

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

9-2010

Reference group effects in the measurement of personality and attitudes

Marcus Crede

University at Albany, State University of New York

Michael Ramsay BASHSHUR

Singapore Management University, mbashshur@smu.edu.sg

Sarah Niehorster

University at Albany, State University of New York

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research



Part of the [Organizational Behavior and Theory Commons](#)

Citation

Crede, Marcus; BASHSHUR, Michael Ramsay; and Niehorster, Sarah. Reference group effects in the measurement of personality and attitudes. (2010). *Journal of Personality Assessment*. 92, (5), 390-399. Available at: https://ink.library.smu.edu.sg/lkcsb_research/3136

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Reference Group Effects in the Measurement of Personality and Attitudes

MARCUS CREDÉ, MICHAEL BASHSHUR, AND SARAH NIEHORSTER

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left can be found in the text of the paper for reference. In addition, please review your paper as a whole for correctness.

- Q1:** Au: Please provide school department for author Bashshur affiliation.
- Q2:** Au: Please provide school department for author Niehorster affiliation.
- Q3:** Au: Please note that the *Publication Manual of the American Psychological Association* (6th ed.) style used for this journal does not capitalize names of theories or laws: thus, “social comparison theory” here.
- Q4:** Au: Please note that the *Publication Manual* (APA) states that “combination forms such as *he/she* [or his/her] are awkward and distracting. Consider using the plural instead”: for example, “Individuals describing **their** level . . . base **their** description partly on **their** standing” or use a more specific noun such as “test taker” here and throughout.
- Q5:** Au: Tests, scales, inventories, etc., not original to the study in question must include a citation (and corresponding reference) on first mention in text. I have added the citation and reference for the MMPI (original version). If you would like to cite and reference another source, please do so.
- Q6:** Au: Unless that was the exact choice on a demographic questionnaire, this journal prefers “White” to “Caucasian” (the word “Caucasian” indicates some geographical ancestry that is not always accurate). Later in the article you do use “White.” Please advise.
- Q7:** Au: Reworded throughout to be in the active voice, with “we” referring to you and your coauthors only. If in any instance you and your coauthor did not perform said parts of the study, please specify exactly who did.
- Q8:** Au: Please provide the degrees of freedom for the *F* statistic (i.e., fill in for the “X”).
- Q9:** Au: Please provide the degrees of freedom for the *t* statistic.
- Q10:** Au: Journal style does not capitalize names of models unless in a title.
- Q11:** Au: Please present the lower and upper numerical and verbal anchors for the 10-item scales and the 11-item scale, e.g. “a 5-item scale ranging from 0 (*Never*) to 5 (*Always*).”
- Q12:** Au: Please provide the anchors for the 40-item scale.
- Q13:** Au: Please provide the chapter page range for Hulin, 1991.

TABLE OF CONTENTS LISTING

The table of contents for the journal will list your paper exactly as it appears below:
Reference Group Effects in the Measurement of Personality and Attitudes
Marcus Credé, Michael Bashshur, and Sarah Niehorster

ARTICLES

Reference Group Effects in the Measurement of Personality and Attitudes

MARCUS CREDÉ,¹ MICHAEL BASHSHUR,² AND SARAH NIEHORSTER³

¹*Department of Psychology, State University of New York at Albany*

²*Universitat Pompeu Fabra*

³*State University of New York at Albany*

Reference-group effects (discovered in cross-cultural settings) occur when responses to self-report items are based not on respondents' absolute level of a construct but rather on their level relative to a salient comparison group. In this article, we examine the impact of reference-group effects on the assessment of self-reported personality and attitudes. Two studies illustrate that a reference-group effect can be induced by small changes to instruction sets, changes that mirror the instruction sets of commonly used measures of personality. Scales that specified different reference groups showed substantial reductions in criterion-related validities for academic performance, self-reported counterproductive behaviors, and self-reported health outcomes relative to reference-group-free versions of those scales.

Test takers' responses to self-report inventories of personality have been shown to be influenced by a wide variety of phenomena other than the individual test taker's absolute standing on the construct being assessed. These phenomena include response biases such as faking (e.g., Zickar & Robie, 1999), socially desirable responding (e.g., Detrick & Chibnall, 2008), and yea-saying and nay-saying (e.g., Knowles & Condon, 1999), as well as the order of questions (e.g., Benton & Daly, 1991; Gaskell, Wright, & Muirheartaigh, 1995; Schwarz & Hippler, 1995) and even the format and nature of the instructions and response options used for the inventory (e.g., Schwarz, 1990, 1999; Schwarz, Hippler, Deutsch, & Strack, 1985). As a result, test developers have attempted to design tests and items in a manner that reduces the impact of these artifacts (e.g., inclusion of validity scales, balancing positive and negative item content).

An additional possible influence on the manner in which individuals respond to self-report inventories of personality has been identified by researchers in the cross-cultural domain, where it is referred to as the reference-group effect (RGE; Heine, Buchtel, & Norenzayan, 2008; Heine, Lehman, Peng, & Greenholtz, 2002; McCrae, Terracciano, Realo, & Allik, 2007; Takano & Sogon, 2008). The RGE reflects the influence of comparison others on the responses to individual self-report items and has primarily been found to influence responses to measures of individualism–collectivism. In this study, we are the first to examine whether the RGE also extends to self-report measures of personality and what the impact of any RGE is likely to have on the criterion validity of scores on self-report measures of personality.

In the cross-cultural literature, it is a widely held belief that East Asians have higher levels of collectivism and lower levels of individualism than North Americans (Takano & Osaka, 1999), but studies that have measured these two constructs in East Asia and North America have frequently found this belief unsupported. Indeed, of 76 such studies reviewed by Heine et al. (2002), 34 reported data that indicated higher levels of collectivism and/or lower levels of individualism in North America than in East Asia (i.e., a pattern opposite to the expected pattern). The RGE explanation for this phenomenon is based on an extension of social comparison theory (Festinger, 1954). Festinger argued that individuals have an innate need to evaluate themselves and that this evaluation will, in the absence of objective information, be based on a comparison with others around them. These are generally spontaneous and unintentional comparison processes that can shape self-perceptions without the individual's awareness (Heine et al., 2002). This comparison is generally made with similar others, typically individuals with whom the individual has some familiarity. RGE researchers have extended the logic of the social comparison theory to argue that this comparison process can influence how respondents rate themselves on Likert-style scales. Given that the act of responding to the items of a self-report inventory requires self-evaluation on the construct being assessed, similar comparison processes may be activated while responding to an inventory of, say, collectivism (or other individual difference constructs). An individual describing his/her level of collectivism will therefore base his/her description partly on his/her standing relative to the perceived collectivism of a comparison group (e.g., people from the same national culture). Heine et al. (2002) argued that the mean of the response-option continuum on a Likert-style item represents the perceived average standing of the reference group on the constructs; and further, that it is the respondent's relative standing to this reference group that determines whether the respondent

Received August 26, 2009; Revised February 16, 2010.

Address correspondence to Marcus Credé, Department of Psychology, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222; Email: mcrede@albany.edu

80 selects a response option that is above or below this midpoint. Thus, it is thought that East Asians responding to a Likert-style inventory assessing their level of individualism will base their responses not on their absolute level of individualism but rather on their level of individualism relative to a salient reference group (i.e., other East Asians). Similarly, North Americans responding to a measure of collectivism are likely to base their responses to items on their perceived standing relative to a norm group comprised of other North Americans. Given the use of different reference groups, a direct comparison of the mean scores of North Americans and East Asians is thought to have questionable validity (Heine et al., 2002). Unfortunately, the RGE not only suggests that cross-cultural comparisons of variables assessed using Likert-type scales are likely to have low validity but also suggests that the measurement of other individual differences may be similarly affected, even if the measurement only occurs within a single culture. This is because the social comparison mechanisms that underpin the RGE are also likely to manifest themselves in other domains that require individuals to make judgments about their standing on a dimension. One such domain is the measurement of personality.

Typical self-report inventories of personality traits contain a series of descriptive phrases or statements that are thought to be reflective of the assessed personality traits. Individuals responding to these inventories are asked to indicate the degree to which these statements represent accurate descriptions of themselves via the use of Likert scales. Social comparison theory would suggest that responses to such inventories are influenced by a process in which individuals compare their own standing on the trait in question to comparison others, in a manner akin to the process influencing responses to inventories of collectivism or individualism. Indeed, it is possible that the RGE is stronger for self-report measures of constructs that fall outside of the domain of cross-cultural psychology.

Cross-cultural researchers typically assume that test takers within a culture make use of the same reference group (i.e., East Asians in general, North Americans in general), but the assumption of a universal reference group may not be valid for the measurement of other individual differences such as personality or attitudes. Indeed, it could be argued that reference groups other than the national population are more likely to be used by test takers in many research and applied settings because more specific reference groups are likely to be more salient or cognitively accessible (e.g., availability heuristic; Tversky & Kahneman, 1973) or, in social comparison theory terms, because the reference group is similar to the individual in some important way (Festinger, 1954; Goethals & Darley, 1977). For example, a job applicant may base his/her responses to a personality inventory more on his/her standing relative to the perceived standing of other applicants or job incumbents rather than his/her standing relative to the general population of his/her country. In this case, the setting in which the information is being collected (e.g., employment selection setting) suggests that an individual's responses are likely to be compared to the responses of other applicants (or of job incumbents), thereby triggering the use of individuals from that particular reference group. Another possible reason for the use of more specific reference groups may simply be that the test taker has a better understanding of his/her standing relative to the specific reference group than his/her standing relative to the population as a whole (and can therefore also access that information more readily). For

example, the responses of a high school student may be based on that student's perceived standing relative to other high school students or even his or her immediate peer group.

The use of different reference groups may, of course, not only be implicitly activated by the nature of the setting or the purpose of data collection but can also be explicitly cued by the instructions or response formats that accompany an inventory. This is especially important to consider given that prominent inventories of personality already differ with respect to whether a reference group is specified in the instruction set. For example, the instruction set of the widely used International Personality Item Pool (IPIP; Goldberg et al., 2006) specifies a reference group (scales such as these are hereafter referred to as "reference-group measures") which is comprised of people known to the respondent and of the same age and gender. The instruction set provided for IPIP scales reads as follows: "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." A similar instruction set is used by the similarly widely used scales developed by Saucier (1994). The instructions given for many other self-report measures, such as the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1943), do not specify a reference group at all (such scales are hereafter referred to as "reference-group-free measures").

Understanding the possible impact of different reference-group relevant instruction sets is important for a theoretical understanding of the processes that determine responses to self-report inventories. These effects are also important for determining whether the scores on inventories that differ in the use of specified reference groups can be meaningfully compared to each other and what reference groups (if any) should be specified to maximize criterion-related validity.

REFERENCE GROUP EFFECTS ON SCALE SCORES

If responses to psychological inventories are influenced by the mechanisms outlined by the RGE (Heine et al., 2002), then an individual test taker's scores on a measure should vary depending on the reference group being utilized. That is, test takers' scores are a function of their relative standing on the trait, and, as reference groups change, test takers' scores on a trait measure should also change. For example, test takers may indicate high levels of conscientiousness relative to their friends, low levels relative to their parents, and average levels relative to people of the same age and gender.

Hypothesis 1: Versions of a scale that specify different reference groups will produce scores that are meaningfully different from each other and different from a reference-group-free version of the same scale.

REFERENCE GROUP EFFECTS ON VALIDITY

The use of reference groups is likely to represent a significant threat to the criterion validity of inventory scores. Consider, for example, the hypothetical example in which test takers are asked to base their responses to the items of a measure of conscientiousness on their level of conscientiousness relative to their friends (the reference group). Each test taker is likely to have a group of friends that is different in terms of the distribution of conscientiousness, and test takers' responses will therefore reflect neither their absolute level of conscientiousness nor their level of conscientiousness relative to a common reference group. That is, two test takers may have identical levels of

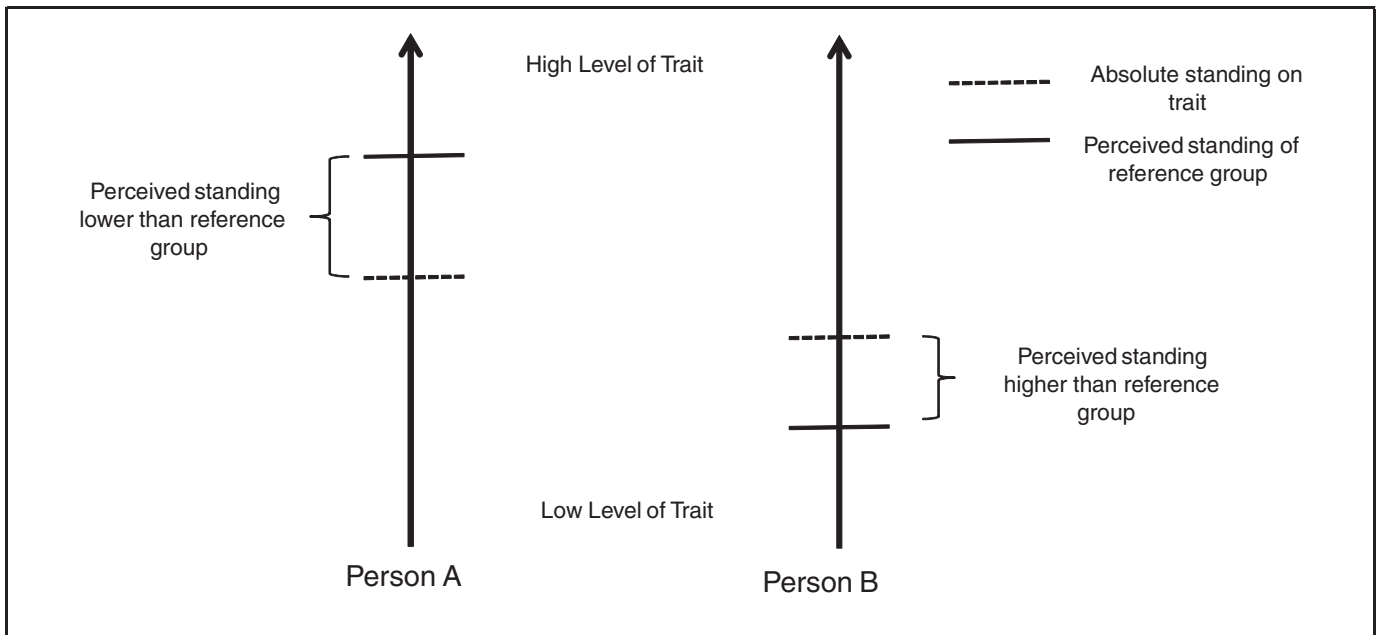


FIGURE 1.—Illustration of the effect of reference-group comparisons on personality trait scale scores.

conscientiousness but respond to the items in a fundamentally different way based on their use of different peer reference groups. Figure 1 illustrates this issue in more detail. Person A has a higher absolute level of the trait in question than Person B but uses a reference group that has even higher average levels of the trait, whereas Person B uses a reference group that has even lower average levels of the trait. The result is that Person A will arrive at a judgment that he/she has low levels of the trait relative to his/her reference group, whereas Person B will arrive at a judgment that he/she has high levels of the trait relative to his/her reference group.

Reference groups do, of course, vary in the degree to which they objectively differ across test takers and we anticipate that the threat to validity would be lower for reference groups that are so broad as to ensure that perceptions of the reference group are relatively similar across test takers (e.g., people in general) than for reference groups whose objective standing on the construct in question will vary to a greater degree from person to person (e.g., your immediate family, close friends).

Hypothesis 2a: Reference-group measures of a construct will exhibit lower correlations with relevant criteria than a reference-group-free measure of the same construct.

Hypothesis 2b: The difference in the criterion-related validity between a reference-group-free measure and a reference-group measure of a construct will increase as the utilized reference group becomes more specific to the individual test taker.

STUDY 1

In Study 1 of this article, we explored (a) whether the RGE can be activated through changes to the instructions of personality and attitudes inventories and (b) whether the RGE produces meaningful differences in criterion-related validities between measures of the same construct for criteria that have previously been shown to exhibit correlations with the personality and

attitude variables. In particular, we examined the relationship of conscientiousness with behaviors that have previously been shown to exhibit relationships with conscientiousness: procrastination (e.g., Steel, 2007) and health behaviors (e.g., Martin, Friedman, & Schwartz, 2007).

Method

Participants. The sample for this study consisted of 163 undergraduate students drawn from the psychology subject pool at a large, public Midwestern university. The sample was largely Caucasian (83%) and female (55%), with an average age of 18.6 years of age.

Measures

Conscientiousness. Conscientiousness was assessed with a 10-item measure from the International Item Pool (Goldberg et al., 2006). We administered this scale with one reference-group-free format and four additional versions that included an explicit reference group. In these explicit reference group versions, respondents were instructed to indicate their level of conscientiousness in comparison to (a) their immediate family, (b) people of the same age and gender, (c) close friends and peers, and (d) people in general. Respondents were asked to indicate their level of agreement with items using a 5-point response scale ranging from 1 (*Far less accurate for me than for my immediate family*) to 5 (*Far more accurate for me than for my immediate family*). The scales were scored so that high scores reflect high levels of conscientiousness. The order of the four reference-group conditions was varied using a Latin-squares design to counteract potential ordering effects. We always administered the reference-group-free version of the scale first to ensure that responses to this scale were not affected by a previously cued comparison with a reference group.

TABLE 1.—Correlations between measures of conscientiousness (Study 1).

No.	Measure	<i>M</i>	<i>SD</i>	1	2	3	4	5
1	Conscientiousness: No reference	3.60	.64	<i>.83</i>				
2	Conscientiousness: Family reference	2.93	.59	.44	<i>.83</i>			
3	Conscientiousness: Same age/gender reference	3.60	.62	.64	.24	<i>.86</i>		
4	Conscientiousness: Peer group reference	3.25	.48	.39	.28	.39	<i>.83</i>	
5	Conscientiousness: People in general reference	3.34	.63	.66	.46	.66	.43	<i>.86</i>

Note. *N* = 164. Internal consistency estimates (α) are presented along the diagonal in italics.

Procrastination. We assessed procrastination using the 20-item Procrastination Scale developed by Lay (1986). Respondents are asked to indicate their level of agreement with items using a 4-point response scale ranging from 1 (*Strongly Disagree*) to “4 (*Strongly Agree*). We scored the scale so that high scores reflect low levels of procrastination.

Health behaviors. Health behaviors were assessed with the 40-item Health Behavior Checklist (Vickers, Conway, & Hervig, 1990) that is comprised of four subscales: Preventative Health Behaviors, Accident Control, Traffic Risk Behaviors, and Substance Risk. Respondents are asked to indicate their level of agreement with each statement using a 5-point response format ranging from 1 (*Disagree Strongly*) to 5 (*Agree Strongly*). Subscales were scored so that high scores reflect high levels of healthy behaviors.

Subjective weighting of reference groups. At the end of the experiment, we asked participants to indicate the relative importance of each of the four reference groups in arriving at their responses to the reference-group-free condition by distributing 100 points across the four reference groups.

Results

Table 1 presents the zero order correlations, reliability estimates, and means and standard deviations of the five measures of conscientiousness. Reliability estimates are high for scores on all measures of conscientiousness, and a comparison of the reliability estimates of the five measures of conscientiousness in which we used the procedure developed by Feldt and Ankenmann (1998) found no significant differences. Scores on the reference-group-free version of conscientiousness were most strongly correlated with scores on the version making use of the

“people in general” reference group and the “same age and gender” reference group. Interestingly, the subjective weights provided by participants indicated the lowest mean weights given to these two reference groups: 18% ($SD = 11.5$) for people of same age and gender and 13% ($SD = 13.3$) for people in general. Significantly ($p < .01$) higher subjective weights were given for the family reference group and the peer group reference group with mean weights of 33.6% ($SD = 16.9$) and 35.3% ($SD = 12.9$), respectively. Importantly, none of the correlations with the reference-group-free version of the scale were large enough to suggest that all participants relied on a single reference group when arriving at responses to the items of the reference-group-free version of the scale. Further, if participants are consciously aware of the reference group that they are actually using when making comparisons in the reference-group-free version, then the relationship between the score produced by a measure of conscientiousness with that reference group and the score produced by the reference-group-free version of the scale should be moderated by the subjective weight given to the scale. We therefore used moderated regression analysis to examine whether the subjective weight given to a reference group moderated the relationship between the reference-group version of the scale and the reference-group-free version of the scale. No significant moderation effect was detected for any of the four reference groups. This indicates that participants were not consciously aware of the reference group comparisons that most strongly influenced their responses to the reference-group-free version of the scale. Alternately, it may be that participants do not use reference-group comparisons in arriving at responses to the reference-group-free version of the scale but are using some other strategy for arriving at responses to items (e.g., assessing their absolute standing on the trait by a consideration of their behaviors that does not rely on comparisons with others).

Mean differences across reference group conditions. The mean score for conscientiousness was lowest for the immediate family reference group, with the highest scores being observed for the reference-group-free condition and the for the same-age-and-gender condition. Repeated-measures analysis of variance (ANOVA) was used to examine the mean differences across all five measures of conscientiousness. In support of Hypothesis 1, we observed a significant effect, $F(X, XXX) = 59.84$, $p < .01$, and relatively large overall effect size (partial eta squared = .29), illustrating that the use of different reference groups results in meaningfully different scores on measures of conscientiousness. Focused contrasts of the reference-group-free measure with each of the four reference-group measures of conscientiousness yielded effect size estimates (Cohen’s d) of 1.09,

TABLE 2.—Correlations of scores on reference-group-free and reference-group measures of conscientiousness with five criteria (Study 1).

Criteria	Criterion Reliability (α)	No Reference Group	Family Reference	Same Age and Gender Reference	Peer Group Reference	People in General Reference
Procrastination	.88	.44 _a	.32 _{bc}	.24 _{bc}	.19 _c	.34 _b
Preventive health	.67	.09 _a	-.01 _a	.08 _a	.02 _a	.09 _a
Accident control	.63	.20 _a	-.02 _{bc}	.11 _{bc}	.02 _{bc}	.15 _b
Traffic risk	.58	-.43 _a	-.36 _a	-.44 _a	-.17 _b	-.35 _a
Substance risk	.62	-.43 _a	-.30 _{ab}	-.27 _b	.00 _c	-.34 _{ab}

Note. *N* = 164. Correlations with different subscripts across conditions signify statistically significant ($p < .05$) differences in correlation coefficients with the criterion. Calculations are based on the Meng, Rosenthal, and Rubin (1992) procedure for comparing correlated correlation coefficients.

TABLE 3.—Incremental validity of four reference-group measures of conscientiousness over the reference-group-free measure of conscientiousness (Study 1).

	Incremental Validity Over Reference-Group Free Measure of Conscientiousness							
	Immediate Family		Same Age and Gender		Friends and Peers		People in General	
	ΔR	p	ΔR	p	ΔR	p	ΔR	p
Procrastination	.02	.04	.003	.44	.000	.78	.005	.32
Preventive health behaviors	.003	.51	.001	.73	.000	.83	.002	.61
Accident control	.014	.12	.001	.68	.004	.44	.000	.78
Traffic risk behaviors	.035	.01	.045	<.01	.000	.99	.007	.25
Substance risk	.015	.09	.000	.88	.032	.01	.005	.32

Note. All ΔR values are adjusted R values.

0.00, 0.63, and 0.41 for the family, same age and gender, peer group, and people in general reference groups, respectively. A Cohen’s d of around 0.50 is often interpreted as signifying a medium effect, whereas Cohen’s d of greater than 0.80 is often interpreted as signifying a large effect (e.g., Cohen, 1988).

Table 2 presents the correlations between scores on the five measures of conscientiousness and the five criteria. In support of Hypothesis 2a, the correlations observed for the reference-group-free version of the Conscientiousness scale were highest for all five criteria and significantly different to many of the correlations observed for scores on the other versions of the scale. In support of Hypothesis 2b, the lowest criterion-related validities were observed for scores on the scale that made use of peers as a reference group. Low internal consistency estimates were found for scores on the four subscales of the Health Behavior Checklist, which is not surprising given the relatively smaller number of items in the scales and the low base rate of the phenomena assessed by individual items.

We assessed the discriminant validity of the various measures of conscientiousness using confirmatory factor analysis by comparing the chi-square values for a single-factor solution of the items assessing the two most highly correlated measures of conscientiousness (the version in which respondents were asked to “compare yourself with people of the same age and gender” and the version in which respondents were asked to “compare yourself with people in general”) with a two-factor solution in which items from the two different versions loaded onto separate factors. A two-factor solution was found to exhibit significantly better fit ($\Delta\chi^2 = 10.93$, $\Delta df = 1$, $p < .01$). The mean scores on these two measures were also significantly different, $t(XX) = 6.48$, $p < .01$.

Although scores on the reference-group-free version of the Conscientiousness scale exhibited the highest correlations with

the various criteria, it is still possible that scores on the various reference-group versions of the same scale explained variance in the criteria that was unaccounted for by the reference-group-free version. To examine this possibility, we used hierarchical regression to assess whether scores on scales with a specified reference group captured unique variance in relevant criteria above and beyond the variance accounted for by the reference-group-free measure. We also examined the reverse model (i.e., the amount of incremental variance accounted for by scores on the reference-group-free measure). Results for this analysis across the full range of predictor-criterion pairs are provided in Table 3 and Table 4. The results illustrate that, for the five criteria examined, scores on measures of conscientiousness that use reference groups provide relatively little incremental validity over scores on a reference-group-free measure of conscientiousness (see Table 3). Scores on the measure of conscientiousness that use the immediate family reference group explained the largest amount of incremental variance in the five criteria, reaching statistically significant levels ($p < .05$) for two of the five criteria. On the other hand, scores on the reference-group-free measure of conscientiousness explained much greater incremental variance in 15 of the 20 reverse models examined (see Table 4).

Discussion

Results from Study 1 show that specifying different reference groups in the instructions for a measure of conscientiousness produces significantly different scores, but also that these scores explain little variance in five relevant criteria over and above the variance explained by a reference-group-free measure of conscientiousness. These findings suggest that the instructions for self-report measures of traits should not specify a reference group given the subsequent loss of criterion-relevant information exhibited in this study.

TABLE 4.—Incremental validity of the reference-group-free measure of conscientiousness over four reference-group measures of conscientiousness (Study 1).

	Incremental Validity of Reference-Group Free Measure of Conscientiousness Over							
	Immediate Family		Same Age and Gender		Friends and Peers		People in General	
	ΔR	p	ΔR	p	ΔR	p	ΔR	p
Procrastination	.110	<.01	.137	<.01	.155	<.01	.078	<.01
Preventive health behaviors	.011	.18	.003	.49	.009	.24	.002	.59
Accident control	.055	<.01	.031	.02	.044	<.01	.019	.08
Traffic risk behaviors	.092	<.01	.037	<.01	.155	<.01	.070	<.01
Substance risk	.113	<.01	.115	<.01	.219	<.01	.077	<.01

Note. All ΔR values are adjusted R values.

STUDY 2

We conducted Study 2 to replicate the practically important findings of Study 1 and also to determine whether the results from Study 1 generalize to a wider array of personality constructs and other important criteria. In addition, we wanted to determine whether the effect of reference groups influenced measures of job attitudes that are important to the study of work. Some referent, comparator, or frame of reference is central to many theories of job attitudes (Hulin & Judge, 2003). For example, the Cornell Model of job attitudes (Hulin, 1991; Smith, Kendall, & Hulin, 1969) explicitly models the effect of broader social and economic settings as referents that determine the weights given to the cost and benefits of a given job. Similarly, Thibaut and Kelley's (1959) comparison level model involves a comparison of current outcomes to past outcomes (actually or vicariously experienced) as well as possible alternative outcomes; and Locke's (1976) value-percept model stresses the comparison of what an individual wants to what an individual has as a key influence on job attitudes. Comparisons with others are also important components of fairness judgments that appear to be strong determinants of job attitudes (Colquitt, Conlon, Wesson, Porter, & Yee, 2001).

To allow a thorough examination of whether personality and attitude measurement with and without reference groups captures unique variance in important outcomes, we expanded the criteria set to variables that have been shown to exhibit meaningful relationships with the selected personality and attitude constructs in either college (e.g., Chamorro-Premuzic & Furnham, 2003) or organizational settings (Marcus, Lee, & Ashton, 2007; Meyer, Stanley, Herscovich, & Topolnytsky, 2002). Specifically, we examined relationships among three sets of variables. The first set, self-control and integrity, have both previously been shown to be useful predictors of counterproductive behaviors (e.g., Marcus & Schuler, 2004; Ones, Viswesvaran, & Schmidt, 1993), and we therefore examined their relationship with counterproductive behaviors in a student setting. Second, achievement striving has frequently illustrated significant relationships with the grade performance of college students (e.g., Barling & Charbonneau, 1992; Lievens, DeCorte, & Schollaert, 2008), and we therefore also examined the presence of the RGE for the relationship between achievement striving and GPA. Third, commitment constructs have been shown to exhibit significant relationships with withdrawal intentions and cognitions (see Meyer et al., 2002, for a review), and we therefore also examined the RGE for the relationship between the construct of commitment to college and cognitions about withdrawal from college.

Based on the results from Study 1, we hypothesized that scores on measures of personality and attitudes that are free of a reference group will correlate more strongly with scores on the respective criteria than scores on measures of personality and attitudes that make use of a reference group (i.e., Hypothesis 2a). We also hypothesized that scores on reference-group-free measures of personality and attitudes will be significantly different from scores on a reference-group measure of the same personality or attitude construct (i.e., Hypothesis 1). In Study 2, we tested these broad hypotheses with only two reference-group relevant instruction sets: first, a reference-group-free instruction set; and second, instructions that asked respondents to indicate their standing relative to people in general. We chose these two reference group conditions because they represent perhaps the

most widely used instruction sets for self-report measures of personality. Although reference groups are typically not specified for attitude measures, we believe that responses to attitude measures may still be influenced by individuals' assessment of their standing relative to the population in general. For example, employees may decide that they have only average absolute levels of commitment to the organization but that their level of commitment is still significantly higher than the level of commitment that other people have toward their organization. That is, the response of the individual to a commitment item is likely to be influenced by both the perception of absolute standing and relative standing. Given that we examined only two reference-group conditions in Study 2, we were not able to reexamine Hypothesis 2b.

Method

Participants. Participants were 276 undergraduate students recruited from a psychology research pool at a large university in the northeastern United States. The average age of the sample was 19.0 years ($SD = 2.86$), and participants were mainly White (68.5%), female (70.7%), and freshmen (57.6%).

Measures

Commitment. We assessed commitment to college using an adapted version of Allen and Meyer's (1990) 18-item measure of affective commitment, normative commitment, and continuance commitment. We asked respondents to indicate the degree to which various statements accurately described their attitude toward the university using 5-point response options ranging from 1 (*Very Inaccurate*) to 5 (*Very Accurate*). We adapted items to assess commitment to the specific university attended by participants. Affective commitment (e.g., "I feel like part of the family at this university") denotes an emotional attachment to the university, normative commitment (e.g., "I have a duty to remain with this university") reflects attachment due to a sense of obligation, and continuance commitment refers to attachment that is the result of a perceived lack of alternatives (e.g., "If I left this university I would struggle to be accepted by a better university"). Internal consistency estimates for the reference-group-free instruction set were $\alpha = .84, .75,$ and $.68,$ respectively, for affective, normative, and continuance commitment; and $\alpha = .83, .79,$ and $.71,$ respectively, for the reference-group instruction set. We fixed the order of the two measures such that the reference-group-free version was administered first.

Personality. We measured three personality traits using scales from the IPIP (Goldberg et al., 2006). We assessed integrity and self control each with 10-item scales, and an 11-item scale was used to assess achievement striving. Internal consistency estimates for the reference-group-free instruction set were $\alpha = .73, .88,$ and $.76,$ respectively, for integrity, achievement striving, and self control; and $\alpha = .75, .86,$ and $.78,$ respectively, for the reference-group instruction set. We fixed the order of the two measures such that the reference-group-free version was always administered first.

Counterproductivity. We assessed counterproductivity using a 40-item scale developed by Hakstian, Farrell, and Tweed (2002). We asked respondents to indicate the degree to which various statements accurately described their behavior as students using 5-point response options ranging from 1

TABLE 5.—Correlations between variables (Study 2).

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1 Affective commitment	2.92	.89	.86																							
2 Normative commitment	2.73	.82	.71	.77																						
3 Continuance commitment	2.90	.76	.34	.54	.72																					
4 Integrity	3.86	.51	.14	.00	.03	.73																				
5 Achievement striving	3.86	.67	.18	.10	.04	.48	.89																			
6 Self-Control	2.84	.69	-.10	-.19	-.17	.42	.24	.78																		
7 Reference affective commitment	2.58	.75	.71	.45	.26	.14	.11	-.15	.85																	
8 Reference normative commitment	2.59	.72	.59	.58	.32	.03	.03	-.21	.80	.82																
9 Reference continuance commitment	2.59	.67	.24	.41	.64	-.02	-.06	-.23	.44	.56	.74															
10 Reference integrity	3.64	.55	.05	-.05	.03	.54	.36	.32	.02	-.03	-.13	.80														
11 Reference achievement striving	3.48	.69	.03	-.03	-.07	.36	.66	.27	.02	-.08	-.16	.52	.89													
12 Reference Self control	3.24	.64	.16	-.25	-.07	.23	.27	.56	-.19	-.24	-.20	.46	.38	.79												
13 CP cheating	2.07	.96	.06	-.02	.02	-.26	-.17	-.19	.04	-.07	.03	-.08	-.07	-.06	.85											
14 CP substance abuse	2.34	1.18	.00	.05	.02	-.39	-.34	-.56	.06	.13	.08	-.30	-.32	-.43	.40	.85										
15 CP low personal standards	3.39	.90	0.03	.03	.07	-.29	-.25	-.57	.012	.12	.11	-.14	-.22	-.30	.34	.58	.75									
16 CP property theft	1.41	.65	.09	.01	-.03	-.30	-.33	-.17	.07	.08	.08	-.22	-.28	-.12	.47	.33	.17	.70								
17 CP duplicity	2.01	.94	.00	-.04	.03	-.26	-.15	-.27	.02	.01	.09	-.18	-.15	-.18	.50	.35	.36	.27	.63							
18 CP misrepresentation	1.95	.84	.01	-.04	-.04	-.21	-.17	-.24	.02	-.05	-.04	-.06	-.07	-.15	.71	.48	.41	.39	.45	.76						
19 CP work avoidance	2.55	.94	.014	.08	-.09	-.27	-.15	-.24	.13	.13	.08	-.24	-.20	-.15	.30	.39	.45	.25	.33	.24	.65					
20 CP petty personal gain	1.74	.70	.01	-.13	-.15	-.36	-.16	-.16	.00	-.09	-.11	-.19	-.17	-.15	.54	.33	.25	.32	.47	.49	.41	.69				
21 CP indolence	2.16	.73	.05	.10	.01	-.44	-.47	-.41	.12	.12	.08	-.24	-.37	-.21	.43	.54	.53	.38	.46	.44	.50	.43				
22 Withdrawal intentions	2.58	1.27	-.51	-.40	-.32	-.22	-.10	-.07	-.43	-.46	-.26	-.19	-.07	.01	.11	.08	-.01	.07	.05	.07	.01	.017	.10	.87		
23 Current college GPA	3.08	.54	.01	-.05	-.13	-.05	.30	.10	-.15	-.11	-.18	.07	.24	.19	-.16	-.11	-.20	-.24	.00	-.10	-.04	-.06	-.26	.10	--	
24 Gender (female = 2, male = 1)	1.67	.47	-.10	-.01	.05	-.014	.07	-.13	-.06	.04	.09	.02	.05	.12	-.07	-.03	.12	-.02	-.01	-.11	.08	-.11	-.05	.08	.17	--

Note. *N* = 276. CP = counterproductivity. Internal consistency estimates (α) are presented along the diagonal.

(*Very Inaccurate*) to 5 (*Very Accurate*). The scale measures nine facets of counterproductive behavior: cheating ($\alpha = .83$), substance abuse ($\alpha = .82$), low personal standards ($\alpha = .73$), property theft ($\alpha = .72$), duplicity ($\alpha = .63$), misrepresentation ($\alpha = .72$), work avoidance ($\alpha = .63$), petty personal gain ($\alpha = .72$), and indolence ($\alpha = .55$).

Withdrawal intentions. We used three items with 5-point response options ranging from 1 (*Very Inaccurate*) to 5 (*Very Accurate*) to assess participants' intention to withdraw from college. The items were the following: I have thought about leaving (the university), I have made plans to leave (the university), and I have tried to find another university that I can attend. The internal consistency estimate for scores on these three items was $\alpha = .87$.

GPA. We gathered GPA data by asking students to self-report their GPA. Meta-analytic findings by Kuncel, Credé, and Thomas (2005) showed that self-reported GPA correlates highly with actual GPA.

Results

Table 5 presents the zero order correlations, reliability estimates, and means and standard deviations of the variables included in Study 2. The reliabilities for all measures of personality and commitment are sound ($\alpha > .70$), with only scores on some of the subscales of the counterproductivity scales exhibiting more modest internal consistency estimates, which is not surprising given the relatively small number of items in the scales and the low base rate of the phenomena assessed by individual items (e.g., cheating on tests). The correlations between scores with and without a reference group specified for the six commitment and personality scales ranged from $r = .46$ (integrity) to $r = .67$ (affective commitment). In support of Hypothesis 1, in repeated-measures ANOVAs, we found significant ($p < .01$) changes in all six of the examined variables across the reference-group and reference-group-free conditions, with effect sizes (Cohen's d) of 0.41 (affective commitment), 0.18 (normative commitment), 0.43 (continuance commitment), 0.41 (integrity), 0.53 (achievement striving), and -0.60 (self-control).

In support of Hypothesis 2a, scores on the reference-group-free predictor measures exhibited higher validities than scores on the reference-group predictor measures for 21 out of the 22 hypothesized predictor-criteria relationships. Using the Meng, Rosenthal, and Rubin (1992) procedure for comparing correlated correlations, nine of the validity coefficients for the reference-group-free measure were significantly larger ($p < .01$) than the corresponding validity coefficient for the reference-group measure. In the only instance for which the reference-group measure was more strongly related to the criteria than the reference-group-free measure (normative commitment), the difference in correlation was not significant. The largest differences between validity coefficients for reference-group measures and reference-group-free measures were observed for self-control and integrity.

Incremental validity. We used usefulness analysis to assess whether scores on scales with and without a specified reference group capture unique variance in relevant criteria. Results for this analysis across the full range of predictor-criterion pairs are provided in Table 6 and Table 7. Results indicate that, in general,

TABLE 6.—Usefulness analysis of two versions of three commitment types and achievement striving with regard to withdrawal cognition and GPA (Study 2).

Term Entered Second in Hierarchical Regression	Criteria	
	Withdrawal Cognition	College GPA
	ΔR	ΔR
Affective commitment	.106**	
Affective commitment: Reference	.004	
Normative commitment	.038**	
Normative commitment: Reference	.024**	
Continuance commitment	.02**	
Continuance commitment: Reference	.000	
Achievement striving		.015*
Achievement striving: Reference		.012

Note. All R values are adjusted values. Affective commitment = reference-group-free measure of affective commitment; affective commitment reference = measure of commitment with reference group specified in instructions.
 $p < .05$. ** $p < .01$.

scores on measures that included mention of a reference group in the instruction set provide little incremental validity over scores on the same measure assessed without mention of a reference group. Even in those rare cases in which statistically significant amounts of incremental variance are accounted for by scores on the reference-group measure, the absolute amount of incremental variance was relatively small. This effect was evident both for measures of attitudes and for measures of personality. Reverse models showed largely opposing effects. That is, scores on reference-group-free measures explained substantial amounts of incremental variance over reference-group measures of the same construct in almost all cases.

Discussion

Study 2 extends Study 1 to the measurement of both attitudes and personality traits while also examining a wider variety of criteria. The results from Study 2 are largely similar to those of Study 1 in that RGE can be produced by small changes in the instruction sets provided for inventories of personality and attitudes and that the use of reference groups in the instruction set appears to significantly reduce the criterion-related validity of scores on measures of both personality traits and attitudes.

TABLE 7.—Usefulness analysis for two versions of integrity and self-control with regard to counterproductivity (Study 2).

Criteria	Variable Entered Second in Hierarchical Regression			
	Integrity: Integrity	Integrity: Reference	Self-Control: Self-Control	Self-Control: Reference
Cheating	.072**	.005	.036*	.003
Substance abuse	.065**	.012	.147**	.019*
Low personal standards	.046**	.000	.239**	.001
Theft	.038*	.005	.016	.001
Duplicity	.03*	.002	.042**	.001
Misrepresentation	.043**	.004	.034*	.001
Work avoidance	.03*	.012	.037*	.000
Petty personal gain	.096**	.000	.010	.005
Indolence	.137**	.000	.122**	.001

Note. All ΔR values are adjusted R values. Integrity = reference-group-free measure of integrity; integrity reference group = measure of integrity with reference group specified in instructions.
 $p < .05$. ** $p < .01$.

GENERAL DISCUSSION

The results of the two studies we discussed in this article suggest that reference group judgments exert a significant influence on responses to the types of inventories commonly used to assess self-rated personality. These findings contribute further to understanding how instruction sets can influence the manner in which individuals respond to inventories. Three findings appear particularly relevant. First, an individual's score on an inventory is substantially influenced by the reference group that is specified in the instruction set; that is, different reference groups result in substantial within-person variability in scores. Second, there are substantial differences across individuals in the types of reference groups they report using when responding to self-report inventories, and there was substantial variation across participants in how important they rated each of the specified reference groups. Third, and perhaps most important, the use of reference groups resulted in significant and, at times, dramatic decrements in the criterion-related validity of scores on standard inventories of personality. Scores on measures that did not specify a reference group exhibited higher criterion-related validities than scores on measures that did specify a reference group across nearly all of the criteria examined.

We believe that there are two main reasons why scores on trait inventories that are based on reference-group comparisons are less strongly related to criteria than are scores on inventories of the same construct that are not based on reference-group judgments. First, test takers are unlikely to view the specified reference group in the same manner, in effect setting up distinct scaling standards for each individual. That is, individual A's perception of the "same age and gender" reference group is likely to be meaningfully different to individual B's perception of the "same age and gender group" reference group—even if the two individuals are, in fact, of the same age and gender. Second, there is little theoretical reason why an individual's standing on a trait relative to a reference group should be relevant to the behaviors engaged in by that individual—unless the behavior is specifically constrained by the utilized reference group (e.g., relative levels of extroversion predicting the amount of time that an individual speaks within a peer reference group). In most organizational and research settings, the absolute levels of a trait are likely to be substantially more predictive of the enactment of related behaviors.

Many researchers make use of self-report inventories that specify a reference group in the instruction set provided to research participants, the widely used IPIP scales being perhaps the most obvious example. It is also likely that similar instructions sets are utilized in selection setting (e.g., "indicate how well the statement describes the applicant relative to other applicants for this job"). This approach does not appear to be optimal in terms of obtaining scores with maximum levels of criterion validity, and we suggest that reference-group instruction sets only be included with careful consideration of the appropriateness of such comparisons to the research question or selection setting. Our findings would suggest that instruction sets without reference groups are likely to be preferable to instruction sets that refer test takers to a particular reference group. The implications of these findings are not only restricted to settings in which reference-group judgments are explicitly solicited. Findings from the cross-cultural area (e.g., Heine et al., 2008, 2002) would suggest that reference-group judgments

are used by at least some test takers even when not explicitly solicited in the instruction sets of inventories. That is, test takers are likely to naturally refer to reference groups when deciding on responses to inventory items, especially if the setting in which data is being collected or other questions that are asked activate a particular reference group. The availability heuristic (Tversky & Kahneman, 1973), for example, would suggest that data gathered in a work setting may activate a coworker reference group, data collected in a selection context may activate a fellow applicant reference group, and data collected in a school setting may activate a reference group comprised of fellow students. The recent review by Goodman and Haisley (2007) also establishes the substantial evidence that individuals in work settings make use of a variety of reference groups when evaluating themselves or their jobs. Future research may benefit by examining if the criterion-related validity of trait inventories can be further improved with an explicit request to test takers to base their responses on their perceived absolute level of the trait in question, thereby ameliorating the tendency to implicitly base their responses on a salient reference group. Future research should also examine how situational factors and individual differences influence the use of different reference groups. It is possible, for example, that extroverted individuals rely on reference group comparisons to a greater degree than introverted individuals in arriving at responses to individual items.

Our findings also suggest that scores on measures using specific reference groups (e.g., immediate family; Study 1) may provide meaningful levels of incremental validity over scores on a reference-group-free measure of the same construct. Further investigations into whether specific reference group measures may provide criterion-relevant information appear warranted, although such incremental validity is likely to be highly dependent on the nature of the criterion. We would expect incremental validity of scores affected by RGE to be largest for criteria that also reflect relative standing. For example, scores on a measure of achievement striving assessed with a peer reference group for high school students might be particularly strongly related to high school rank.

Our findings also have implications for organizational selection processes that are increasingly incorporating personality and other self-report data. The selection process in which individuals are explicitly being compared to other applicants or (in the case of promotion decisions) other incumbents seems particularly prone to RGE—even if no reference group is specified in the instruction sets. Incumbents applying for a promotion are likely to be aware of the fact that they are being compared to other employees and may therefore naturally revert to reference-group-based judgments. If RGE in selection contexts lead to a similar degradation of criterion-related validities as witnessed in this study, then organizations may benefit significantly from efforts that attempt to reduce the degree to which responses to inventories in the selection context are based on social comparison processes. Such efforts could involve simply not using instructions sets that include mention of a reference group or explicitly asking applicants to indicate their absolute level of the construct being assessed. Such relatively minor changes to the instruction format of rating forms may not only increase the criterion-related validity of self-ratings but also the validity of ratings provided by others (e.g., spouse, coworker, supervisor).

REFERENCES

- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63, 1–18.
- 725 Barling, J., & Charbonneau, D. (1992). Disentangling the relationship between the achievement striving and impatience-irritability dimensions of Type A behavior, performance, and health. *Journal of Organizational Behavior*, 13, 489–377.
- 730 Benton, J. E., & Daly, J. L. (1991). A question order effect in a local government survey. *Public Opinion Quarterly*, 55, 640–642.
- Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality*, 17, 237–250.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- 735 Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O., & Yee, N. K. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86, 425–445.
- Detrick, P., & Chibnall, J. T. (2008). Positive response distortion by police officer applicants: Association of Paulhus Deception Scales with MMPI-2 and Inward Personality Inventory Validity scales. *Assessment*, 15, 87–96.
- 740 Feldt, L. S., & Ankenmann, R. D. (1998). Appropriate sample size for comparing alpha reliabilities. *Applied Psychological Measurement*, 22, 170–178.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- 745 Gaskell, G. D., Wright, D. B., & Muirheartaigh, C. (1995). Context effects in the measurement of attitudes: A comparison of the consistency and framing explanations. *British Journal of Social Psychology*, 34, 383–393.
- Goethals, G., & Darley, J. M. (1977). Social comparison theory: An attributional approach. In J. M. Suls and R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 259–278). Washington, DC: Hemisphere Publishing Co.
- 750 Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.
- 755 Goodman, P. S., & Haisley, E. (2007). Social comparison processes in an organizational context: New directions. *Organizational Behavior and Human Decision Processes*, 102, 109–127.
- 760 Hakstian, A. R., Farrell, S., & Tweed, R. G. (2002). The assessment of counterproductive tendencies by means of the California Psychological Inventory. *International Journal of Selection and Assessment*, 10, 58–86.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.
- 765 Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science*, 19, 309–313.
- 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, 903–918.
- 770 Hulin, C. L. (1991). Adaptation, persistence, and commitment in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2). Palo Alto, CA: Consulting Psychologist.
- Q13 775 Hulin, C. L., & Judge, T. A. (2003). Job attitudes. In I. B. Weiner (Series Ed.) & W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Vol. Eds.), *Handbook of Psychology: Vol. 12. Industrial and organizational psychology* (pp. 255–276). Hoboken, NJ: Wiley.
- Knowles, E. S., & Condon, C. A. (1999). Why people say “yes”: A dual process theory of acquiescence. *Journal of Personality and Social Psychology*, 77, 379–386.
- 780 Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75, 63–82.
- Lay, C. (1986). At last, my research article on procrastination. *Journal of Research in Personality*, 20, 474–495.
- Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268–279.
- 785 Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297–1343). Chicago, IL: Rand McNally.
- 790 Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology*, 60, 1–34.
- Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, 89, 647–660.
- 795 Martin, L. R., Friedman, H. S., & Schwartz, J. E. (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology*, 26, 428–436.
- 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, 987–991.
- 800 Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172–175.
- Meyer, J., Stanley, D., Herscovich, L., & Topolnysky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior*, 61, 20–52.
- 805 Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. *Journal of Applied Psychology*, 78, 679–693.
- 810 Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506–516.
- Schwarz, N. (1990). What respondents learn from scales: Informative functions of response alternatives. *International Journal of Public Opinion Research*, 2, 274–285.
- 815 Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105.
- Schwarz, N., & Hippler, H. J. (1995). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly*, 59, 93–97.
- 820 Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and subsequent judgments. *Public Opinion Quarterly*, 49, 388–395.
- 825 Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *Measurement of satisfaction in work and retirement*. Chicago, IL: Rand-McNally.
- Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133, 65–94.
- 830 Takano, Y., & Osaka, E. (1999). An unsupported common view: Comparing Japan and the U. S. on individualism/collectivism. *Asian Journal of Social Psychology*, 2, 311–341.
- Takano, Y., & Sogon, S. (2008). Are Japanese more collectivistic than Americans? Examining conformity in in-groups and the reference-group effect. *Journal of Cross Cultural Psychology*, 39, 237–250.
- 835 Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York, NY: Wiley.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- 840 Vickers, R. R., Conway, T. L., & Hergiv, L. K. (1990). Demonstration of replicable dimensions of health behaviors. *Preventive Medicine*, 19, 377–401.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, 84, 551–563.
- 845