SUBMISSION TYPE

Poster

TITLE

Patterns of Missing Data in Personality Ratings by Self and Informants

ABSTRACT

PRESS PARAGRAPH

WORD COUNT

(up to 3,000)

Researchers have long grappled with the challenge of missing data results from a wide range of sources (e.g., participant attrition, skipping over questions, or subscales, administration or data entry errors). Although these sources of missing data are commonplace in psychological studies, they are particularly vexing in studies that rely on multiple informants to provide information about a target (Blozis et al., 2013; Phares, 1992). The focus in the current paper is on exploring patterns of missing data across different rating sources of a target (i.e., from the target, and two informant raters, one of which was not selected by the target him or herself). Here, we predict that the desirability of personality item itself will predict the likelihood of “skipping” a question about oneself, but also among targets rating someone they know well. Moreover, we suggest people may be more likely to provide a non-response to an item with a more negative connotation. This pattern in the data would suggest that the data are not missing completely at random, but rather point toward a systematic bias – one that calls for our attention in the area of personality assessment.

**Missingness Mechanisms: Reasons for Missing Data**

There are three ways that data can be missing. First, missing data can be completely at random (MCAR), which means there are not systematic differences between observed and missing data. Stated another way, “missingness” is entirely independent of the observed and the missing data. Data can also be missing at random (MAR), in which missingness is independent of the missing data, but it is dependent on the observed data. Blozis et al. (2013) provide a clear example of how this might occur. If studying father reports of marital satisfaction and some are missing, they could be “missing at random,” for fathers with low levels of self-esteem who are less likely to report their marital satisfaction. Thus, those missing data depend not on someone’s marital satisfaction, but rather their self-esteem. Data can also be missing not at random (MNAR). Here, missingness is tied to the missing data even after accounting for dependencies of the missingness on the observed data.

**How to Address Missing Data**

Exploring or examining data for missingness is a continued challenge. Researchers have utilized a variety of methods to explore (e.g., complete case analysis) and manage it (e.g., inverse probability weighting, multiple imputation). Full information maximum likelihood (FIML) estimation in structural equation models is often used where some of the data from multiple informants are missing (Cui et al., 2010). Each of these choices makes assumptions about the missingness of the data. When they are missing for some systematic reason(s) - MNAR, the source of this issue cannot be downplayed, but rather, need to be addressed to prevent the potential of biasing results and conclusions (Little & Rubin, 2019).

From this point forward, need to paraphrase and condense and tailor a bit. This is directly from our submitted paper.

Self-report measures of psychological constructs are commonly criticized for their vulnerability to response biases (e.g., Spector, 2006). These perspectives regarding self-evaluations are likely not surprising to the reader who is familiar with the issue of social desirability in personality assessment (e.g., Asendorpf & Ostendorf, 1988; Dilchert et al., 2006; Edwards, 1957; Paunonen & LeBel, 2012) or common method variance across measurement domains (e.g., Kammeyer-Mueller et al., 2010; Podsakoff et al., 2003). Many researchers propose that by collecting informant ratings, they circumvent some of the psychometric concerns traditionally associated uniquely with self-ratings (see, for example, Chang et al., 2012; Kolar et al.,1996; Podsakoff et al., 2012; Robins & John, 1997). Although by tradition these informants are sought as sources of objectivity, their susceptibility to similar inflation prompts as self-raters is, in fact, unknown (and perhaps more dangerously, on occasion *assumed* to be different than self-raters; e.g., Chang et al., 2012; Podsakoff et al., 2003). The current study explores whether or not these biases have a systematic influence on which types of questions respondents choose not to answer.

**Theoretical Basis: Electing Non-response**

Alicke and Sedikides (2011) documented what they have termed *self-enhancement* and *self-protection* motives that pervade healthy conceptualizations of the self. Self-enhancement refers to “a tendency to claim greater standings on a characteristic, or more credit, than is objectively warranted” (p. 2). Self-protection, on the other hand, refers to “tactics that are adopted to avoid falling below a desired standard” (p. 2). These motives suggest that a target’s view of their own behavior may be reasonably expected to be positively biased. Alicke and Sedikides (2011) additionally, however, theorize that these motives also extend to others in whom one is invested (e.g., friends and family). This theoretical foundation regarding the way in which people, whether consciously or unconsciously, view themselves also therefore frames how we might view others in whom we have some personal investment. We next consider how these self-enhancement and self-protection motives might complicate our efforts to accurately measure personality via questionnaire administration.

**The Use of Informant Ratings in Personality Assessment**

The historical focus on the detection and mitigation of method effects in psychological assessment grew in magnitude with Campbell and Fiske’s (1959) MTMM approach and others have continued to broaden the scope of investigation since (e.g., Kammeyer-Mueller et al., 2010; Podsakoff et al., 2012; Podsakoff et al., 2003). Within these psychometric frameworks, non-self raters (who throughout the paper we refer to as “informants”) may be considered unique and/or useful sources of information – primarily in the elucidation of measurement artifacts. Many studies, in fact, have leveraged the informant rater methodology in an attempt to bypass the widely acknowledged measurement deficiencies in self-reports (e.g., Berry et al., 2012; Borkenau, & Ostendorf, 1989; Connelly & Ones, 2010; Connolly et al., 2007; John & Robins, 1993; Kammeyer-Mueller et al., 2010; Kolar et al., 1996; Leising et al., 2015; Letzring et al., 2006; Mount et al., 1994; Oh et al., 2011; Paunonen, & Kam, 2014). Hofstee (1994) has suggested that aggregated judgments provided by multiple informants offer the *best* assessment of a target, and others have confirmed that there is convergence across raters in their judgement of others’ personalities (e.g., Connelly & Ones, 2010). These orientations that hint at the utility of informant ratings, however, are dependent on self and informants having *different* motives, or, more practically, *not* being susceptible to the same evaluative influences.

Within these psychometrically-oriented investigations, the use of informants is a frequently recommended safeguard against artifactually-inflated inter-measure associations (see, for example, Chang et al., 2012; Podsakoff et al., 2012 although see also Kammeyer-Mueller et al., 2010 for a counter-perspective). Podsakoff et al. (2003), for example, recommend obtaining measures of independent and dependent variables from different sources to prevent any one individual rater from introducing associations due to intra-person factors such as consistency motifs or mood states. In line with this recommendation, Chang et al. (2012) applied meta-analysis to informant ratings, treating these sources as distinct methods of assessment. Conway and Lance (2010), however, argue quite strongly that researchers (and reviewers) have often misconstrued rating *source* as a measurement *method*. They further contend that informants/rating sources may, in fact, be prone to the same biases as self-raters – we aim to explicitly evaluate this possibility in the current study.

**Informants versus Selves – Divergent or Comparable Response Orientations**

As mentioned above, it is currently unknown whether targets and informants are susceptible to the same biasing influences in their responses. Theory suggests that people have an unconscious desire to present themselves in the way they would like others to perceive them (Hogan & Blickle, 2018). Not unexpectedly, similar motives, when considered from the perspective of fellow members of one’s “in group” (social identity theory; Tajfel & Turner, 1979), would compel us to view those in our inner circles in a positive light – thus, seeing and being seen through “rose-colored glasses” (e.g., Murray et al., 1996). However, a contrasting perspective regarding methodologies to separate “substance” vs. “style” elements in the personality and common method variance literatures argues that the use of informant ratings is a viable strategy to mitigate measurement artifacts.

Tilting the debate a bit, there does seem to be disparately situated yet converging evidence that suggests that informants may be biased in their rating of known targets. In addition to theoretical orientations such as that posed by Alicke and Sedikides (2011) as well as social identity theory (Tajfel & Turner, 1979), there exists a history of consideration, for example, from within the job performance domain. Self-ratings of job performance or organizational citizenship behaviors have long been criticized within the organizational literatures as representing a methodological flaw. Subsequently, many investigations and meta-analyses have investigated the comparative equivalence of self- and informant-ratings of these constructs, with the implication being that self-informant agreement represents evidence that self-ratings are “ok” (see, for example, the meta-analyses of Carpenter et al., 2014 and Heidemeier & Moser, 2009). That is, rating alignment is taken as a comfort. This implication of course depends on the quality of the comparative standard. Alternatively, evidence of equivalence could indicate that the rating sources are merely susceptible to similar flaws.

In fact, others *have* acknowledged this potential limitation in using informant ratings. Researchers for some time now have suggested methods by which we might tease apart ratings of targets’ actual characteristics and perceivers’ evaluation of the positivity or negativity of item/characteristic (e.g., Leising et al., 2015; Peabody, 1967). Within other applications, researchers have documented that respondents provide socially desirable ratings of both significant others (e.g., Brown & Han, 2012), and children (e.g., Lench et al., 2006). This tendency to rate others that we like more favorably has in some contexts been referred to as the “letter of recommendation problem” (Klonsky et al., 2002). This potential bias in *liking* is distinct from simply *knowing* a target – Leising and colleagues (2010), in fact, demonstrated that informants who likedtheir targets more also rated them more positively. Furthermore, Wessels et al. (2018) showed that liking targets was associated with greater positivity bias, and knowing was associated with less positivity bias. We note here that standard assessment protocols in which the target identifies potential additional informants would seem to be particularly susceptible to such biases.

**Current Study and Hypotheses**

H1: Item social desirability negatively predicts non-response.

H2: We have 3 candidates to explain missingness – two of which are traditional indices: 1) reading grade level of the item (12 – round to nearest grade level), 2) presence or absence of infreuquent vocabulary (min to max??), 3) novel conribution is hwo much item is saturatuated with item desirability (8 of them). Expecatation is that the strongest effect will be for the Edwards category.

H?: Prediction about differences based on relation to the self? Less likely to miss as get further away from target?

H?: Based on the additional analysis using quanteda package – not sure on detail here.

Food for thought – I copied this directly from authors’ discussion section so delete this: “When data are NMAR, information about the missing data process should be incorporated

into an analysis to avoid biased parameter estimates (Little & Rubin, 2002). Strategies for

dealing with data that are NMAR may rely on known analytic procedures, including mean

and covariance structure analysis, but may also include a model for the missingness. Major

frameworks in this area include selection models and pattern-mixture models (Demirtas &

Schafer, 2003; Kenward, 1998; Little, 1995; Xu & Blozis, 2010). A major complication in

the estimation of NMAR models concerns the identification of parameters that

correspondence to the dependence of the missingness on the missing data (Little & Rubin,

2002). These models also require that a model for the missing data process be explicitly

defined, an added challenge as the process that generated the missing data in a given

problem is not likely to be known.” (Blozis et al., 2013, p. 11).

**Method**

**Participants and Procedure**

Eight hundred and eighty-eight undergraduate students from 3 US universities were recruited via an online research participation website in exchange for course credit. Upon completion of personality and demographic questions, participants were asked to provide the email address of someone else who could make ratings of their personality. An invitation was automatically generated to the person identified, and this “once-removed” rater was then asked to rate the personality of the original participant using the same item set. Once more, this rater was asked to provide the email address of a second person to make ratings of the original participant (we refer to this source as “twice-removed”). From our 888 original student participants, we realized 460 once-removed raters and 131 twice-removed raters for an experiment-wide totalof 1,479 respondents. These decreasing sample sizes were expected based on the nature of the nomination process (e-mail invitation) and lack of researcher control regarding opportunities for response rate enhancement. Conservative screening of the data resulted in 306 raters matched across self, once-removed, and twice-removed sources (all *n*’s = 102)[[1]](#footnote-1). Upon survey completion, participants were thanked and compensated via course credit (original student) or the opportunity to take part in a drawing for one of three gift cards (once- and twice-removed raters).

**Materials**

*Item Social Desirability*. Ratings of the social desirability content of 2,413 personality indicators from the International Personality Item Pool (http://www.ipip.org) were made along a 9-point scale ranging from: (-4) *extremely undesirable* to (+4) *extremely desirable* (see, for example, Edwards, 1957). Our graduate student raters (*n* = 6) were provided with definitions of both ends of the social desirability continuum prior to making ratings. Inter-rater consistency estimates using a 2-way random model was .96.

*Personality measure(s).* Four stratified random samplings of the 2,413 items were conducted. The strata were defined via the 9-point rating scale: stratum #1 = social desirability values below -3.01, stratum #2 = -3 to -2.01, stratum #3 = -2 to -1.01, stratum #4 = -1 to -.01, stratum #5 = 0 to 1, stratum #6 = 1.01 to 2, stratum #7 = 2.01 to 3, and stratum #8 = above 3.01 (e.g., *extremely desirable*). Ten items were randomly sampled four times from within each social desirability-defined stratum, and the very large number of original items permitted sampling without replacement. Thus, participants completed only one of four completely different versions of the inventory (all 320 administered items were completely unique). All participants made typical personality ratings on a 5-point Likert-type agreement scale ranging from *strongly disagree* to *strongly agree* to 80 items. These items were presented in fully randomized orderings. In addition to the five substantive response categories, a *not applicable* option was also provided.

**Results**

As a check on possible differences in familiarity with the target, all once- and twice-removed raters were asked both how well they knew as well as how frequently they interacted with the target. Possible responses to the “how well do you know the target” question included: *hardly*, *moderately well*, *very well*, and *extremely well*. Response options for frequency of interaction were: *less than once per month*, *about once per month*, *about once per week*, and *daily*. There were slight differences on responses to these questions across groups (“well-known” (3) = 10.03, *p* = .018; “frequently” (3) = 16.91, *p* = .001), but the differences were not sufficiently meaningful to justify inclusion of these variables into our analyses (87% of twice-removed and 95% of once-removed indicated they knew the target either *very well* or *extremely well*, although 14% of twice-removed respondents indicated they only interacted with the target once a month or less frequently compared to 2% of once-removed respondents).

Missing analyses add here

Plus analysis of the question content?? “Quantitative Analysis of Textual Data” package “quanteda”: <https://www.r-bloggers.com/advancing-text-mining-with-r-and-quanteda/>

**Discussion**

**References**

Alicke, M. D., & Sedikides, C. (2011). Historical overview and conceptual framework. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of Self-Enhancement and Self-Protection.* (pp.1-19). New York: Guilford Press.

Asendorpf, J., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74,* 955-966. https://doi.org/10.1037/0022-3514.74.4.955

Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology, 97,* 613-636. https://doi.org/10.1037/a0026739

Blozis, S. A., Ge, X., Xu, S., Natsuaki, M. N., Shaw, D. S., Neiderhiser, J. Scaramella, L., Leve, L., & Reiss, D. (2013). Sensitivity analysis of multiple informant models when data are not missing at random. *Structural Equation Modeling, 20*(2), 283–298. https://doi.org/10.1080/10705511.2013.769393

Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self and peer reports. *European Journal of Personality, 3,* 31-45. https://doi.org/10.1002/per.2410030105

Brown, J. D., & Han, A. (2012). My better half: Partner enhancement as self-enhancement. *Social Psychological and Personality Science, 3*(4), 479-486. https://doi.org/10.1177/1948550611427607

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105. https://doi.org/10.1037/h0046016

Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior, 35,* 547-574. https://doi.org/10.1002/job.1909

Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait-multimethod approach. *Journal of Personality and Social Psychology, 102*, 408-426. https://doi.org/10.1037/a0025559

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and other observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment, 15,* 110-117. https://doi-org/10.1111/j.1468-2389.2007.00371.x

Connelly, B. S., & Ones, D. S. (2010). Perspective on personality: Meta-analytic integration of observers’ accuracy and predictive validity. *Psychological Bulletin, 136,* 1092–1122. https://doi.org/10.1037/a0021212

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business in Psychology, 25,* 325-334. https://doi.org/10.1007/s10869-010-9181-6

Cui, M., Durtschi, J. A., Donnellan, M. B., Lorenz, F. O., & Conger, R. D. (2010). Intergenerational transmission of relationship aggression: A prospective longitudinal study. *Journal of Family Psychology, 24,* 688–697. https://doi.org/10.1037/a0021675

Dilchert, S., Ones, D. S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science, 48,* 209-225.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology, 9,* 353-370. https://doi.org/[10.1037/0021-9010.94.2.353](https://doi-org.ezproxy.lib.uwstout.edu/10.1037/0021-9010.94.2.353)

Hofstee, W. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8,* 149-162. https://doi.org/10.1002/per.2410080302

Hogan, R., & Blickle, G. (2018). Socioanalytic theory: Basic concepts, supporting evidence and practical implications. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences: The science of personality and individual differences* (p. 110–129). Sage Reference. https://doi.org/10.4135/9781526451163.n5

John, O. P., & Robins, R. W., (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61,* 521-551. https://doi.org/10.1111/j.1467-6494.1993.tb00781.x

Kammeyer-Mueller, J., Steel, P. D. G., & Rubenstein, A. (2010). The other side of method bias: The perils of distinct source research designs. *Multivariate Behavioral Research, 45,* 294-321. https://doi.org/10.1080/00273171003680278

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9,* 300-311.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64,* 311-337. https://doi.org/10.1111/j.1467-6494.1996.tb00513.x

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology, 98,* 668-682. https://doi.org/10.1037/a0018771

Leising, D., Scherbaum, S., Locke, K. D., & Zimmermann, J. (2015). A model of ‘‘substance’’ and ‘‘evaluation’’ in person judgments. *Journal of Research in Personality, 57,* 61-71. https://doi.org/10.1016/j.jrp.2015.04.002

Lench, H. C., Quas, J. A., & Edelstein, R. S. (2006). My child is better than average: The extension and restriction of unrealistic optimism. *Journal of Applied Social Psychology, 36*(12), 2963-2979. https://doi.org/10.1111/j.0021-9029.2006.00137.x

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91,* 111-123. https://doi.org/10.1037/0022-3514.91.1.111

Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data. (3rd Ed). https://doi.org/10.1002/9781119482260

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79,* 272-280. https://doi.org/10.1037/a0038436

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology, 70,* 79-98.

Oh, I-S, Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96,* 762-773. https://doi.org/10.1037/a0021832

Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology, 7*(4, Pt.2)*,* 1-18. https://doi.org/10.1037/h0025230

Phares, V. (1992). Where’s Poppa? The relative lack of attention to the role of fathers in child and adolescent psychopathology. *American Psychologist. 1992; 47:*656–664. https://doi.org/10.1037//0003-066x.47.5.656

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63,* 539-569. https://doi.org/10.1146/annurev-psych-120710-100452

Podsakoff, P. M., MacKenzie, S. B., Lee, J-Y., & Podsadkoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903. https://doi.org/10.1037/0021-9010.88.5.879

R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of Personality Psychology* (pp. 649-679). San Diego, CA: Academic Press.

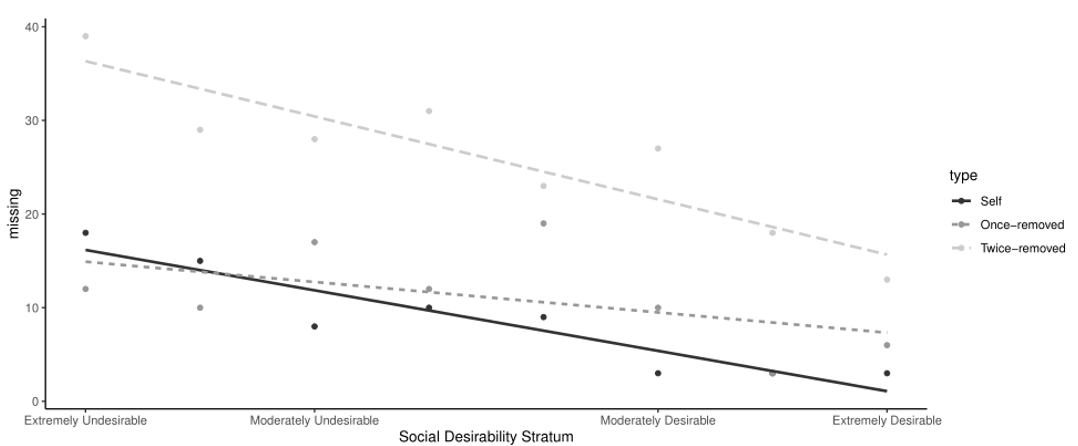
Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9,* 221-232. https://doi.org/10.1177/1094428105284955

Tajfel, H., & Turner, J. C. (1979). An integrative theory of inter-group conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Wessels, N. M., Zimmermann, J., Biesanz, J. C., & Leising, D. (2018). Differential associations of knowing and liking with accuracy and positivity bias in person perception. *Journal of Personality and Social Psychology.* https://doi.org/10.1037/pspp00002

**Figure 1**

Smoothed (Regression-predicted) Missing Value Estimates Across Social Desirability Strata

****

1. 1 There were two screening variables embedded within once- and twice-removed administrations, asking “who is supposed to be rated” prior to item administration and “who was rated” upon completion of the survey. Several individuals did not answer at least one of these questions with the correct source (the original self-rater) and these cases were excluded from analysis. [↑](#footnote-ref-1)