Personality and Individual Differences 196 (2022) 111708

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Distinguishing between trait desirability and item desirability in predicting item scores: Is informant evaluation of personality free from social desirability?

Tianxue Cui [a], Chester Chun Seng Kam [a,*], Emily Hongzhen Cheng [a], Man Yee Ho [b]

[a] The University of Macau, China
[b] City University of Hong Kong, China

ARTICLE INFO

Keywords:
Social desirability
Response distortion
Self-rating
Informant rating

ABSTRACT

It is often assumed that informant, or peer evaluation is more trustworthy than self-evaluation because the former is less vulnerable to social desirability response style. The current study examines this assumption in three independent samples, conceptually distinguishing between two types of social desirability: desirability due to *item* characteristics and due to a person's *trait* characteristics. We found that self-ratings (Studies 1–3) and peer ratings (Studies 2–3) are equally liable to item desirability in two cultures (Canada and China) with the relation further moderated by rater's trait desirability. Results challenge the popular assumption that informant ratings are impervious to social desirability. Relationship closeness has a moderating role, with closer targets rated more favorably than more distant ones. Results demonstrate the importance of conceptually distinguishing between item and trait desirability and are discussed in terms of a motivational account of response inflation.

1. Introduction

Personality assessment inventories have been developed for use in clinical, educational, and organizational settings. The use of inventories for predicting academic success (Poropat, 2009), task performance (Debusscher et al., 2016), and job performance (Alhendi, 2019b) has been well documented in both academic and applied settings. In their daily lives, people routinely evaluate themselves and others, formally and informally. Job interviewees, for example, judge their own ability and provide letters of recommendation from third parties.

Psychologists and members of organizations alike are concerned with people's tendency to present themselves over-positively in clinical assessments and interview, and on more formal psychometric instruments such as personality tests (Crowne & Marlowe, 1964; Paulhus, 2002). The tendency, commonly called social desirability, may motivate raters to behave in a culturally acceptable manner, regardless of their true attitudes or feelings (Crowne & Marlowe, 1964). At least six decades of research (e.g., Edwards, 1953; Vazire, 2010) has been devoted to understanding the nature of social desirability (Helmes & Holden, 2003; Holden, & Fekken, 1989; Paulhus, 1984; Vazire, 2010b), creating social desirability inventories (Crowne & Marlowe, 1964; Paulhus et al., 2003) and examining their validity (Dunlop et al., 2020; Kam et al., 2015), investigating the effect of social desirability in various settings (Protzko et al., 2019; Richman et al., 1999), and inventing novel methods to minimize social desirability (Bäckström et al., 2009; Bäckström & Björklund, 2021). Social desirability not only changes respondents' answers to an uncertain degree (Paulhus & Buckels, 2012), but it also obscures the genuine relationships between variables (Ganster et al., 1983; Ones et al., 1996; Paunonen & LeBel, 2012). Despite the tremendous amount of research effort on the topic, the socially desirable response process is still not fully understood.

1.1. Trait desirability and item desirability

A possible reason for the slow progress is that research (e.g., Edwards, 1953) does not always differentiate between two implicit aspects of desirability: person-level and item-level. As a personal, *trait* characteristic, social desirability was conceptualized as a tendency for an individual to behave in a culturally sanctioned manner across situations. Trait desirability was often measured with scales such as Marlowe-Crowne Social Desirability (Crowne & Marlowe, 1960) and the Balanced Inventory of Desirable Responding (Paulhus, 1991). Scale scores are found to be highly consistent over time, with rank-order correlations above 0.60 over six or more months (Haberecht et al.,

* Corresponding author at: The University of Macau, Taipa, Macau, China.
E-mail address: chesterkam@um.edu.mo (C.C.S. Kam).

https://doi.org/10.1016/j.paid.2022.111708
Received 20 February 2022; Received in revised form 28 April 2022; Accepted 30 April 2022
Available online 13 May 2022

2015; Schmitt & Steyer, 1993) and above 0.40 over 3 years (Lönnqvist et al., 2007). Given its longitudinal stability, person-level desirability is likely driven by an individual's chronic need for social approval or self-enhancement (i.e., the drive to perceive oneself favorably; Alicke & Sedikides, 2009b; Paulhus, 2002). Henceforth, we use the term "trait (social) desirability" to describe an individual's chronic tendency (i.e., trait) to present himself or herself unrealistically in an culturally sanctioned manner.

As a *momentary* characteristic, social desirability was conceptualized as a drive to distort one's image positively in a motivating context. Research showed that job applicants amplify their positive and downplay their negative qualities (Smith & Ellingson, 2002). Participants who were requested to "fake good" rated themselves much more favorably than those who did not receive this instruction or were asked to "fake bad," with effect size at least one standard deviation between conditions (Lambert et al., 2016).

However, rating distortion can also occur due to *item* characteristics. Edwards (1953) showed that the positive correlation between an item's average self-rating and its desirability reached as high as 0.87, implying that items' desirability dictates their probability of endorsement (see Thomas and Kilmann, 1975, for a similar finding). According to Podsakoff et al. (2003), "items may be written … to reflect more social desirable attitudes, behaviors, or perceptions" (p. 882). Some researchers (e.g., Kam & Meyer, 2015) have suggested that items have a given amount of inherent desirability, with some items (e.g., "I am sociable") inherently more desirable than others (e.g., "The first thing that I always do in a new place is to make friends"). Henceforth, we use the term "item (social) desirability" (i.e., "item evaluativeness") to describe an item's demand characteristic (Podsakoff et al., 2003).

Research has shown that trait and item desirability can *independently* predict item scores. Participants were asked to report trait desirability (using a social desirability scale) and item (subjective) desirability of scale items (Gove & Geerken, 1977; Randall & Fernandes, 1991). Each type of desirability was found to have unique predictive power in explaining self-ratings.

These studies, however, ignore a possible *interactive* process between the two types of desirability. Such an interaction is at least theoretically possible: The relationship between item desirability and item scores is likely positive, but more so for participants with high trait desirability due to their stronger motivation to distort item ratings (e.g., Bäckström et al., 2009). If the two desirability types interact, it is not theoretically acceptable to fuse them into a single construct ("social desirability"). The absence of distinction between the two types of desirability may inhibit proper theorizing of how they may interact to predict response distortion. The current research is motivated in part to investigate if and how the two types interact.

It is also motivated by a desire to examine the extent to which item desirability and informant (peer) ratings are associated. The general public likely believes that other-ratings are relatively immune to socially desirability; it follows that an evaluation by an expert referee (e.g., in a letter of recommendation) is taken as reasonably objective. On this assumption, theorists may predict that social desirability distorts informant ratings less than self-ratings (e.g., Funder, 1995). Again, this is based on the assumption that an informant's motivation to inflate ratings is weaker than the self's (e.g., Asendorpf & Ostendorf, 1998b; Colvin et al., 1995; Hayes & Dunning, 1997b; Kenny & West, 2010; John & Robins, 1993; McCredie & Kurtz, 2020).

As an example, McCredie and Kurtz (2020) stated:

> Relative to neutral traits, self-perception of evaluative traits is more likely to be distorted in an ego-protective manner than other-perception of those traits…leading to a self-other knowledge asymmetry in which informants are more accurate raters of these traits than the self. (p. 2).

This claim is consistent with some findings that individuals—at least in the West—rated themselves more favorably than their peers rated them (e.g., Heine & Renshaw, 2002; Lanthier, 2000; but see also Simms, et al., 2010; Stachowski & Kulas, 2020, for contrary evidence). Belief about the relative integrity peer ratings' still exists and thus deserves a thorough investigation.

1.2. The present research

The current research has three main purposes: (1) To examine the correlation between item desirability and self-reported personality ratings across several personality instruments (Studies 1–3); (2) to explore the possible interaction between trait and item desirability in predicting self-ratings, where trait desirability is operationalized as scores on three different scales (Studies 2–3); and (3) to test the assumption that peer ratings are less biased than self-ratings (Funder, 1995; McCredie & Kurtz, 2020). To enhance generalizability, both Western (Studies 1–2) and Eastern (Study 3) cultures were studied.

Based on previous research, we expected to find that item desirability correlates positively and strongly with personality self-ratings (Hypothesis 1, or H1; in Studies 1-3). We also hypothesized that item and trait desirability interact to predict self-ratings, with individuals highest in trait desirability most influenced by item desirability (H2; in Studies 2-3). Based on the assumption of weaker motivation for the informant to inflate ratings than the self, it is worth examining whether item desirability predicts and whether item desirability and trait desirability interact to predict peer ratings (Research Question 1, or RQ1; in Studies 2-3), and whether the results are consistent across Chinese and Canadian cultures (RQ2; in Studies 1-3).

If the two types of desirability interact to predict peer ratings, the question becomes why would peers be motivated to inflate the personality ratings of the individual. One possible moderator is relationship closeness: People may be more motivated to inflate ratings of targets to whom they are close (e.g., a romantic partner; Ones et al., 1996b) and deflate ratings of targets they dislike (Zimmermann et al., 2017b). Therefore, as an exploratory investigation, we checked the possible three-way interaction among item desirability, trait desirability, and relationship closeness (RQ3; in Study 3).

2. Study 1

Part of Study 1's data were from Kam (2013).

2.1. Method

2.1.1. Participants

Participants were 1254 Canadian introductory psychology university students (873 females, 380 males, 1 unidentified; $M_{age} = 18.38$, $SD_{age} = 2.27$) who completed a mass testing survey for course credit (1 = *strongly disagree*; 5 = *strongly agree*).

2.1.2. Measures

2.1.2.1. Big Five Personality. The NEO domain of the International Personality Item Pool (IPIP) (Cronbach's αs =0.75 to 0.87) was used to examine the Big Five personality traits (Goldberg et al., 2006). Each factor consists of 10 items, half of which were reverse-keyed.

2.1.2.2. BIDR. We used the BIDR (Paulhus, 1991) to measure two components of social desirability: impression management (IM; Cronbach's α = 0.76) and self-deception (SDE; Cronbach's α =0.66). The overall desirability score is reported for our purpose. Two items (one about sexy books or magazines and another about doubting oneself as a competent lover) were excluded, as the Ethics Board found them invasive. BIDR and IPIP items were randomized together.

2.1.2.3. Self-esteem. The Rosenberg (1965) self-esteem scale contains

10 items (Cronbach's α = 0.90).

2.1.2.4. Item desirability. A separate group of participants from the same source (5 women, 3 men; M_{age} = 18.50, SD_{age} = 0.53), following the same instructions as in Study 1, rated item desirability of all 60 personality items (M = 4.16; SD = 1.91, range = [1.13, 6.63]). The inter-judge reliability among items was high (mean r_{wg} = 0.98, range r_{wg} = 0.94–1.00).

2.2. Data analysis

To investigate H1, we correlated item desirability with item scores. To test H2, cross-classified random effects models with maximum likelihood estimator were employed with SPSS 26.0. In the main effect model, item desirability and trait desirability (both grand-mean centered) were entered into the model to predict item scores:

$$Y_{ij} = \gamma_{00} + \gamma_{10}\ \text{item desirability}_i + \gamma_{01}\ \text{trait desirability}_j + u_{0i} + u_{0j} + \varepsilon_{ij}$$

where Y_{ij} represents the score of item i within the cross-classification of participant j; γ_{00} is the intercept of participants' item score; γ_{10} is the fixed slope of the desirability of item i; γ_{01} is a fixed slope of participant j's trait desirability; u_{0i} represents the residual random intercept for item i; u_{0j} is the residual random intercept for participant j; and ε_{ij} represents the residual error.

In the second step, the interaction model adds the interaction term between item desirability and trait desirability that has the fixed slope γ_{11}:

$$Y_{ij} = \gamma_{00} + \gamma_{10}\ \text{item desirability}_i + \gamma_{01}\ \text{trait desirability}_j + \gamma_{11}\ \text{item desirability}_i\text{*trait desirability}_j + u_{0i} + u_{0j} + \varepsilon_{ij}$$

2.3. Results

A significant positive correlation was found between item mean scores and item desirability (r = 0.72, 95% CI [0.58,0.82], $p < .001$; see Fig. 1a). H1 was supported.

2.3.1. Predictors of self-ratings

The main effects of trait desirability, $\gamma_{01} = -0.08$, $t(1246.39) = -6.74$, $p < .001$, and item desirability, $\gamma_{10} = 0.25$, $t(58) = 7.88$, $p < .001$, were statistically significant (see Table 1). The main effect was nonetheless qualified by the significant interaction between item and trait desirability, $\gamma_{11} = 0.32$, $t(73315.21) = 61.21$, $p < .001$. Simple effect analysis showed that item desirability always positively predicted self-rated personality item scores, but the magnitude varied with level of trait desirability: the simple slope for participants with high trait desirability, $\gamma_{10} = 0.37$, $Z = 11.67$, $p < .001$, was steeper than it was for those with low trait desirability, $\gamma_{10} = 0.13$, Z = 4.06, $p < .001$; see Fig. 2a). H2 was supported.

3. Study 2

The data for Study 2 are from Paunonen (1982). The data include peer ratings in addition to self-ratings.

3.1. Method

3.1.1. Participants

Participants were 45 pairs of roommates at a large Canadian university (35 pairs of women). The mean age of all the students was 19.18 (SD_{age} = 0.84). Participants rated themselves and their roommates on a paper-and-pencil questionnaire on separate days.

3.1.2. Measures

3.1.2.1. PRF-E and social desirability. The PRF-E (Jackson, 1974) measures 21 behavior-situation types of personality traits (e.g., affiliation, dominance, understanding), in addition to a self-rated (trait) desirability scale. Sixteen items were included for each dimension (including the social desirability dimension); nearly half were reversed-keyed. Participants were asked to respond to the 352 items on a 9-point Likert scale (1 = *extremely uncharacteristic of the person being rated*; 9 = *extremely characteristic of the person being rated*). The self-rated version (Cronbach's αs ranged from 0.69 to 0.91) and peer-rated version (Cronbach's αs ranged from 0.76 to 0.95) had parallel items (except of course the rating target).

3.1.2.2. Item desirability. Helmes et al. (1977) reported item desirability for the PRF-E (on a 9-point Likert scale) using a large sample of college students (114 women and 100 men). Although unreported by Helmes et al., inter-rater reliability should be high given the already high consensus with a much smaller sample of judges in our Studies 1.

3.2. Results

3.2.1. Self-ratings

The findings of the Study 1 were replicated: item desirability was positively correlated with self-rated item scores (H1), r = 0.76, 95% CI [0.70, 0.80], $p < .001$ (Fig. 1b). In the multilevel anlysis (Table 1), self-rated item mean score was predicted by item desirability, $\gamma_{10} = 0.76$, $t(334) = 21.07$, $p < .001$, but not by trait desirability, $\gamma_{01} = 0.02$, $t(88) = 0.93$, $p = .357$. The main effect was again qualified by the interaction between the types of desirability, $\gamma_{11} = 0.23$, t (29814) = 23.92, $p < .001$. The slope of item desirability was steeper for participants with high trait desirability, $\gamma_{10} = 1.00$, $Z = 26.66$, $p < .001$, than for those with low trait desirability (H2), $\gamma_{10} = 0.52$, $Z = 13.99$, $p < .001$ (Fig. 2b).

3.2.2. Peer-ratings

Contrary to expectation, there was also a strong positive correlation between item desirability and peer-rated item scores (RQ1), r = 0.68, 95% CI [0.62, 0.74], $p < .001$ (Fig. 1c). Comparison of correlations using path analysis in the program Amos 24.0 revealed that the correlation associated with self-rated items was not significantly stronger than the one associated with peer-rated items, rs = 0.76 vs 0.68, $\Delta\chi^2$ (1) = 2.71, p = .100, meaning that peer ratings were not less affected by item desirability.

Contrary to expectation, the interaction term between item and trait desirability (see Table 1) also significantly predicted peer-rated item means (RQ2), $\gamma_{11} = 0.18$, t (29814) = 18.36, $p < .001$, again with the slope for item desirability steeper for individuals with high trait desirability, $\gamma_{10} = 0.80$, $Z = 21.21$, $p < .001$, than for those with low trait desirability, $\gamma_{10} = 0.43$, $Z = 11.49$, $p < .001$ (Fig. 2c). The results for peer rating thus mirrored the results for self-rating.

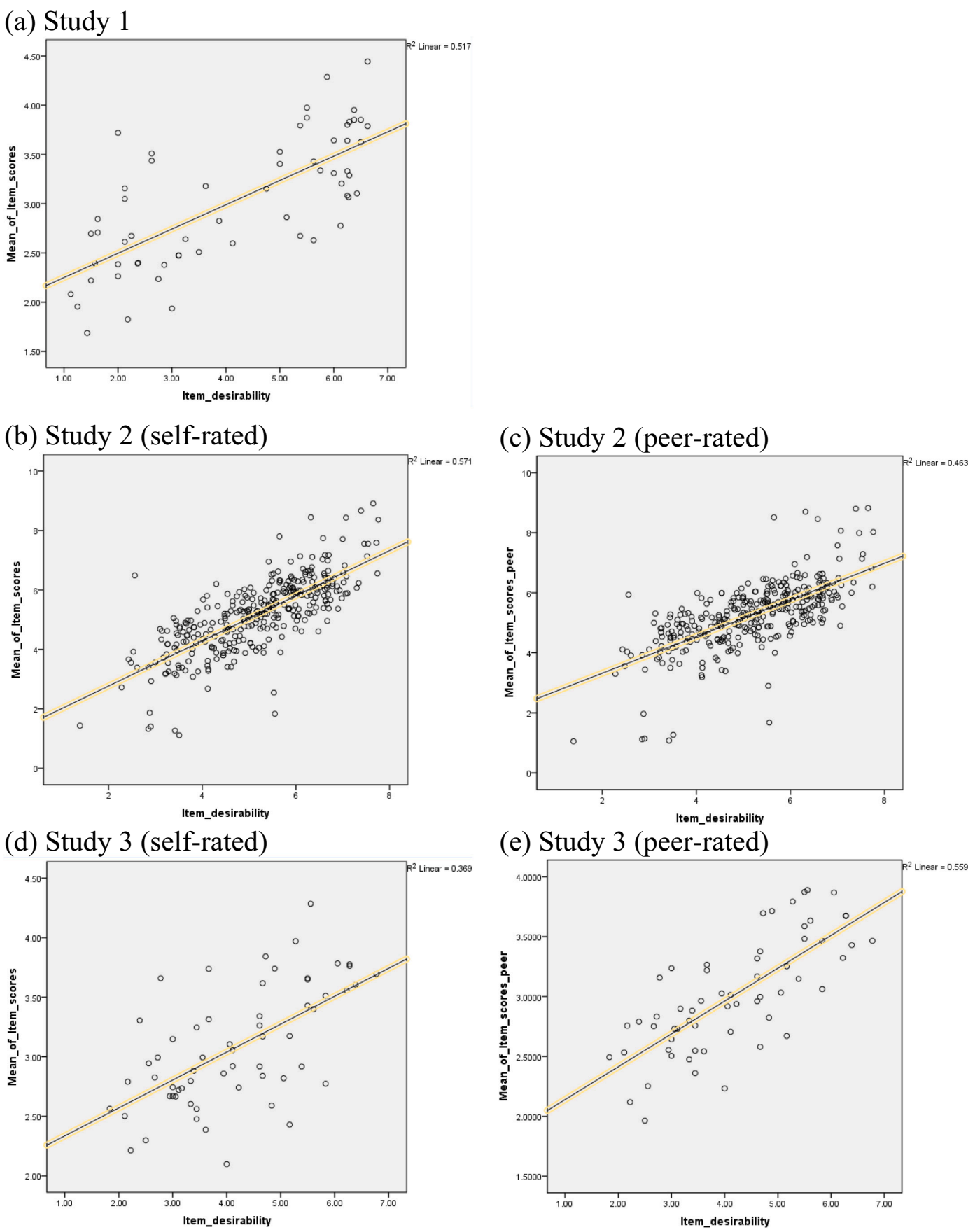


Fig. 1. Relationship between item desirability and self/peer-rated mean of item scores.
Note. The horizontal axis represents item-level desirability and the vertical axis represents the self/peer-rated mean of item scores in each study.

Table 1
Multilevel modeling using item desirability (ID), trait desirability (TD), and relationship closeness to predict item scores.

	Study 1	Study 2		Study 3	
	Self-rated scores	Self-rated scores	Peer-rated scores	Self-rated scores	Peer-rated scores
	γ Estimate (SE)	γ Estimate (SE)	γ Estimate (SE)	γ Estimate (SE)	γ Estimate (SE)
Main effect					
Version (Study 3 only)[a]				−0.02 (0.01)	0.005 (0.01)
ID	**0.25***** (0.03)	**0.76***** (0.04)	**0.61***** (0.04)	**0.20***** (0.03)	**0.23***** (0.03)
TD	−0.08 (0.01)	0.02 (0.03)	0.01 (0.03)	0.001 (0.05)	0.06 (0.04)
Closeness (Study 3 only)					0.02 (0.02)
Two-way interaction					
ID×TD	**0.32***** (0.01)	**0.23***** (0.01)	**0.18***** (0.01)	**0.46***** (0.02)	**0.21***** (0.02)
ID×Closeness					**0.12***** (0.01)
TD × Closeness					0.01 (0.05)
Three-way interaction					
ID×TD × Closeness					**0.07***** (0.02)

[a] Self-rating first = 1, peer rating first = −1. ID = Item desirability. TD = Trait desirability. Numbers within parenthesis represent standard error (SE). Statistically significant coefficients were bolded for clarity.
*** $p < .001$.

4. Study 3

The data for this study are from Kam and Sun (2020). The study includes a measure of participants' closeness with the rated target. Theoretically, the relationship between item ratings and item desirability (discovered in Study 2) should be moderated by relationship closeness, because individuals should judge their friends more positively than they do a stranger (Hayes & Dunning, 1997; Leising, Erbs, & Fritz, 2010).

4.1. Method

4.1.1. Participants

Participants were 218 pairs of roommates from various universities in Shanghai, China ($N = 436$; 289 women, 146 men, 1 unidentified) who completed an online survey (with items on a 5-point Likert scale: 1 = *strongly disagree*; 5 = *strongly agree*) for $100 RMB (approximately US $16). The mean age of all students is 21.02 ($SD_{age} = 3.93$).

4.1.2. Measures

4.1.2.1. HEXACO (self-report and peer-report). The official Chinese version of the HEXACO PI-R personality inventory (Ashton & Lee, 2009b), posted on the HEXACO website, measures six factors of personality. The self-report (Cronbach's αs = 0.69–0.74) and the peer-report versions (Cronbach's αs = 0.74–0.86) are parallel in wording, each with 60 items. The students completed both versions, with order of versions counterbalanced.

4.1.2.2. MCSD. Trait desirability was measured by the 33-item MCSD (Crowne & Marlowe, 1960; Cronbach's $\alpha = 0.62$). The MCSD was embedded and randomized within the self-report HEXACO items.

4.1.2.3. Relationship closeness. At the end of the survey, participants reported their subjective affective closeness with their roommate, using the Friendship Qualities Scale (4 items; Bukowski et al., 1994; Cronbach's $\alpha = 0.81$).

4.1.2.4. Item desirability. A separate group of Chinese university students ($N = 18$, 10 women; $M_{age} = 24.17$, $SD_{age} = 2.66$) rated the item desirability of the HEXACO PI-R (self-report). Self- and peer-report versions were strictly parallel in wording; therefore, item desirability was identical. For consistency, the same procedure was used as in Studies 1 and 2. The inter-judge reliability for the 60 items was still high (mean $r_{wg} = 0.97$, range $r_{wg} = 0.94$–1.00), far above the recommended cutoff value of 0.70. The average item desirability rating was 4.17 ($SD = 1.29$) with an extensive range ([1.83, 6.78]).

4.2. Results

4.2.1. Self-ratings

We replicated the previous studies: item desirability correlated significantly with self-rated item means (H1), $r = 0.61$, 95% CI [0.42, 0.75], $p < .001$ (see Fig. 1d). Replicating the results of Study 2, item desirability, $\gamma_{10} = 0.20$, $t(58) = 5.82$, $p < .001$, but not trait desirability, $\gamma_{01} = 0.03$, t (424.89) = 0.03, $p = .02$, predicted self-ratings (see Table 1). The interaction between item and trait desirability was significant (H2), $\gamma_{11} = 0.46$, t (25,095.09) = 26.43, $p < .001$, again replicating Study 2: the relationship between item desirability and item scores varied with trait desirability, with the slope for participants with high trait desirability, $\gamma_{10} = 0.32$, $Z = 8.66$, $p < .001$, stronger than the slope for those with low trait desirability, $\gamma_{10} = 0.08$, $Z = 2.24$, $p = .025$ (see Fig. 2d).

4.2.2. Peer-ratings

We replicated our finding in Study 2 that item desirability correlated significantly with peer-rated item means (RQ1), $r = 0.75$, 95% CI [0.64, 0.83], $p < .001$ (Fig. 1e). Using the same procedure as in Study 2 the correlation comparison showed that the correlation associated with self-rated items was not statistically different from that associated with peer-rated items, $rs = 0.61$ vs. 0.75, $\Delta\chi^2$ (1) =0.22, $p = .636$.

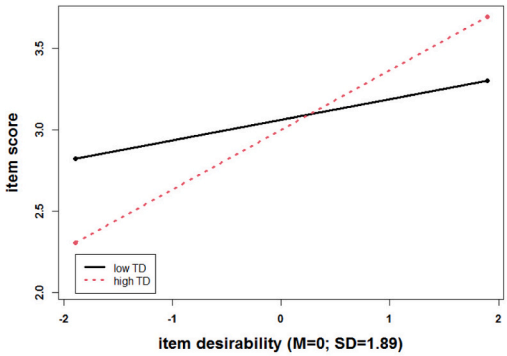
We then examined item desirability, trait desirability, and closeness, together with their two-way interaction terms, in predicting peer-rating item scores (see Table 1). Item desirability significantly predicted peer-rated item means, $\gamma_{10} = 0.23$, $t(58.00) = 8.53$, $p < .001$, but participants' own trait desirability, $\gamma_{01} = 0.06$, t (422.67) = 1.40, $p = .163$, and closeness ($\gamma_{02} = 0.03$, t (423.26) = 1.57 $p = .117$), did not. These main results were qualified by two significant interactions. First, the interaction between item desirability and trait desirability (RQ2), $\gamma_{11} = 0.21$, t (25,088.03) = 12.22, $p < .001$, replicated the finding in Study 2. The interaction showed the same pattern as that in the self-rating: the relationship between item desirability and trait desirability was stronger for individuals high in trait desirability, $\gamma_{10} = 0.29$, $Z = 10.39$, $p < .001$, than for individuals low in trait desirability, $\gamma_{10} = 0.18$, $Z = 6.49$, $p < .001$ (Fig. 2e). We therefore reasoned that—at least to some extent—when they make personality judgments, participants treat their peers as they do themselves.

Second, the interaction between closeness and trait desirability, $\gamma_{12} = 0.12$, t (25,088.30) = 20.04, $p < .001$, was unexpected. The relationship between item desirability and closeness was stronger for individuals who are close to the roommates, $\gamma_{02} = 0.32$, $Z = 11.64$, $p < .001$, than for individuals who are more distant from their roommates, $\gamma_{02} = 0.14$, $Z = 5.13$, $p < .001$ (Fig. 2f).

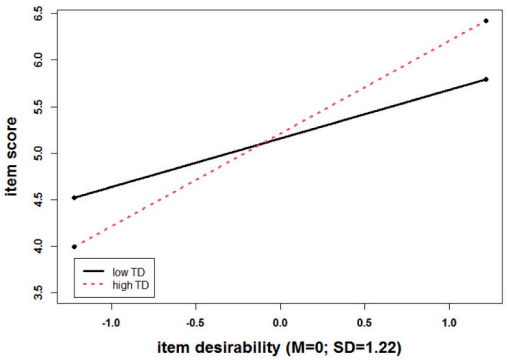
4.2.3. Exploratory analysis

To better understand the significant interaction between item and

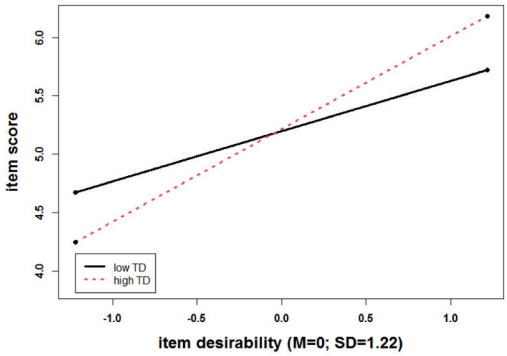
(a) Moderating effect of TD in Study 1



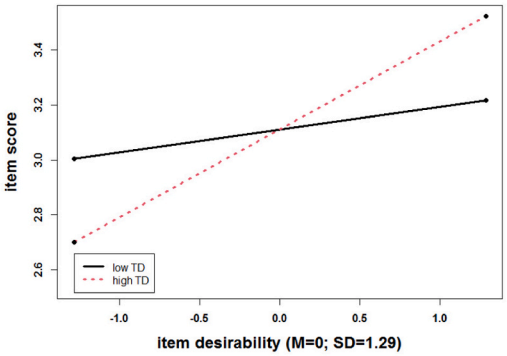
(b) Moderating effect of TD in Study 2 (self-ratings)



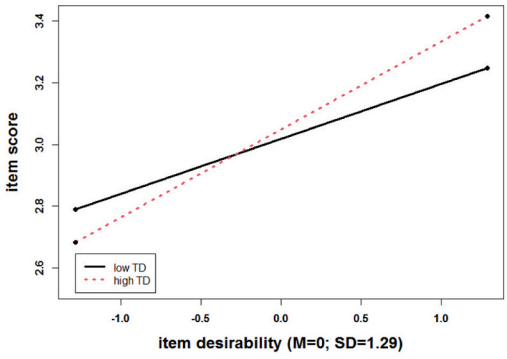
(c) Moderating effect of TD in Study 2 (peer ratings)



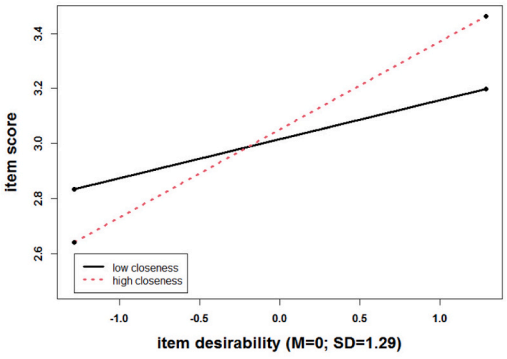
(d) Moderating effect of TD in Study 3 (self-ratings)



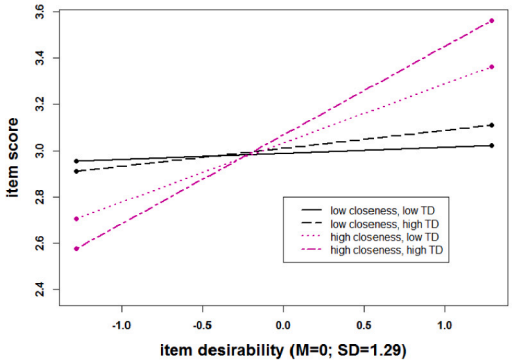
(e) Moderating effect of TD in Study 3 (peer ratings)



(f) Moderating effect of closeness in Study 3 (peer ratings)



(g) Moderating effect of TD and closeness in Study 3 (peer ratings)



(caption on next page)

Fig. 2. Simple slopes of item desirability in predicting self/peer-rated mean of item scores with different levels of moderators (trait desirability in Studies 1–3, closeness in Study 3).
Note. TD = Trait desirability. Item desirability has been grand-mean centered. High (low) trait desirability has been defined as one standard deviation above (below) the mean.
(For interpretation of the references to color in this figure legend, the reader is referred to the online version of this chapter.).

trait desirability, we conducted an exploratory analysis, probing a three-way interaction involving item desirability, trait desirability, and closeness (RQ3). This interaction was also statistically significant (see Table 1), $\gamma_{14} = 0.07$, t (25,087.04) = 3.52, $p < .001$. We followed up by probing the relationship between item desirability and peer's trait desirability for two levels of closeness: one SD above the mean (4.35) and two SDs below the mean (2.25). These two values represent a high and low levels of closeness, respectively, given that the midpoint of the closeness scale is 3 and the mean of closenss is fairly high in the sample (3.65).

At the high level of closeness, participants with high trait desirability had a stronger positive relationship between item desirability and ratings ($\gamma_{10} = 0.38$, $Z = 13.76$, $p < .001$) than did those with low trait desirability, $\gamma_{10} = 0.26$, $Z = 9.12$, $p < .001$. At the low level of closeness, the interaction effect was attenuated. Participants with high trait desirability still had a stronger positive relationship between item desirability and ratings ($\gamma_{10} = 0.08$, $Z = 2.53$, $p = .011$) than did those with low trait desirability, $\gamma_{10} = 0.03$, $Z = 0.90$, $p = .369$, but the difference between the unstandardized slopes was smaller. Thus, at high trait desirability, the relationship between item desirability and item ratings becomes stronger with increasing closeness (high vs low closeness: $\gamma_{10} = 0.38$ vs 0.08; see Fig. 2g).

5. Discussion

Previous research on social desirability did not always distinguish between trait desirability and item desirability. Accordingly, the potential of these aspects to interact was not taken into account. Lack of attention to possible interactions has led to confusion in the construct's theoretical development and roadblocks in empirical research. Another concern is that researchers tended to assume that social desirability affects self-ratings more than peer ratings.

The current study offers correctives to some of these notions. We demonstrated the benefits of differentiating between trait and item desirability, and showed that peer ratings are far from immune to social desirability. Across culturally distinct samples (Canada and China), we found that item desirability correlates substantially with peer ratings (e.g., $rs > 0.65$) and that trait desirability moderates the relationship between item desirability and item scores. These results highlight the interactivity between item and trait desirability in both self- and peer ratings. In their ongoing research and theoretical work, investigators of social desirability are encouraged to differentiate between the two types.

5.1. Peer ratings: effects of trait and item desirability

Investigating how item desirability affects peer ratings, we found a three-way interaction involving item desirability, trait desirability, and relationship closeness: Participants strong in trait desirability and emotionally close to their roommates appeared to inflate personality ratings on desirable items. The moderating effect of trait desirability may point to a motivational source as opposed to a cognitive, non-motivational oversight: Participants may consciously or unconsciously inflate ratings of a close roommate—probably similar to how they would rate themselves. Researchers should thus beware of the previously untested assumption that peer ratings are immune to social desirability.

A difficulty with peer ratings is that, except in rare circumstances, researchers do not have the luxury of randomly assigning raters. Instead, participants nominate their own raters, and it is likely to be someone they like or feel close to. In the current study, roommates would probably not participate if their relationship was negative. Given relatively strong relationship closeness ($M = 3.66$ out of 5 in the current study), peer ratings may be just as distorted as self-ratings. Consistent with the findings by Zimmermann et al. (2017), dislike was consistently associated with lower positivity bias across self-reports and informant reports of personality measures. Whereas some academic researchers regard peer ratings as the more accurate (McCredie & Kurtz, 2020), the present finding invites them to reconsider the assumption.

There is ongoing debate as to which type of rating, self- or other-, has higher validity (e.g., Connelly & Ones, 2010; Paunonen & Kam, 2014; Vazire, 2010). On the one hand, self-raters have more opportunities to observe their behavior, resulting in more accurate judgments. On the other, in some occasions social desirability may distort self-ratings much more than informant ratings, resulting in less accurate judgments. Our findings do not speak to which rating type is more valid, but they do show that *both* self- and informant ratings can be distorted by social desirability. The relationship between social desirability and informant ratings is further complicated by its interaction (discussed earlier) with relationship closeness.

Some questions remain unanswered. If informant ratings are biased by social desirability, why are they better predictors of performance than self-ratings in a meta-analysis? (Connelly & Ones, 2010). Is the relationship between a rating and a criterion distorted by social desirability, and if so, how? (Ones et al., 1996). Theoretically, disentangling the process of how social desirability explains response distortion man help resolve the debate of who is the better judge of personality: self or observer (Vazire, 2010). We believe separation of the two types of desirability, together of course with routine measurement of both, will be key for explaining if and how desirability distorts response (Leising et al., 2015).

The present findings shed light on response bias of self-report and others' report in personality assessment inventories. Although using both self- and informant reports provides more comprehensive information about personality traits and motivation, it is not immune to bias. The implication is that plausible ways need to be found to reduce socially desirable responding. For example, at the item level, framing items neutrally might be effective (Bäckström et al., 2009); at the person level, warning against faking (e.g., Kluger & Colella, 1993) may partially reduce their socially desirability motivation. Relationship closeness plays a role in influencing social desirability. Further studies can consider including both close and not-so-close others as informants and use scenario-based assessment to understand and measure individual personality differences that characterize people across time. In a selection context, interviewers should beware of recommendation letters or personality reports from a referee close to the job candidate. These reports can be as biased as the candidate's self-description.

5.2. Potentially shedding new light on previous findings

In previous research, the distinction between trait and item desirability was regularly overlooked. An example from our own research illustrates the importance of the distinction. Some research supported the argument that social desirability response style is a major cause of score differences between regular- and reverse-keyed items (e.g., Kulas et al., 2018), while other research failed to find support for this hypothesis (e.g., Kam & Meyer, 2015). This discrepancy in the empirical findings was possibly caused by the differential operationalization of social desirability.

Using structural equation modeling, Kam and Meyer (2015) extracted unique variances (i.e., method latent factors) related to regular- and reverse-keyed items from various personality and social psychology measures. They then correlated those unique variances with social desirability scales (i.e., trait desirability). Finding the correlation in a low range (*rs* from −0.08 to 0.20), the researchers concluded that social desirability was unlikely to cause the method factors. The finding was replicated in another study (e.g., Kam, 2018). These empirical results, however, were contradicted by later studies that directly measured item desirability. Kulas et al. (2018), for instance, discovered a strong correlation between item desirability and loadings of the method factors, implying that participant response was heavily influenced by social desirability. Unlike Kam and colleagues (Kam & Meyer, 2015; Kam, 2018), Kulas et al. (2018) concluded that social desirability severely confounded item responses in self-report surveys.

If previous research had recognized two types of social desirability, they might have found that method factors could be predicted by not only trait desirability but also by its interaction with item desirability. The contrary findings could then be reconciled. In light of the possible interactive effect between the two types, those researchers could now realize that social desirability can explain more item variance than previously assumed. In short, social desirability researchers have a more complete understanding of their construct.

5.3. Limitations and future directions

Several limitations should be noted. First, the generalizability of the findings in this study deserves further examination. Although we conducted the study in both Western and Eastern countries, in both places the respondents were highly educated university students in modern cities, and data from the Canadian sample was collected much earlier than the Chinese sample. Although there is little theoretical reason to expect dramatically different findings, replication in less developed nations with other participant characteristics (such as job interviewees) and a recent sample will enhance the generalizability of the findings. Second, although not critical to the purpose of the current study, the exploratory finding of the three-way interaction involving item desirability, trait desirability, and closeness requires replication with a different sample. Third, although replication across multiple measures (NEO-IPIP, PRF-E, HEXACO) is undoubtedly a strength of the study, due to the low number of items in each personality dimension we were unable to examine personality variables as potential moderators. It is possible that the relationship between desirability and item ratings is stronger for some personality dimension due to differential variation of item desirability across traits.

5.4. Conclusion

The current study contributes to the literature on social desirability by demonstrating that peer ratings can be just as vulnerable to social desirability bias as self-ratings. Conceptually, social desirability comprises trait and item desirability, which appear to interact and influence both self-ratings and peer ratings. Making a distinction between two types of social desirability appears to be a promising strategy for future investigation.

Funding

The current research was financially supported by the Multi-Year Research Grant from the University of Macau to the second author.

References

Alhendi, O. (2019). Personality traits and their validity in predicting job performance at recruitment: A review. *International Journal of Engineering and Management Sciences, 4*(3), 222–231.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*(1), 1–48. https://doi.org/10.1080/10463280802613866

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality & Social Psychology, 74*(4), 955–966. https://doi.org/10.1037/0022-3514.74.4.955

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340–345. https://doi.org/10.1080/00223890902935878

Bäckström, M., & Björklund, F. (2021). Socially desirable responding in experience sampling: Consequences for personality research. *Journal of Individual Differences, 42*(4), 157–166. https://doi.org/10.1027/1614-0001/a000342

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43*(3), 335–344. https://doi.org/10.1016/j.jrp.2008.12.013

Bukowski, W. M., Hoza, B., & Boivin, M. (1994). Measuring friendship quality during pre- and early adolescence: The development and psychometric properties of the friendship qualities scale. *Journal of Social and Personal Relationships, 11*(3), 471–484. https://doi.org/10.1177/0265407594113011

Colvin, C. R., & Funder, D. C. (1995). Overly positive self-evaluations and personality: negative implications for mental health. *Journal of Personality & Social Psychology, 68*, 1151–1162.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092–1122. https://doi.org/10.1037/a0021212

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. https://doi.org/10.1037/h0047358

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.

Debusscher, J., Hofmans, J., & Fruyt, F. D. (2016). Do personality states predict momentary task performance? The moderating role of personality variability. *Journal of Occupational and Organizational Psychology, 89*, 330–351. https://doi.org/10.1111/joop.12126

Dunlop, P. D., Bourdage, J. S., de Vries, R. E., McNeill, I. M., Jorritsma, K., Orchard, M., & Choe, W. K. (2020). Liar! Liar!(when stakes are higher): Understanding how the overclaiming technique can be used to measure faking in personnel selection. *Journal of Applied Psychology, 105*(8), 784.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*(2), 90–93. https://doi.org/10.1037/h0058073

Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self- judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego, CA: Academic Press.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656–672.

Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal, 26*, 321–331.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public domain personality measures. *Journal of Research in Personality, 40*, 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Gove, W. R., & Geerken, M. R. (1977). Response bias in surveys of mental health: An empirical investigation. *American Journal of Sociology, 82*(6), 1289–1317. https://doi.org/10.1086/226466

Haberecht, K., Schnuerer, I., Gaertner, B., John, U., & Freyer-Adam, J. (2015). The stability of social desirability: A latent change analysis. *Journal of Personality, 83*(4), 404–412. https://doi.org/10.1111/jopy.12112

Hayes, A. F., & Dunning, D. (1997). Construal processes and trait ambiguity implications for self–peer agreement in personality judgment. *Journal of Personality and Social Psychology, 72*, 664–677.

Heine, S. J., & Renshaw, K. (2002). Interjudge agreement, self-enhancement, and liking: Cross-cultural divergences. *Personality & Social Psychology Bulletin, 28*(5), 578–587. https://doi.org/10.1177/0146167202288002

Helmes, E., & Holden, R. R. (2003). The construct of social desirability: One or two dimensions? *Personality and Individual Differences, 34*(6), 1015–1023. https://doi.org/10.1016/S0191-8869(02)00086-7

Helmes, E., Reed, P. L., & Jackson, D. N. (1977). Desirability and frequency scale values and endorsement proportions for items of personality research form-e. *Psychological Reports, 41*(2), 435–444. https://doi.org/10.2466/pr0.1977.41.2.435

Holden, R. R., & Fekken, G. C. (1989). Three common social desirability scales: Friends, acquaintances, or strangers? *Journal of Research in Personality, 23*(2), 180–191. https://doi.org/10.1016/0092-6566(89)90022-6

Jackson, D. N. (1974). *Personality research form manual*. Port Huron, MI: Research Psychology Press.

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

Kam, C. (2013). Probing item social desirability by correlating personality items with balanced inventory of desirable responding (BIDR): A validity examination. *Personality & Individual Differences, 54*(4), 513–518. https://doi.org/10.1016/j.paid.2012.10.017

Kam, C. C. S. (2018). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research, 47*(3), 574–597. https://doi.org/10.1177/0049124115626177

Kam, C. C. S., & Meyer, J. P. (2015). Implications of item keying and item valence for the investigation of construct dimensionality. *Multivariate Behavioral Research, 50*(4), 457–469. https://doi.org/10.1080/00273171.2015.1022640

Kam, C. C. S., Risavy, S. D., & Perunovic, W. Q. E. (2015). Using over-claiming technique to probe social desirability ratings of personality items: A validity examination. *Personality & Individual Differences, 74*, 177–181. https://doi.org/10.1016/j.paid.2014.10.017

Kam, C. C. S., & Sun, S. (2020). Method factor due to the use of reverse-keyed items: Is it simply a response style artifact? *Current Psychology*, (4), 1–9. https://doi.org/10.1007/s12144-020-00645-z

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self and other perception: A meta-analysis. *Personality and Social Psychology Review, 12*, 196–213. https://doi.org/10.1177/1088868309353414

Kluger, A. N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology, 46*(4), 763–780. https://doi.org/10.1111/j.1744-6570.1993.tb01568.x

Kulas, J. T., Klahr, R., & Knights, L. (2018). Confound It! Social desirability and the "reverse-scoring" method effect. *European Journal of Psychological Assessment, 35*(6), 855–867. https://doi.org/10.1027/1015-5759/a000459

Lambert, C. E., Arbuckle, S. A., & Holden, R. R. (2016). The marlowe-crowne social desirability scale outperforms the BIDR impression management scale for identifying fakers. *Journal of Research in Personality, 61*, 80–86. https://doi.org/10.1016/j.jrp.2016.02.004

Lanthier, R. P. (2000). Intersibling agreement for Goldberg's big five adjective markers. *Psychological Reports, 86*, 400–406. https://doi.org/10.2466/pr0.2000.86.2.400

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

Lönnqvist, J.-E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality, 75*, 291–322. https://doi.org/10.1111/j.1467-6494.2006.00440.x

McCredie, M. N., & Kurtz, J. E. (2020). Prospective prediction of academic performance in college using self- and informant-rated personality traits. *Journal of Research in Personality, 85*, Article 103911. https://doi.org/10.1016/j.jrp.2019.103911

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*(1), 79–98.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). The role of social desirability in personality testing for personnel selection. *Journal of Applied Psychology, 81*(6), 660–679. https://doi.org/10.1037/0021-9010.81.6.660

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598–609. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307–317. https://doi.org/10.1037/0022-3514.60.2.307

Paulhus, D. L. (2002). Social desirable responding: The evolution of a construct. In H. I. Brown, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Paulhus, D. L., & Buckels, E. (2012). Classic self-deception revisited. In S. Strachey, & T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 363–378). New York, NY: The Guilford Press.

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of personality and social psychology, 84*(4), 890–904. https://doi.org/10.1037/0022-3514.84.4.890

Paunonen, S. V. (1982). *Behavioral consistency and individual differences in predictive structure*. Doctoral dissertation. Retrieved from ProQuest Dissertations Publishing (Accession No. NK54116).

Paunonen, S. V., & Kam, C. (2014). The accuracy of roommate ratings of behaviors versus beliefs. *Journal of Research in Personality, 52*, 55–67. https://doi.org/10.1016/j.jrp.2014.07.006

Paunonen, S. V., & LeBel, E. P. (2012). Socially desirable responding and its elusive effects on the validity of personality assessments. *Journal of Personality and Social Psychology, 103*(1), 158–175. https://doi.org/10.1037/a0028165

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Poropat, W. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychology Bulletin, 135*(2), 322–338. https://doi.org/10.1037/a0014996

Protzko, J., Zedelius, C. M., & Schooler, J. W. (2019). Rushing to appear virtuous: Time pressure increases socially desirable responding. *Psychological Science, 30*, 1584–1591.

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10*(11), 805–817. https://doi.org/10.1007/BF00383696

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*(5), 754–775. https://doi.org/10.1037/0021-9010.84.5.754

Rosenberg, M. (1965). *Society and the adolescent child*. Princeton, NJ: Princeton University Press.

Schmitt, M. J., & Steyer, R. (1993). A latent state-trait model (not only) for social desirability. *Personality and Individual Differences, 14*, 519–529. https://doi.org/10.1016/0191-8869(93)90144-R

Simms, L. J., Zelazny, K., Yam, W. H., & Gros, D. F. (2010). Self-informant agreement for personality and evaluative person descriptors: Comparing methods for creating informant measures. *European Journal of Personality, 24*(3), 207–221. https://doi.org/10.1002/per.763

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*, 211–219. https://doi.org/10.1037/0021-9010.87.2.211

Stachowski, A. A., & Kulas, J. T. (2020). The persnickety pervasiveness of rating enhancement in personality assessment: The self, the other, and then again yet another. *European Journal of Psychological Assessment, 3*, 1–13. https://doi.org/10.1027/1015-5759/a000610

Thomas, K. W., & Kilmann, R. H. (1975). The social desirability variable in organizational research: An alternative explanation for reported findings. *Academy of Management Journal, 18*, 741–752.

Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*(3), 243–262. https://doi.org/10.1177/1745691610369465

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (soka) model. *Journal of Personality and Social Psychology., 98*(2), 281–300. https://doi.org/10.1037/a0017908

Zimmermann, J., Schindler, S., Klaus, G., & Leising, D. (2017). The effect of dislike on accuracy and bias in person perception. *Social Psychological and Personality Science, 9*(1), 80–88. https://doi.org/10.1177/1948550617703167