

Nature and Impact of Reference Group Effects in Personality Assessment Data

Madeline R. Lenhausen, Christopher J. Hopwood & Wiebke Bleidorn

To cite this article: Madeline R. Lenhausen, Christopher J. Hopwood & Wiebke Bleidorn (2022): Nature and Impact of Reference Group Effects in Personality Assessment Data, Journal of Personality Assessment, DOI: [10.1080/00223891.2022.2132504](https://doi.org/10.1080/00223891.2022.2132504)

To link to this article: <https://doi.org/10.1080/00223891.2022.2132504>



Published online: 19 Oct 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



This article has been awarded the Centre for Open Science 'Open Data' badge.






This article has been awarded the Centre for Open Science 'Open Materials' badge.



This article has been awarded the Centre for Open Science 'Preregistered' badge.



Nature and Impact of Reference Group Effects in Personality Assessment Data

Madeline R. Lenhausen¹ , Christopher J. Hopwood² , and Wiebke Bleidorn² 

¹Department of Psychology, University of California, Davis; ²Department of Psychology, University of Zurich

ABSTRACT

People have a tendency to engage in social comparisons when evaluating and reporting on personality. This tendency and variation in who people compare their personality to is known as the reference group effect and has been largely discussed in cross-cultural research. However, reference group effects have implications beyond cross-cultural research and should be considered when collecting and interpreting personality data. In the present study, we examined the nature and impact of reference group comparisons on the Big Five personality traits in a sample of $N = 1194$ participants. Specifically, we examined what reference groups participants most believed they compared their personality to, and which reference group was actually the most impactful on trait scores. We found that most people believed they compared their personality to *people in general*. However, the most influential reference group was people the *same age* as the participants. Moreover, we found that people mostly engaged in between- as opposed to within-person comparisons when evaluating their own personality. Overall, our findings highlight that people have relatively little insight into the comparisons they engage in when make judgments on personality. Discussion focuses on theoretical and practical implications of our findings in light of personality assessment data.

ARTICLE HISTORY

Received 9 May 2022
Accepted 25 September 2022

Responding to personality items requires people to know *what* their personality is like and *how* to translate their personality perceptions into responses for personality questionnaires. Though much research has investigated the extent to which people know *what* their personality is like (Vazire & Carlson, 2010), considerably less research has focused on *how* people report on their personality (Carvalho et al., 2020), and still less on whether they know *how* they report on their personality. The overall goal of this study was to examine people's insight into how they go about rating themselves on questionnaires.

The nature of questionnaire items encourages people to compare themselves to others. There are different potential groups of others, or reference groups, and thus variation in which reference group people compare themselves can influence scores. This variation is referred to as the reference group effect (RGE; Heine et al., 2002). Understanding which reference groups people use when responding to personality questionnaires, and the extent to which they are aware of these comparisons, would shed light on the processes underlying personality reporting as well as people's insight into their own personality scaling. In the present preregistered study, we investigated a) which reference group people say they use when responding to a Big Five (John et al., 2008) personality questionnaire, b) which reference group comparison most closely matches their actual unprompted responses, and c) how the selection of different reference groups impacts personality assessment data.

Social comparisons

Responding to personality items is a complex event that evokes a series of cognitive processes including introspection, memory retrieval of relevant personality information, interpretation of personality information, application of personality information to the personality item, and, for the context of the present study, social comparisons (Böckenholt, 2012; Bornstein, 2007, 2009; Festinger, 1954; Rogers, 1974). Because there is no ground truth about the level of a personality trait that is normative, acceptable, or even possible, people responding to personality questionnaires presumably use internalized standards against which to rate themselves (Festinger, 1954; Heine et al., 2002; Wood et al., 2012). The selection of particular groups, such as humans in general, people from the respondent's own culture, or people who are like the respondent in other ways (e.g., gender or age) enables a benchmark against which the respondent can rate their own thoughts, feelings, and behaviors.

These reference group comparisons can impact personality assessment data. For instance, it is established that women tend to experience more anxiety, on average, than men (Costa et al., 2001; Weisberg et al., 2011). Thus, if a woman receives a question like "do you experience anxiety" and considers her own anxiety relative to other women, she may rate herself lower in anxiety than she would have if she compared herself to people in general. The use of different reference groups by different people can thus introduce error into personality assessment data and complicate interpretation. This phenomenon has been used to explain

findings that cultures do not differ as much expected in personality traits (Heine et al., 2002). For example, RGE could lead to an absence of average differences between people from highly extraverted and highly introverted cultures, if people within each culture are only comparing themselves to other people within that same culture. Indeed, research on social comparisons suggests that people do differ in who they compare themselves to, largely preferring people in their immediate social network (Pachur & Schulze, 2022).

Awareness of comparisons

The selection of reference groups is typically thought to be unconscious and spontaneous (Baldwin & Holmes, 1987; Blanton & Stapel, 2008; Credé et al., 2010; Festinger, 1954; Heine et al., 2002; Latane & Darley, 1970). In contrast, personality questionnaires explicitly ask individuals to evaluate their own personality, which brings the individual's self-perceptions to the forefront of their awareness (Bornstein, 2009). This is particularly the case when the instructions specify a reference group, although many do not. Even when they do, it is possible that respondents may not have access to actual levels of a trait or behavior in certain populations (e.g., humans in general). Given this combination of factors, RGEs likely reflect a combination of both conscious consideration ("how extraverted am I compared to other people I know?") and unconscious impulse ("I think I am pretty extraverted"). Thus far, research has not established the degree to which people are consciously aware of the reference group they use while responding to personality items.

Credé et al. (2010) collected self-reports of conscientiousness in a "reference group-free" condition (no reference group prompted) as well as four additional reference group conditions (immediate family, people of same age and gender, close friends and peers, people in general). They found that scores on the reference-group-free condition of conscientiousness were most similar to conscientious scores for the "people in general" and "same age and gender" reference group conditions. Notably, these were the reference groups participants reported as *least* important when reporting on their personality, and instead reported "immediate family" and "close friends and peers" reference groups as more important when responding to personality items. These findings suggested that participants were not aware of the reference groups they compare themselves to when reporting on their own personality, particularly conscientiousness.

The present study

In this pre-registered study (<https://osf.io/qm5sd/>), we aimed to go beyond existing research by addressing three questions.

First, *who do people believe they compare themselves to when answering questions about their personality?* We expected to find individual differences in primary reference group choice, such that particular reference groups would be preferred over others, on average. We examined the following reference group categories: 1) people in general, 2) close

others, 3) same age, 4) same gender, 5) past self, and 6) ideal self. Our selection of reference categories reflected three considerations. First, we selected reference categories that had been used in previous RG manipulation studies (e.g., Credé et al., 2010). Second, we included reference categories that reflected instructions in established personality trait questionnaires. For example, the International Personality Item Pool (IPIP; Goldberg, 1999; Goldberg et al., 2006) prompts participants to: "Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age". Likewise, certain items (but not all) in the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) prompt participants to consider a reference group such as "I have fewer fears than most people" (McCrae et al., 2007). Third, we added an additional category that, if used, had important implications for research on personality trait development (i.e., past self), with particular focus on within-person change. We predicted that most people would report using the *close others* (e.g., romantic partner, close friends/family) reference group when responding to personality items (H1). Past research found that people mostly preferred "peers" and "family" for reference groups (Credé et al., 2010) likely because these close others are the people most readily available and present in an individual's everyday life.

Second, *who do people actually compare themselves to when answering questions about their personality?* We addressed this question by quantifying the similarity between participants' personality scores where no reference group was prompted to personality scores when different reference groups (i.e., people in general, close others, same age, same gender, past self, ideal self) were specifically prompted. We expected that the personality scores from the unprompted condition would align strongest with personality scores from the reference group condition that most people believed they compared their personality to (H2). For instance, if we found support for H1, we would expect that unprompted personality scores would be most similar to scores when respondents were prompted to use *close others* as their reference group.

Third, *how do individual differences in choice of reference group impact personality data?* We explored this question by calculating individual similarity between trait scores from the unprompted condition to the reference group condition corresponding to each participant's reported primary reference group. High similarity above and beyond the similarity with other reference groups would imply that people are aware of the reference group they refer to when reporting on their own personality, and that this impacts their scores.

Method

Sample

We recruited 1227 participants with a target sample of 1000 participants from Prolific— an online data collection service. We asked one exclusionary item to exclude participants who were not paying attention during the study: "Please select 'Somewhat agree' for this question. Thank you for paying

attention.” We removed participants who did not finish the study and/or failed the attention check ($n = 68$), giving us a total of $N = 1194$, 51.42% female, 46.73% male, 1.84% non-binary/other, $M_{age} = 36.91$, $SD_{age} = 12.92$. This study was granted exemption (IRB#1706865-2) by the UC Davis Institutional Review Board because it only included surveys of which ethical approval is not required. Informed consent was obtained electronically within the survey.

Measures

Personality

We used the 20-item Mini International Personality Item Pool (Mini-IPIP; Donnellan et al., 2006) to score participants' Big Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, openness) on a 9-point Likert scale ranging from 1 = “Strongly disagree” to 9 = “Strongly agree”. We used the original questionnaire to score participants' unprompted personality. We used a modified version of the questionnaire with specific reference groups mentioned for each item to score participants' reference-prompted personality.

Perceived reference group

We assessed participants' perceived reference group with the following item: “When people answer questions about themselves, they usually compare themselves to others. Please indicate which group you compared yourself to the most when completing the previous items.” Reference groups were: *other people in general*, *people I know well* (e.g., *romantic partner*, *close friends/family*), *people my same age*, *people my same gender*, *what I was like in the past* (*my past self*), and *what I would like to be like* (*my ideal self*). Participants were also given the option to select “other” and specify a different reference group. The order of reference groups listed was randomized. After the initial choice of primary reference group, we asked participants if they use any other reference groups with the following item: “Sometimes people compare themselves to multiple groups when answering questions about themselves. Please indicate any other groups you compared yourself to when completing the previous items.” They were then given the option of selecting any of the reference groups they did not select initially. Finally, we asked participants to rate all reference groups they compared themselves to using a 10-point Likert scale with the following item “Please rate each of the following groups, where 1 indicates that you never compared yourself to the group when answering questions about yourself, whereas 10 indicates that you always compared yourself to members of the group. Please do not select the same response option more than once.”

Procedure

We first administered the Mini-IPIP personality questionnaire with no reference group prompt to measure participants' unprompted personality scores. Following the unprompted personality assessment, we asked participants to

provide reference group ratings as described above. Next, we readministered the Mini-IPIP six times with the corresponding reference group prompted for each item. For example, the first re-administration of the questionnaire indicated *people in general*, the second re-administration indicated *people I know well* and so on. For example, the original item “Am the life of the party” had 6 different versions corresponding to each reference group such as “Am the life of the party in comparison to my past self.” Thus, the respondents completed 120 trait questions across reference group conditions (6 reference groups x 20 items). Reference group condition was randomized across participants.

Analyses

Aim 1: Perceived reference group

We used a χ^2 test to examine differences in observed frequencies of reference group choices, against the null hypothesis that all reference groups would be equally preferred. We calculated binomial tests between each reference group distribution and used Bonferroni adjustment to correct for multiple pairwise comparisons, subsequent to a significant χ^2 . This allowed us to gauge whether the reference group *close others* was most preferred across participants (H1), and if not, which reference group was the most preferred.¹

Aim 2: General impacts of reference groups

We used two methods to ascertain the reference group condition that produced personality scores most similar to unprompted personality scores, in general (H2). First, we assessed the mean squared differences between individuals' unprompted Big Five trait scores and each of the Big Five trait scores within each reference group condition. Smaller mean squared differences indicate greater similarity in *absolute levels* of unprompted and prompted trait scores. Second, we examined both individual trait and profile correlations. For individual traits, we calculated the correlation between unprompted trait scores and scores from each reference group condition (e.g., correlation between openness in the unprompted condition and openness in the *ideal self* condition). This indicates the similarity of different reference group prompts to unprompted responding in terms of the *rank-ordering of individuals*. Next, we computed both overall and distinct profile correlations between the unprompted Big Five trait scores and each of the prompted reference group trait scores. Distinct profile correlations test the association between mean-centered scores, thus representing similarity in deviance from the normative profile (Biesanz & Human, 2010; Furr, 2008).² These indicate the *similarity of trait configurations* across conditions. We compared mean-level differences and correlations using 95% confidence

¹We additionally examined predictors of primary reference group choice using participants' unprompted Big Five trait scores, age, and gender. However, only age emerged as a significant predictor (see Tables S1-S4).

²This is a deviation from our original analysis plan as we did not pre-register it, and as such is exploratory. We thank a reviewer for suggesting this analysis.

intervals (CIs). Non-overlapping CIs would indicate that the mean squared differences or correlations are different from one another in their degree of similarity to the unprompted personality scores.

Aim 3: Individual differences in reference group choice

Whereas our second aim focused on the similarity of data from different reference group prompts to unprompted scores for all participants, our third aim focused on how individual differences in reference group choice impact personality assessment data. Specifically, we sought to test whether the data from the reference groups people reported using was most similar to unprompted data. This would suggest that people have insight regarding the reference groups they select, and that variation in reference group selection could introduce trait-irrelevant variation in personality data. To test Aim 3, we divided participants into subsamples based on their primary reference group choice. We then investigated the similarity between unprompted trait scores and the trait scores from each participant’s perceived reference group condition using mean-level squared differences and individual trait and profile correlations, as in Aim 2.

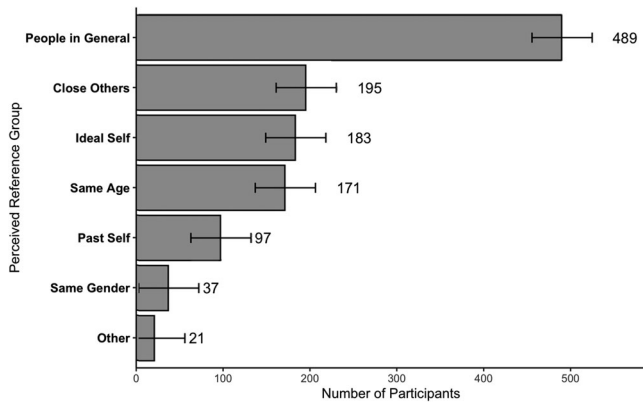


Figure 1. Distribution of participants' reported primary reference group.

Table 1. Mean squared differences between participants' unprompted trait scores and trait scores in each reference group condition.

	M ² Diff (SD)	95% CI	M ² Diff (SD)	95% CI	M ² Diff (SD)	95% CI	M ² Diff (SD)	95% CI	M ² Diff (SD)	95% CI
	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
^g People in General	1.40 (2.75) _{api}	[1.24, 1.55]	1.47 (2.74) _{pi}	[1.32, 1.63]	1.03 (1.93) _{cpi}	[0.92, 1.14]	1.18 (2.66) _{pi}	[1.03, 1.33]	1.15 (2.39) _{pi}	[1.02, 1.29]
^c Close Others	1.57 (2.92) _{api}	[1.40, 1.73]	1.77 (2.89) _{api}	[1.61, 1.94]	1.29 (2.30) _{gapi}	[1.16, 1.42]	1.46 (2.92) _{api}	[1.29, 1.63]	1.39 (2.72) _{api}	[1.24, 1.55]
^a Same Age	1.07 (2.32) _{gcdpi}	[0.94, 1.20]	1.25 (2.78) _{cpi}	[1.09, 1.41]	0.92 (1.99) _{cpi}	[0.80, 1.03]	0.92 (1.84) _{cdpi}	[0.82, 1.03]	0.91 (2.06) _{cdpi}	[0.80, 1.03]
^d Same Gender	1.39 (2.78) _{api}	[1.23, 1.55]	1.52 (2.99) _{pi}	[1.35, 1.69]	1.05 (2.13) _{pi}	[0.93, 1.17]	1.21 (2.81) _{api}	[1.05, 1.36]	1.18 (2.48) _{api}	[1.04, 1.32]
^p Past Self	3.31 (4.43) _{gcd}	[3.06, 3.56]	4.10 (5.08) _{gcdi}	[3.81, 4.39]	3.27 (5.01) _{gcd}	[2.99, 3.56]	5.50 (6.97) _{gcdi}	[5.10, 5.89]	4.93 (5.67) _{gcdi}	[4.60, 5.25]
^l Ideal Self	3.27 (6.14) _{gcd}	[2.92, 3.62]	3.03 (5.02) _{gcdp}	[2.75, 3.32]	2.78 (4.70) _{gcd}	[2.51, 3.04]	2.97 (5.32) _{gcdp}	[2.67, 3.27]	3.34 (5.93) _{gcdp}	[3.01, 3.68]

Note. Subscripts (labeled in first column) indicate which values differ from one another based on non-overlapping 95% CIs.

Table 2. Profile and individual trait correlations between participants' unprompted trait scores and trait scores in each reference group condition.

	r	95% CI	r	95% CI	r	95% CI	r	95% CI	r	95% CI	r	95% CI	r	95% CI
	Overall Profile		Distinctive Profile		Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
^g People in General	.94 _{cpi}	[.94, .95]	.89 _{cpi}	[.88, .90]	.83 _{api}	[.80, .84]	.77 _{pi}	[.75, .79]	.83 _{pi}	[.81, .85]	.82 _{cpi}	[.80, .84]	.79 _{cpi}	[.77, .81]
^c Close Others	.92 _{gadpi}	[.91, .93]	.85 _{gadpi}	[.84, .87]	.80 _{api}	[.77, .82]	.74 _{api}	[.71, .76]	.79 _{api}	[.77, .81]	.77 _{gapi}	[.74, .79]	.74 _{gapi}	[.71, .77]
^a Same Age	.95 _{cdpi}	[.95, .96]	.90 _{cdpi}	[.89, .91]	.86 _{gcdpi}	[.85, .88]	.79 _{cpi}	[.76, .81]	.84 _{cpi}	[.82, .85]	.85 _{cdpi}	[.83, .86]	.82 _{cdpi}	[.80, .84]
^d Same Gender	.94 _{capi}	[.93, .95]	.88 _{capi}	[.87, .89]	.81 _{api}	[.79, .83]	.75 _{pi}	[.72, .77]	.81 _{pi}	[.79, .83]	.80 _{api}	[.78, .82]	.77 _{api}	[.75, .79]
^p Past Self	.22 _{gcdi}	[.16, .27]	.21 _{gcdi}	[.15, .26]	.47 _{gcd}	[.42, .51]	.48 _{gcd}	[.43, .52]	.34 _{gcdi}	[.28, .39]	-.04 _{gcdi}	[-.10, .02]	.24 _{gcdi}	[.18, .29]
^l Ideal Self	.78 _{gcdp}	[.75, .80]	.68 _{gcdp}	[.65, .71]	.54 _{gcd}	[.49, .57]	.56 _{gcd}	[.52, .60]	.63 _{gcdp}	[.60, .67]	.56 _{gcdp}	[.52, .60]	.48 _{gcdp}	[.44, .53]

Note. Subscripts (labeled in first column) indicate which values differ from one another respective to each profile/trait based on non-overlapping 95% CIs. All correlations were significant at $p < .001$, with an exception to past self for neuroticism which was nonsignificant.

Results

Perceived reference group

Reference group frequencies in the overall sample are shown in Figure 1, ranked by most to least chosen. Results of the χ^2 test indicated that some reference groups were preferred over others. Pairwise binomial tests demonstrated that the most frequently chosen reference group was *people in general*. The second most common groups were *close others*, *ideal self*, and *same age*. People used past self and same gender relatively infrequently, and few people chose any other reference group.

General impacts of reference group

Mean squared differences and individual trait correlations between unprompted trait scores and trait scores from each reference group condition are shown in Tables 1 and 2. Results were relatively consistent across traits, suggesting that reference group effects do not depend on which trait is being assessed. Across all traits, the smallest mean squared difference scores and strongest correlations (indicating highest similarity to unprompted trait scores) were found for the reference group condition *same age*. Although confidence intervals overlapped across groups, *same age* had the smallest mean-difference scores and the strongest correlations for all five traits and both the overall and distinctive profile. This is particularly interesting given that the results of Aim 1 suggest that people think they compare themselves to *people in general*. This provides strong evidence that, whatever groups people might think they are comparing themselves to, they tend to produce responses that indicate they are comparing themselves to people their own age when responding to personality assessment items.

The *people in general*, *close others*, and *same gender* prompts clustered together as a group as somewhat less similar to unprompted scores, although again confidence

Table 3. Mean squared differences between participants' unprompted trait scores and trait scores in each reference group condition by primary RG choice.

	M^2 Diff (SD)	95% CI	M^2 Diff (SD)	95% CI	M^2 Diff (SD)	95% CI	M^2 Diff (SD)	95% CI	M^2 Diff (SD)	95% CI
RG Condition by RG Choice	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
People in General (n = 489)										
People in General	1.12 (2.13)	[0.94, 1.31]	1.37 (2.64)	[1.14, 1.61]	0.87 (1.72)	[0.71, 1.02]	1.03 (2.85)	[0.78, 1.29]	1.06 (2.27)	[0.86, 1.27]
Close Others	1.39 (2.35)	[1.19, 1.60]	1.78 (2.83)	[1.53, 2.03]	1.17 (2.27)	[0.97, 1.37]	1.45 (3.16)	[1.17, 1.73]	1.26 (2.50)	[1.04, 1.49]
Same Age	0.88 (1.71)	[0.73, 1.03]	1.07 (1.96)	[0.90, 1.24]	0.83 (1.69)	[0.68, 0.98]	0.75 (1.57)	[0.61, 0.89]	0.81 (1.93)	[0.64, 0.98]
Same Gender	1.24 (2.35)	[1.04, 1.45]	1.42 (2.31)	[1.21, 1.62]	0.95 (1.90)	[0.78, 1.12]	1.00 (2.71)	[0.76, 1.24]	1.02 (2.28)	[0.82, 1.23]
Past Self	3.68 (4.56)	[3.28, 4.09]	4.34 (5.27)	[3.87, 4.81]	3.30 (5.16)	[2.84, 3.76]	5.53 (6.75)	[4.93, 6.12]	5.28 (5.82)	[4.76, 5.80]
Ideal Self	2.96 (5.10)	[2.51, 3.42]	3.01 (4.18)	[2.63, 3.38]	2.64 (4.00)	[2.29, 3.00]	3.03 (5.63)	[2.53, 3.53]	3.65 (5.86)	[3.13, 4.17]
Close Others (n = 195)										
People in General	1.29 (2.14)	[0.99, 1.60]	1.24 (2.25)	[0.93, 1.56]	1.12 (2.31)	[0.80, 1.44]	1.17 (2.13)	[0.87, 1.47]	1.07 (1.69)	[0.83, 1.31]
Close Others	1.75 (3.59)	[1.25, 2.26]	1.59 (2.49)	[1.24, 1.94]	1.49 (2.50)	[1.13, 1.84]	1.33 (2.38)	[1.00, 1.67]	1.51 (3.18)	[1.07, 1.96]
Same Age	1.37 (3.19)	[0.92, 1.82]	1.24 (2.49)	[0.89, 1.59]	0.92 (1.97)	[0.65, 1.20]	0.90 (1.59)	[0.67, 1.12]	0.95 (1.59)	[0.73, 1.17]
Same Gender	1.49 (2.78)	[1.10, 1.89]	1.50 (2.78)	[1.11, 1.89]	1.11 (2.15)	[0.80, 1.41]	1.07 (1.60)	[0.85, 1.30]	1.23 (2.11)	[0.94, 1.53]
Past Self	3.16 (4.29)	[2.55, 3.76]	4.40 (5.82)	[3.58, 5.22]	3.88 (5.59)	[3.09, 4.66]	5.55 (7.27)	[4.53, 6.57]	4.41 (5.03)	[3.70, 5.11]
Ideal Self	3.56 (7.59)	[2.49, 4.62]	2.87 (5.17)	[2.14, 3.59]	3.18 (5.40)	[2.42, 3.94]	2.64 (4.04)	[2.07, 3.20]	3.05 (5.84)	[2.23, 3.87]
Same Age (n = 171)										
People in General	1.35 (2.90)	[0.91, 1.78]	1.63 (2.81)	[1.21, 2.06]	0.99 (1.46)	[0.77, 1.21]	1.38 (2.50)	[1.00, 1.75]	1.22 (3.22)	[0.74, 1.70]
Close Others	1.48 (2.68)	[1.08, 1.88]	1.86 (3.45)	[1.35, 2.38]	1.21 (1.98)	[0.91, 1.51]	1.60 (2.66)	[1.20, 1.99]	1.46 (3.15)	[0.98, 1.93]
Same Age	0.94 (1.79)	[0.67, 1.20]	1.48 (2.69)	[1.07, 1.88]	0.90 (1.64)	[0.66, 1.15]	0.83 (1.42)	[0.62, 1.04]	0.86 (1.56)	[0.62, 1.09]
Same Gender	1.13 (1.83)	[0.86, 1.41]	1.46 (2.20)	[1.13, 1.79]	1.05 (1.79)	[0.78, 1.32]	1.20 (1.97)	[0.91, 1.50]	1.14 (2.01)	[0.84, 1.44]
Past Self	3.05 (3.82)	[2.48, 3.63]	3.56 (4.63)	[2.86, 4.25]	2.89 (3.79)	[2.32, 3.46]	5.08 (6.52)	[4.11, 6.06]	4.20 (5.33)	[3.40, 5.00]
Ideal Self	3.42 (6.40)	[2.46, 4.38]	3.06 (4.34)	[2.41, 3.71]	2.65 (4.49)	[1.98, 3.33]	2.90 (4.81)	[2.18, 3.62]	2.72 (4.31)	[2.08, 3.37]
Same Gender (n = 37)										
People in General	1.62 (2.39)	[0.85, 2.39]	1.57 (3.37)	[0.48, 2.65]	0.89 (1.40)	[0.44, 1.34]	1.30 (2.15)	[0.61, 1.99]	1.54 (2.37)	[0.78, 2.30]
Close Others	1.65 (2.69)	[0.79, 2.52]	1.33 (1.94)	[0.71, 1.96]	1.16 (1.70)	[0.61, 1.70]	1.30 (1.97)	[0.66, 1.93]	1.47 (1.67)	[0.93, 2.01]
Same Age	0.81 (1.39)	[0.37, 1.26]	1.01 (1.45)	[0.54, 1.47]	0.69 (1.38)	[0.25, 1.14]	1.13 (2.41)	[0.35, 1.90]	1.02 (1.71)	[0.47, 1.57]
Same Gender	1.65 (2.49)	[0.85, 2.45]	1.03 (1.97)	[0.40, 1.67]	0.79 (1.44)	[0.33, 1.25]	0.58 (0.83)	[0.32, 0.85]	1.01 (1.57)	[0.50, 1.51]
Past Self	2.27 (3.40)	[1.17, 3.36]	3.49 (4.52)	[2.04, 4.95]	2.22 (3.39)	[1.13, 3.31]	5.24 (7.79)	[2.73, 7.76]	4.96 (6.94)	[2.73, 7.20]
Ideal Self	4.02 (9.08)	[1.10, 6.95]	2.27 (6.64)	[0.13, 4.41]	2.40 (4.11)	[1.08, 3.73]	2.65 (5.36)	[0.92, 4.37]	2.09 (4.89)	[0.52, 3.67]
Past Self (n = 97)										
People in General	1.66 (2.89)	[1.09, 2.24]	1.78 (3.21)	[1.14, 2.42]	0.85 (1.41)	[0.56, 1.13]	0.99 (1.43)	[0.71, 1.28]	0.96 (1.35)	[0.70, 1.23]
Close Others	1.49 (2.60)	[0.97, 2.01]	1.60 (2.69)	[1.06, 2.13]	1.41 (2.42)	[0.93, 1.89]	1.43 (2.91)	[0.85, 2.01]	1.26 (1.90)	[0.89, 1.64]
Same Age	1.07 (1.84)	[0.70, 1.43]	1.27 (2.20)	[0.84, 1.71]	0.95 (1.89)	[0.57, 1.32]	1.25 (1.89)	[0.87, 1.62]	1.03 (1.99)	[0.63, 1.42]
Same Gender	1.28 (2.32)	[0.82, 1.74]	1.33 (2.12)	[0.90, 1.75]	0.93 (1.53)	[0.62, 1.23]	1.40 (2.08)	[0.99, 1.82]	1.46 (2.66)	[0.93, 1.99]
Past Self	3.00 (4.41)	[2.13, 3.88]	3.92 (4.83)	[2.96, 4.88]	3.64 (6.21)	[2.41, 4.88]	5.15 (7.40)	[3.67, 6.62]	4.50 (4.80)	[3.55, 5.46]
Ideal Self	3.38 (5.27)	[2.33, 4.43]	2.13 (3.14)	[1.51, 2.76]	2.68 (3.80)	[1.93, 3.44]	3.74 (6.95)	[2.35, 5.12]	3.17 (4.78)	[2.22, 4.12]
Ideal Self (n = 183)										
People in General	2.11 (4.12)	[1.51, 2.71]	1.59 (2.89)	[1.17, 2.01]	1.43 (2.42)	[1.08, 1.78]	1.40 (3.37)	[0.91, 1.89]	1.20 (2.02)	[0.91, 1.49]
Close Others	1.97 (3.83)	[1.41, 2.52]	2.02 (3.17)	[1.56, 2.48]	1.42 (2.41)	[1.07, 1.77]	1.35 (2.75)	[0.95, 1.75]	1.46 (2.66)	[1.08, 1.85]
Same Age	1.37 (3.18)	[0.91, 1.83]	1.60 (4.86)	[0.90, 2.31]	1.15 (2.99)	[0.72, 1.58]	1.22 (2.68)	[0.83, 1.61]	1.04 (3.02)	[0.61, 1.48]
Same Gender	1.85 (4.32)	[1.22, 2.47]	1.92 (5.06)	[1.19, 2.65]	1.39 (3.20)	[0.93, 1.85]	1.79 (4.68)	[1.11, 2.47]	1.35 (3.35)	[0.87, 1.84]
Past Self	3.18 (5.06)	[2.45, 3.92]	3.89 (4.43)	[3.24, 4.53]	3.11 (4.68)	[2.44, 3.79]	5.97 (7.39)	[4.90, 7.04]	5.16 (6.18)	[4.26, 6.05]
Ideal Self	3.19 (5.21)	[2.44, 3.95]	3.76 (7.37)	[2.70, 4.83]	2.98 (6.17)	[2.09, 3.88]	2.75 (4.44)	[2.10, 3.39]	3.70 (7.96)	[2.55, 4.86]
Other (n = 21)										
People in General	1.45 (3.72)	[-0.15, 3.04]	2.20 (3.69)	[0.62, 3.77]	1.68 (3.25)	[0.30, 3.07]	1.67 (2.35)	[0.66, 2.68]	2.99 (6.17)	[0.36, 5.63]
Close Others	1.46 (2.38)	[0.44, 2.47]	2.14 (2.88)	[0.90, 3.37]	1.74 (2.74)	[0.56, 2.91]	3.08 (5.55)	[0.71, 5.46]	2.39 (4.31)	[0.55, 4.23]
Same Age	1.58 (2.95)	[0.32, 2.84]	1.11 (1.20)	[0.59, 1.62]	1.23 (1.70)	[0.50, 1.95]	1.24 (2.03)	[0.37, 2.11]	1.38 (3.03)	[0.08, 2.68]
Same Gender	2.06 (3.79)	[0.44, 3.68]	2.90 (4.54)	[0.96, 4.84]	1.06 (1.13)	[0.57, 1.54]	2.07 (2.21)	[1.12, 3.01]	2.27 (4.33)	[0.42, 4.13]
Past Self	2.74 (2.87)	[1.51, 3.97]	4.07 (4.25)	[2.26, 5.89]	1.63 (2.27)	[0.66, 2.60]	4.92 (5.48)	[2.58, 7.27]	7.22 (6.39)	[4.49, 9.96]
Ideal Self	5.33 (12.97)	[-0.22, 10.88]	4.14 (4.90)	[2.04, 6.23]	2.52 (4.60)	[0.55, 4.49]	3.99 (9.37)	[-0.01, 8.00]	4.14 (5.32)	[1.86, 6.42]

Note. Reference Group (RG) Choice refers to the primary reference group that participants reported using when answering personality items. Participants are divided into subsamples based on this choice.

intervals varied across trait and analysis somewhat. This suggests that these reference groups may also be operating as people respond to personality assessment questions in the absence of specific prompts. In contrast, similarity was rather low between unprompted scores and *ideal self* prompt, and lower still for the *past self* prompt. This suggests that people generally make between-person rather than within-person comparisons when answering personality assessment questions.

Individual differences in reference group choice

To examine how individual differences in reference group choice impacted responding, we first divided participants into subsamples based on their primary (perceived)

reference group choice, and then examined similarity of their trait scores with the unprompted trait scores within each subsample in terms of absolute level (Table 3) and rank order of people and traits (Table 4). We note that results from subsamples reflecting four most common reference groups choices (i.e., *people in general*, *close others*, *same age*, *ideal self*) are most reliable.

Again, the most consistent pattern was that the *same age* prompt produced data that was most similar to unprompted responses (see Figure 2). The mean difference was smallest and correlation largest for the *same age* prompt for nearly all traits in all subgroups, and in the few exceptions the difference between this value and the largest value was not statistically significant. This surprising finding suggests that the tendency for people to compare themselves to people of

Table 4. Trait correlations between participants' unprompted trait scores and trait scores in each reference group condition by primary RG choice.

RG Condition by RG Choice	Extraversion			Agreeableness			Conscientiousness			Neuroticism			Openness		
	<i>r</i>	95% CI	<i>p</i>	<i>r</i>	95% CI	<i>p</i>	<i>r</i>	95% CI	<i>p</i>	<i>r</i>	95% CI	<i>p</i>	<i>r</i>	95% CI	<i>p</i>
People in General (n = 489)															
People in General	.88	[.85, .89]	<.001	.80	[.76, .83]	<.001	.87	[.85, .89]	<.001	.86	[.83, .88]	<.001	.82	[.78, .84]	<.001
Close Others	.83	[.80, .86]	<.001	.76	[.72, .80]	<.001	.82	[.79, .85]	<.001	.79	[.75, .82]	<.001	.77	[.73, .80]	<.001
Same Age	.90	[.88, .91]	<.001	.82	[.79, .85]	<.001	.86	[.83, .88]	<.001	.89	[.87, .90]	<.001	.85	[.83, .88]	<.001
Same Gender	.85	[.82, .87]	<.001	.78	[.74, .81]	<.001	.84	[.81, .87]	<.001	.85	[.82, .87]	<.001	.81	[.77, .84]	<.001
Past Self	.48	[.41, .54]	<.001	.48	[.41, .54]	<.001	.36	[.28, .44]	<.001	.00	[-.09, .09]	.947	.23	[.15, .31]	<.001
Ideal Self	.60	[.55, .66]	<.001	.60	[.54, .66]	<.001	.68	[.63, .73]	<.001	.58	[.52, .64]	<.001	.49	[.41, .55]	<.001
Close Others (n = 195)															
People in General	.83	[.78, .87]	<.001	.78	[.71, .83]	<.001	.81	[.75, .85]	<.001	.81	[.75, .85]	<.001	.81	[.76, .85]	<.001
Close Others	.78	[.72, .83]	<.001	.75	[.68, .80]	<.001	.76	[.69, .81]	<.001	.77	[.70, .82]	<.001	.72	[.65, .78]	<.001
Same Age	.82	[.77, .86]	<.001	.76	[.70, .81]	<.001	.83	[.78, .87]	<.001	.83	[.78, .87]	<.001	.81	[.75, .85]	<.001
Same Gender	.80	[.74, .84]	<.001	.73	[.66, .79]	<.001	.80	[.74, .85]	<.001	.80	[.74, .85]	<.001	.75	[.68, .81]	<.001
Past Self	.46	[.34, .56]	<.001	.47	[.36, .58]	<.001	.18	[.04, .31]	.012	.02	[-.12, .16]	.783	.33	[.20, .45]	<.001
Ideal Self	.50	[.39, .60]	<.001	.60	[.50, .69]	<.001	.61	[.51, .69]	<.001	.59	[.49, .67]	<.001	.54	[.43, .63]	<.001
Same Age (n = 171)															
People in General	.81	[.75, .86]	<.001	.76	[.68, .81]	<.001	.82	[.77, .87]	<.001	.77	[.71, .83]	<.001	.79	[.72, .84]	<.001
Close Others	.77	[.71, .83]	<.001	.72	[.64, .78]	<.001	.79	[.73, .84]	<.001	.74	[.66, .80]	<.001	.74	[.67, .80]	<.001
Same Age	.87	[.83, .90]	<.001	.73	[.65, .79]	<.001	.83	[.78, .87]	<.001	.87	[.82, .90]	<.001	.84	[.78, .88]	<.001
Same Gender	.84	[.78, .88]	<.001	.73	[.65, .79]	<.001	.79	[.73, .84]	<.001	.79	[.72, .84]	<.001	.79	[.73, .84]	<.001
Past Self	.47	[.35, .58]	<.001	.49	[.37, .60]	<.001	.38	[.24, .50]	<.001	-.05	[-.20, .10]	.486	.28	[.13, .41]	<.001
Ideal Self	.49	[.37, .60]	<.001	.55	[.43, .64]	<.001	.61	[.50, .69]	<.001	.54	[.43, .64]	<.001	.57	[.46, .66]	<.001
Same Gender (n = 37)															
People in General	.78	[.61, .88]	<.001	.69	[.46, .83]	<.001	.87	[.76, .93]	<.001	.78	[.60, .88]	<.001	.71	[.51, .84]	<.001
Close Others	.78	[.61, .88]	<.001	.70	[.48, .83]	<.001	.84	[.70, .91]	<.001	.75	[.56, .86]	<.001	.71	[.50, .84]	<.001
Same Age	.89	[.80, .94]	<.001	.78	[.61, .88]	<.001	.90	[.82, .95]	<.001	.77	[.60, .88]	<.001	.82	[.67, .90]	<.001
Same Gender	.77	[.59, .88]	<.001	.82	[.67, .90]	<.001	.88	[.78, .94]	<.001	.90	[.81, .95]	<.001	.82	[.67, .90]	<.001
Past Self	.67	[.45, .82]	<.001	.47	[.18, .69]	.003	.63	[.38, .79]	<.001	.14	[-.19, .44]	.414	.23	[-.10, .51]	.173
Ideal Self	.43	[.12, .66]	.009	.65	[.42, .81]	<.001	.73	[.54, .85]	<.001	.58	[.31, .76]	<.001	.59	[.32, .76]	<.001
Past Self (n = 97)															
People in General	.76	[.66, .83]	<.001	.76	[.66, .83]	<.001	.83	[.75, .88]	<.001	.85	[.79, .90]	<.001	.80	[.72, .86]	<.001
Close Others	.78	[.69, .85]	<.001	.78	[.69, .85]	<.001	.74	[.63, .82]	<.001	.78	[.69, .85]	<.001	.74	[.63, .82]	<.001
Same Age	.85	[.78, .89]	<.001	.82	[.74, .88]	<.001	.80	[.72, .86]	<.001	.80	[.71, .86]	<.001	.78	[.69, .85]	<.001
Same Gender	.80	[.72, .86]	<.001	.80	[.71, .86]	<.001	.80	[.71, .86]	<.001	.77	[.68, .84]	<.001	.70	[.59, .79]	<.001
Past Self	.46	[.28, .60]	<.001	.52	[.36, .65]	<.001	.32	[.13, .49]	.001	-.16	[-.35, .04]	.123	.24	[.04, .42]	.020
Ideal Self	.44	[.26, .59]	<.001	.64	[.50, .74]	<.001	.49	[.32, .63]	<.001	.38	[.20, .54]	<.001	.37	[.18, .53]	<.001
Ideal Self (n = 183)															
People in General	.69	[.60, .76]	<.001	.72	[.65, .79]	<.001	.77	[.71, .83]	<.001	.76	[.70, .82]	<.001	.75	[.68, .81]	<.001
Close Others	.71	[.63, .77]	<.001	.64	[.55, .72]	<.001	.79	[.73, .84]	<.001	.75	[.68, .81]	<.001	.71	[.63, .78]	<.001
Same Age	.78	[.72, .83]	<.001	.72	[.64, .78]	<.001	.78	[.72, .83]	<.001	.78	[.71, .83]	<.001	.77	[.70, .82]	<.001
Same Gender	.72	[.64, .78]	<.001	.65	[.55, .72]	<.001	.75	[.68, .81]	<.001	.66	[.57, .74]	<.001	.71	[.64, .78]	<.001
Past Self	.42	[.29, .53]	<.001	.43	[.30, .54]	<.001	.32	[.18, .44]	<.001	-.20	[-.34, -.06]	.006	.13	[-.02, .27]	.092
Ideal Self	.51	[.39, .61]	<.001	.35	[.21, .47]	<.001	.63	[.54, .71]	<.001	.60	[.50, .69]	<.001	.43	[.30, .54]	<.001
Other (n = 21)															
People in General	.78	[.52, .90]	<.001	.49	[.07, .76]	.026	.48	[.07, .76]	.026	.81	[.58, .92]	<.001	.47	[.05, .75]	.032
Close Others	.83	[.63, .93]	<.001	.41	[-.03, .71]	.065	.51	[.10, .77]	.018	.63	[.27, .83]	.002	.59	[.21, .81]	.005
Same Age	.84	[.64, .93]	<.001	.89	[.74, .95]	<.001	.78	[.52, .91]	<.001	.83	[.62, .93]	<.001	.72	[.42, .88]	<.001
Same Gender	.71	[.41, .88]	<.001	.56	[.17, .80]	.008	.76	[.48, .90]	<.001	.76	[.50, .90]	<.001	.58	[.20, .81]	.006
Past Self	.51	[.10, .77]	.017	.38	[-.07, .69]	.093	.50	[.09, .77]	.021	-.02	[-.45, .41]	.930	.19	[-.27, .57]	.419
Ideal Self	-.08	[-.49, .37]	.733	.08	[-.37, .49]	.739	.29	[-.17, .64]	.207	.38	[-.07, .69]	.093	.05	[-.39, .47]	.827

Note. Reference Group (RG) Choice refers to the primary reference group that participants reported using when answering personality items. Participants are divided into subsamples based on this choice.

their age is so strong that it even overcomes peoples' specific beliefs about the people they are comparing themselves to when responding to personality questions.

The general pattern was, in fact, similar to what we found for the overall sample in Aim 2, in that *people in general*, *close others*, and people the *same gender* clustered together in producing the next most similar results, and within-person comparisons to *past* or *ideal self* produced the least similar results. This was even the case among people who reported using the *past* or *ideal self* for comparisons, suggesting that people actually have relatively little insight into the reference groups they use when completing personality assessment questionnaires. In fact, while there was some evidence for insight regarding reference group choice, it was very modest. As one specific example, the squared

differences and correlations for extraversion do tend to suggest greater similarity between the *people in general* prompted data and unprompted data among people who reported using people in general as a reference group (i.e., 1.12 is the lowest value for that prompt in the Extraversion column of Table 3 and .88 is the highest correlation for that prompt in the Extraversion column of Table 4). That being said, these differences are quite small, and there are many other instances in which similarity is higher for a different prompt than the explicitly recalled reference group. As one example, Extraversion scores between unprompted and *close others* prompted data were more similar among people who said they used people in general (1.35) than among people who said they used close others (1.75; Table 3).

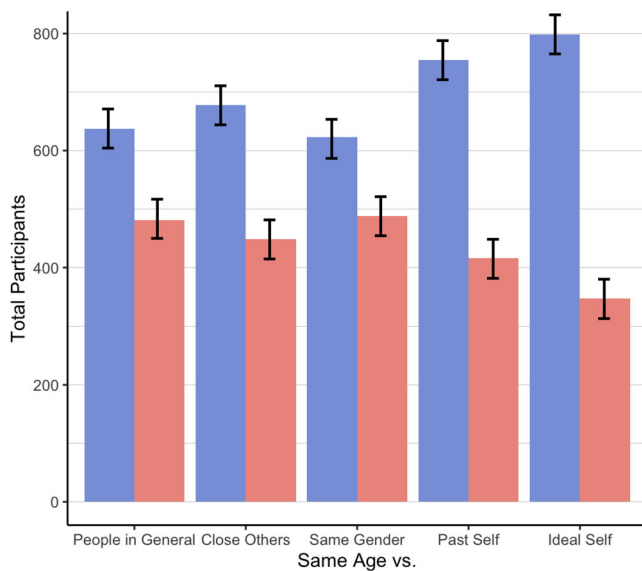


Figure 2. Number of participants whose unprompted scores were most similar to the same age prompted scores (blue/left) vs number of participants whose unprompted scores were most similar when prompted with any of the other five reference groups (red/right). Similarity was represented as the squared difference between people's average unprompted personality score (i.e., the average of all Big Five trait scores) and the average personality score within each reference group condition. No more than 60 people had the same similarity score across multiple reference groups.

Discussion

In this study, we examined which reference groups people say they use when completing a personality trait questionnaire, which reference group seems to exert the most influence on personality assessment data in general, and whether individual differences in reference group choices have specific impacts on personality assessment scores. There were three major findings. First, people are more likely to say they use certain reference groups than others, on average. Second, these stated preferences do not align with the reference groups that are actually the most impactful on trait scores. Third, people prefer between- rather than within-person comparisons. We discuss the theoretical and practical implications of these and further findings of this study below.

The most commonly reported reference group people believed they compared their personality to was *people in general*, contrary to our prediction (H1) and prior research that found close others to be favored most (Credé et al., 2010). The discrepancy between these findings could be due to two factors regarding study design. First, we asked participants to report who they compared their personality to *most* when responding to the personality items which differs from Credé et al.'s instructions that asked participants to rate the *importance* of the reference groups when arriving at their decisions for the personality items. It is possible that participants rated close others as most important in Credé et al.'s study because close others are most important emotionally to the participants. Second, Credé et al. only assessed measures of conscientiousness whereas we assessed all Big Five traits. It may be that participants use different reference groups for different traits, and close others is the most favored reference group when drawing self-perception conclusions about one's own levels of conscientiousness.

However, analyses of the similarity between prompted and unprompted data suggests that the reference group people think they are using is not all that impactful on actual results. In fact, we mostly failed to find support for our hypothesis (H2) that the most similar reference group would be the most chosen reference group (i.e., *people in general*). Instead, the most impactful reference group tended to be people the same age as the participant. This effect generalized across all traits and pertained to both overall similarity of trait scores and to rank ordering of people on trait scores. Most surprisingly, the similarity of scores between unprompted data and *same age* prompted data held even for people who believed they used different reference groups. Although differences in similarity values were often small and, in many cases, not statistically significant, an overall pattern emerged suggesting that the use of same-age peers as a reference group may have effects on personality assessment. Specifically, if people are most commonly comparing their personality to people of their same age, both longitudinal assessments of personality development as well as cross-sectional age differences in personality may be underestimated. Just as cross-cultural differences are difficult to find when everyone compares within-culture (Bleidorn et al., 2016, Heine et al., 2002), personality change may be hard to detect if everyone consistently updates their norms within age groups.

Another consistent pattern was people's tendency to compare themselves to other people rather than other aspects of themselves, even when they believed they were making self-comparisons. In other words, people make between- rather than within-person comparisons when reporting on their own personality. This suggests that personality assessment data primarily provides information on how people perceive and interpret their personality relative to others. This finding has implications for longitudinal studies of personality development, in that, for example, an observed increase in an individual's agreeableness over time is assumed to represent an increase in agreeableness relative to this individual's past levels of agreeableness, not relative to levels of agreeableness among other people. This present finding suggests that the multiple assessments of personality traits in longitudinal research may underestimate the changes in personality traits to the degree that people compare themselves to other people their age rather than to their previous self when reporting their patterns of thoughts, feelings, and behaviors. Thus, when assessing personality change, it may be advisable to explicitly ask participants to consider their previous self or use measures that are explicitly designed to assess within-person change in traits (e.g., Bleidorn et al., 2021; Hopwood et al., 2022).

Another interesting finding to emerge from this study was people's perception of change they experienced compared to their past selves. Our findings suggest that people believe they are less agreeable and open, and more extraverted and neurotic compared to what they were like in the past, with little to no change in conscientiousness. This is in stark contrast to robust personality development data, which evidences that personality development follows a maturation

trend on average, such that people tend to become *more* agreeable and conscientious, and *less* neurotic as they age, with little to no change in extraversion nor openness (Roberts et al., 2006). These differences in findings could partly be explained by expectations people have regarding a past self and how different they should be from the past. Thus, they may be more prone to respond in a way indicative of change than when relating their personality to any entity in the present (e.g., people in general, people their same age).

Overall, our findings suggest that people have relatively little insight into which reference group they are using when responding to questions about their personality. For example, trait scores from *same age* and *same gender* reference groups were two of the most similar to the unprompted scores, however, *people in general* was the most and *same gender* the least chosen reference group that participants believed using. This finding is in line with prior research (Credé et al., 2010) and is critically important when it comes to the interpretation of personality assessment data. Though our findings indicated that *same age* was most impactful on personality data, there was still variability in which reference group produced the most similar trait scores to unprompted scores. Variability in who people compare their personality to combined with their lack of awareness of this process could potentially bias personality data and complicate interpretation. Alternatively, it is possible that people are aware of who they compare their personality to, but the act of intentionally asking them to reflect and report on their personality in a survey introduces a lot of noise into their answers.

Limitations

There are two main limitations to our study. First, although we provided an adequate range of reference groups for participants to choose from, we did not provide an explicit option indicating that they do not use reference groups. This would have provided more information on the insight people had into their own comparison process. Second, we collected a convenience sample from Prolific which limits the generalizability of our findings, especially when considering the reference group effect in terms of cross-cultural comparisons. More generally, a potential limitation of this line of research is the use of practical rather than theoretical considerations in selecting and articulating reference groups in personality questionnaires.

Future directions

To date, there exists very limited research on reference group processes in personality assessment and as such, the present study provides only preliminary evidence that future research may benefit from expanding on in three ways.

First, learning what RG people believe they use may be better addressed by asking an open-ended question about RG comparisons. This could provide more insight into people's beliefs about RG choice as well as give more RG

options to list when inquiring about RG choice. Second, future research could explore people's perceptions of the RGs they use for comparisons, as these RGs may differ in size, nature, and valence. These individual differences in RG perceptions may contribute to different outcomes in personality scores even when comparisons are made using the same RG. Third, scholars should assess not just what reference groups people use, but how they use them. That is, the actual cognitive process and mechanisms of RG comparisons remain unknown. People could compare themselves to RGs relatively as Wood et al. (2012) suggests, such that their personality depends on how they rank within a group of the chosen RG, or these comparisons may happen in a more absolute sense when considering the average personality of the RG. Moreover, people may change their RG comparison depending on each personality item (e.g., perhaps people use coworkers as the RG when answering questions about their conscientiousness but use friends as the RG when answering questions about their extraversion).

Conclusions

How do people make judgments on their personality and report on it? In the present study, we investigated who individuals compare their personality to in order to arrive at their own personality response, and the extent to which they were aware of these comparisons. We found that most people believed that they compare their personality to people in general and people they know well. However, the most preferred reference group participants actually used when rating their personality was people their age, suggesting that people are not aware of the reference group comparisons they make when responding to personality items. Together with the finding that people prefer between- versus within-person comparisons when evaluating their personality, these findings suggest that we may underestimate change when assessing personality in longitudinal studies. These findings have important implications for the collection and interpretation of personality data. Perhaps most critical, our findings suggest that people may not have good insight into how they understand and rate their personality in questionnaires, supporting calls (Credé et al., 2010; Reddock et al., 2011) to provide more specific instructions about the reference group comparisons people should make when reporting on their personality.

Open Scholarship



This article has earned the Center for Open Science badges for Open Data, Open Materials and Preregistered through Open Practices Disclosure. The data and materials are openly accessible at <https://osf.io/qm5sd/>, <https://osf.io/qm5sd/> and <https://osf.io/6nav5>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by a Student Research Grant from the Society for Personality Assessment.

ORCID

Madeline R. Lenhausen  <http://orcid.org/0000-0001-5573-4157>
 Christopher J. Hopwood  <http://orcid.org/0000-0001-6645-8645>
 Wiebke Bleidorn  <http://orcid.org/0000-0003-3795-8143>

Data availability statement

For supplementary material, data that supports these findings, and its corresponding reproducible code, see <https://osf.io/qm5sd/>.

References

- Baldwin, M. W., & Holmes, J. G. (1987). Salient private audiences and awareness of the self. *Journal of Personality and Social Psychology*, 52(6), 1087–1098. <https://doi.org/10.1037/0022-3514.52.6.1087>
- Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: Accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, 21(4), 589–594. <https://doi.org/10.1177/0956797610364121>
- Blanton, H., Stapel, D. A. (2008). Unconscious and spontaneous and ... Complex: The three selves model of social comparison assimilation and contrast. *Journal of Personality and Social Psychology*, 94(6), 1018–1032. <https://doi.org/10.1037/0022-3514.94.6.1018>
- Bleidorn, W., Arslan, R. C., Denissen, J. J., Rentfrow, P. J., Gebauer, J. E., Potter, J., & Gosling, S. D. (2016). Age and gender differences in self-esteem—A cross-cultural window. *Journal of Personality and Social Psychology*, 111(3), 396–410.
- Bleidorn, W., Hopwood, C. J., Back, M. D., Denissen, J. J. A., Hennecke, M., Hill, P. L., Jokela, M., Kandler, C., Lucas, R. E., Luhmann, M., Orth, U., Roberts, B. W., Wagner, J., Wrzus, C., & Zimmermann, J. (2021). Personality trait stability and change. *Personality Science*, 2, 1–20. <https://doi.org/10.5964/ps.6009>
- Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, 17(4), 665–678. <https://doi.org/10.1037/a0028111>
- Bornstein, R. F. (2007). Toward a process-based framework for classifying personality tests: Comment on Meyer and Kurtz (2006). *Journal of Personality Assessment*, 89(2), 202–207. <https://doi.org/10.1080/00223890701518776>
- Bornstein, R. F. (2009). Heisenberg, Kandinsky, and the heteromethod convergence problem: Lessons from within and beyond psychology. *Journal of Personality Assessment*, 91(1), 1–8. <https://doi.org/10.1080/00223890802483235>
- Carvalho, L. D. F., Silveira, F., Oliveira Filho, A. Q. D., & Reis, A. M. (2020). Protocol for investigating the process underlying responses in personality assessments. *Paidéia (Ribeirão Preto)*, 30. <https://doi.org/10.1590/1982-4327e3020>
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.
- Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2), 322–331.
- Credé, M., Bashshur, M., & Niehorster, S. (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment*, 92(5), 390–399. <https://doi.org/10.1080/00223891.2010.497393>
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 192–203.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140. <https://doi.org/10.1177/001872675400700202>
- Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, 76(5), 1267–1316. <https://doi.org/10.1111/j.1467-6494.2008.00521.x>
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903–918. <https://doi.org/10.1037/0022-3514.82.6.903>
- Hopwood, C. J., Bleidorn, W., & Zimmermann, J. (2022). Assessing personality change: Introduction to the special section. *Journal of Personality Assessment*, 104(4), 431–434. <https://doi.org/10.1080/00223891.2022.2041650>
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (Vol. 3, pp. 114–158). Guilford Press.
- Latane, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Appleton-Century-Crofts.
- McCrae, R. R., Terracciano, A., Realo, A., & Allik, J. (2007). On the validity of culture-level personality and stereotype scores. *European Journal of Personality*, 21(8), 987–991. <https://doi.org/10.1002/per.659>
- Pachur, T., & Schulze, C. (2022). Heuristic social sampling. In K. Fiedler, P. Juslin, & J. Denrell (Eds.), *Sampling in judgment and decision making*. Cambridge University Press.
- Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment*, 19(2), 119–131. <https://doi.org/10.1111/j.1468-2389.2011.00540.x>
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1–25. [Database] <https://doi.org/10.1037/0033-2909.132.1.1>
- Rogers, T. B. (1974). An analysis of the stages underlying the process of responding to personality items. *Acta Psychologica*, 38(3), 205–213. [https://doi.org/10.1016/0001-6918\(74\)90034-1](https://doi.org/10.1016/0001-6918(74)90034-1)
- Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, 4(8), 605–620. <https://doi.org/10.1111/j.1751-9004.2010.00280.x>
- Weisberg, Y., DeYoung, C., & Hirsh, J. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, 2, 178. <https://doi.org/10.3389/fpsyg.2011.00178>
- Wood, A. M., Brown, G. D. A., Maltby, J., & Watkinson, P. (2012). How are personality judgments made? A cognitive model of reference group effects, personality scale responses, and behavioral reactions. *Journal of Personality*, 80(5), 1275–1311. <https://doi.org/10.1111/j.1467-6494.2012.00763.x>