**Supplemental Section**

Response options and reliability

Improvements in internal consistency reliability using a personality questionnaire (Cox et al, 2017) and psychopathology measures (Finn et al., 2015), have been observed when the number of response options is increased. Using Monte Carlo methodology, Lozano et al. (2008) simulated responses to 30 items with correlations between them ranging from 0.2 to 0.9. The number of response options employed ranged from two to nine. The results show that as the number of response alternatives increases, both reliability and factorial validity improve. The optimum number of alternatives was between four and seven. With fewer than four response options, the reliability and validity decreased. However, in some instances, despite showing increased reliability when moving beyond dichotomous options, the differences can be quite modest (Finn et al, 2014; Grassi et al., 2007). Grassi et al. examined a popular functional status measure (SF-36) and found modest gains in internal consistency, i.e., for polytomous responses (range 0.55–0.76) verse dichotomous responses (range for 0.52–0.72).

Estimates using modern test theory (allows for the individual person estimates of precision, rather than classical sample averages of reliability) have shown less modest differences. For example, Maydeu-Olivares et al. (2009), using a personality metric, observed that the area under the test information function in the latent trait range increased by 33% when the number of response options was increased from two to three, and by 70% when the number of options was increased from three to five. Using a life-satisfaction scale, the area under the test information function in the latent trait interval increased by 37% when the number of options was increased from two to three, and by 29% when the number of options was increased from three to five. More recently, also employing modern test theory methodology, Lee and Paek (2014) confirmed that, when the number of response categories was reduced from 3 to 2 (across several measures) reliability estimates were reduced; reliability coefficients increased from .67 (dichotomous) to .84 (polytomous). It is noteworthy that the authors further indicated that the negative impact on precision could be mitigated by altering scale length or item discrimination power.

Response options and validity

Broadly speaking, Nunnally and Bernstein (1994) proposed that low internal consistency (i.e., alpha below .70) indicates poor reliability, but also implies poor predictive validity. This association with validity is supported by Preston and Colman (2000), which reported that criterion validity was best with 5 or more categories, with the criterion validity coefficient improving from .83 (dichotomous option) to .89 with polytomous options. However, Maydeu-Olivares observed a negligible effect of the number of response categories on predictive validity. Finally, McCrae et al. (2011) noted that, if two measures of, for example, achievement motivation are both used to predict grade point average in the same sample, it is the measure with higher retest reliability (not internal consistency) that ought—other things being equal—to yield stronger predictive validity. This assertion, that increased response options improves test-retest reliability (metric of stability), was previously supported by Weng et al. (2004).

**References**

Cox A, Pant H, Gilson AN, Rodriguez JL, Young KR, Kwon S, & Weed NC. (2012). Effects of augmenting response options on MMPI–2 RC scale psychometrics. *Journal of Personality Assessment*. 2012; 94: 613–619. doi:10.1080/00223891.2012.700464.

Finn JA, Ben-Porath YS, Tellegen A. Dichotomous versus polytomous response options in psychopathology assessment: Method or meaningful variance? *Psychological Assessment*. 2015; 27:184–193.

Lozano, LM, García-Cueto E, Muñiz J. Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences.* 2008;4(2): 73-79.

<http://dx.doi.org/10.1027/1614-2241.4.2.73>

Finn JA, Ben-Porath YS, Tellegen A. Dichotomous versus polytomous response

options in psychopathology assessment: method or meaningful variance? *Psychol*

*Assess.* 2015;27(1):184-93. doi: 10.1037/pas0000044. Epub 2014 Nov 3. PubMed PMID: 25364870.

Grassi M, Nucera A, Zanolin E, Omenaas E, Anto JM, Leynaert B; European

Community Respiratory Health Study Quality of Life Working Group. Performance

comparison of Likert and binary formats of SF-36 version 1.6 across ECRHS II

adults populations. *Value Health*. 2007;10(6):478-88. PubMed PMID:17970930.

Maydeu-Olivares A, Kramp U, García-Forero C, Gallardo-Pujol D, Coffman D. The

effect of varying the number of response alternatives in rating scales: experimental evidence from intra-individual effects. *Behav Res Methods*. 2009;41(2):295-308. doi: 10.3758/BRM.41.2.295. PubMed PMID: 19363170.

Lee & Paek. In search of the optimal number of response categories in rating scale. *Journal of Psychoeducational Assessment* . 2014; 32 (7): 663 – 673.

Nunnally JC, Bernstein I. Psychometric theory. 3rd ed.. McGraw-Hill; New York: 1994.

Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol (Amst).* 2000;104(1):1-15. PubMed PMID: 10769936.

McCrae RR, Kurtz JE, Yamagata S, Terracciano A. Internal consistency, retest reliability, and their implications for personality scale validity. *Pers Soc Psychol Rev*. 2011;15(1):28-50. doi: 10.1177/1088868310366253.

Weng, Li-Jen. **Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability.** Educational and Psychological Measurement 2004. **64;**956–972. 10.1177/0013164404268674.