, administered by the National Academies of Sciences, Engineering, and Medicine, and a PEO Scholar Award from the PEO Sisterhood.

ORCID iD
Marissa A. Gogniat https://orcid.org/0000-0001-8331-693X

References
Allemand, M., Zimprich, D., & Hertzog, C. (2007). Cross-sectional age differences and longitudinal age changes of personality in middle adulthood and old age. *Journal of Personality*, 75(2), 323–358. https://doi.org/10.1111/j.1467-6494.2006.00441.x

Attarian, H., Brandes, J., Dafer, R., Gerard, E., & Giesser, B. (2015). Sex differences in the study of neurological illnesses. *Behavioural Neurology*, 2015, 676531. https://doi.org/10.1155/2015/676531

Carayannii, V., Stylianopoulou, C., Koulierakis, G., Babatsikou, F., & Koutsis, C. (2012). Sex differences in depression among older adults: Are older women more vulnerable than men in social risk factors? The case of open care centers for older people in Greece. *European Journal of Ageing*, 9(2), 177–186. https://doi.org/10.1007/s10433-012-0216-x

Chapman, B. P., Roberts, B., & Duberstein, P. (2011). Personality and longevity: Knowns, unknowns, and implications for public health and personalized medicine. *Journal of Aging Research*, 2011, 759170. https://doi.org/10.4061/2011/759170

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *Journal of Statistical Software*, 39(8), 1–30.

Christensen, A. J., & Smith, T. W. (1995). Personality and patient adherence: Correlates of the five-factor model in renal dialysis. *Journal of Behavioral Medicine*, 18(3), 305–313.

Costa, P. T., & McCrae, R. R. (1989). NEO-PI/NEO-FFI manual supplement (p. 40). Psychological Assessment Resources.

Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging*, 23(3), 558–566. https://doi.org/10.1037/a0012897

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, 48(1), 26.

Gustavsson, J. P., Eriksson, A.-K., Hilding, A., Gunnarsson, M., & Östensson, C.-G. (2008). Measurement invariance of personality traits from a five-factor model perspective: Multi-group confirmatory factor analyses of the HP5 inventory.

Hazen, J., & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement*, 22(4), 345–355.

Luchetti, M., Terracciano, A., Stephan, Y., & Sutin, A. R. (2015). Personality and cognitive decline in older adults: Data from a longitudinal sample and meta-analysis. *The Journal of Gerontology: Series B, Psychological Sciences and Social Sciences*, 71, 591–601. https://doi.org/10.1093/geronb/gbu184

Mroczek, D. K., & Spiro, A. (2003). Modeling intraindividual change in personality traits: Findings from the normative aging study. *The Journal of Gerontology: Series B*, 58(3), P153–P165. https://doi.org/10.1093/geronb/58.3.P153

Munro, C. A., Winicki, J. M., Schretlen, D. J., Gower, E. W., Turano, K. A., Muñoz, B., Keay, L., Bandeen-Roche, K., & West, S. K. (2012). Sex differences in cognition in healthy elderly individuals. *Aging, Neuropsychology, and Cognition*, 19(6), 759–768. https://doi.org/10.1080/13825585.2012.690366

Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: Questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology*, 7, 321–49. https://doi.org/10.1146/annurev-clinpsy-090310-120435

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46(1), 1–18. https://doi.org/10.1348/014466506X96931

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the big five. *Journal of Personality and Social Psychology*, 101, 862.

Stommel, M., Given, B. A., Given, C. W., Kalaian, H. A., Schulz, R., & McCorkle, R. (1993). Gender bias in the measurement properties of the center for epidemiologic studies depression scale (CES-D). *Psychiatry Research*, 49(3), 239–250. https://doi.org/10.1016/0165-1781(93)90064-N

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T. (2005). Hierarchical linear modeling analyses of NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging*, 20, 493–506. https://doi.org/10.1037/0882-7974.20.3.493

Thombs, B. D., Lewis, C., Bernstein, D. P., Medrano, M. A., & Hatch, J. P. (2007). An evaluation of the measurement equivalence of the Childhood Trauma Questionnaire—Short Form.
across gender and race in a sample of drug-abusing adults. Journal of Psychosomatic Research, 63(4), 391–398. https://doi.org/10.1016/j.jpsychores.2007.04.010

Tjaden, P. G. (2000). Full report of the prevalence, incidence, and consequences of violence against women: Findings from the National Violence Against Women survey. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Van den Broeck, J., Rossi, G., Dierckx, E., & De Clercq, B. (2012). Age-neutrality of the NEO-PI-R: Potential Differential Item Functioning in older versus younger adults. Journal of Psychopathology and Behavioral Assessment, 34(3), 361–369. https://doi.org/10.1007/s10862-012-9287-4

Wagner, J., & Mueller, S. (2020). Personality development in late adulthood. In V. Zeigler-Hill & T. K. Shackelford (eds) Encyclopedia of personality and individual differences. Springer. https://doi.org/10.1007/978-3-319-24612-3_1877

Wetzel, E., Böhnke, J. R., Carstensen, C. H., Ziegler, M., & Ostendorf, F. (2013). Do individual response styles matter? Journal of Individual Differences, 34, 69–81. https://doi.org/10.1027/1614-0001/a000102

Zweig, R. A. (2008). Personality disorder in older adults: Assessment challenges and strategies. Professional Psychology: Research and Practice, 39(3), 298–305. https://doi.org/10.1037/0735-7028.39.3.298