A CONCEPTUAL AND EMPIRICAL REEXAMINATION OF THE MEASUREMENT OF THE ...
Kuncel, Nathan R;Tellegen, Auke
Personnel Psychology; Summer 2009; 62, 2; ABI/INFORM Global
pg. 201

PERSONNEL PSYCHOLOGY
2009, 62, 201–228

A CONCEPTUAL AND EMPIRICAL REEXAMINATION OF THE MEASUREMENT OF THE SOCIAL DESIRABILITY OF ITEMS: IMPLICATIONS FOR DETECTING DESIRABLE RESPONSE STYLE AND SCALE DEVELOPMENT

NATHAN R. KUNCEL AND AUKE TELLEGEN
University of Minnesota

This paper calls into question traditional methods of measuring the social desirability of items and their use in scale construction. First, we make explicit that the proper focus for desirability studies of items and traits are the rated desirabilities of the alternative item responses indicating different trait levels. Second, the results from our first study show that the relation between degree of endorsement of an item and its judged desirability level is often nonlinear and varies across items such that no general model of item desirability can be adopted that will accurately represent the relations across all items, traits, and trait levels. In addition, the nature of these relationships can vary depending on whether desirability is considered in a work or general context. Third, results from a second study indicate specifically that people when instructed to self-present in a maximally desirable manner will choose for some attributes a moderate level of endorsement (e.g., "agree") rather than a more extreme response option (e.g., "strongly agree"). Subjects offer several different reasons for viewing the less extreme response options, which yield more moderate trait level scores, as more desirable. These reasons are linked to perceptions of the more extreme response option as being associated with negative behaviors and concerns about how others will view a more extreme response to the item. Both studies indicate that desirable responding to personality items is more complex than previously believed.

The Measurement of the Social Desirability of Items: A Conceptual and Empirical Reexamination

The nature of social desirability and its effects on personality measurement have long been a heavily studied and often controversial topic (Wiggins, 1968). Some authors have argued that social desirability does not strongly affect personality measurement (Block, 1965; McCrae &

A previous version of this study was presented at the 2001 Annual Conference of the Society for Industrial and Organizational Psychology.

Correspondence and requests for reprints should be addressed to Nathan Kuncel, 75 East River Road, University of Minnesota, Minneapolis, MN 55455; kunce001@umn.edu.

© 2009 Wiley Periodicals, Inc.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Costa, 1983) or criterion-related validity (Ones, Viswesvaran, & Reiss, 1996) whereas others have argued that social desirability is a potentially pervasive response set and that its contribution needs to be minimized or corrected for in personality measurement (e.g., Jackson, 1970, 1974, 1994). In addition, social desirability scales and their variants have been used for flagging potentially invalid personality assessments.

Social desirability continues to be an important topic because of the continued interest in overreporting of desirable characteristics or desirable response style in self-report personality tests in both personnel selection and clinical settings. In this article we define the commonly used term "faking" as deliberate overreporting of favorable characteristics and, as such, a subset of the descriptively more neutral and inclusive desirable-response style. (Ellingson, Sackett, & Hough, 1999; Mills & Kroner, 2005; Stein & Graham, 1999). Although the exact relation between response accuracy and socially desirable responding is still under debate, putative desirable response style measures continue to be used in both research and practice as either direct or proxy measures in both research and practice. In addition, social desirability assessments of items have been and continue to be used for item selection and scale development. Therefore, accurate appraisal of the percieved social desirability of items remains a critical topic.

This study is concerned with the perceived desirability of items, and challenges the traditional methods of estimating the desirability of items. We demonstrate that the failure to assess the desirability of each available response alternative prevents accurate appraisals of the relation between item responses and judged desirability. In addition, our results suggest that respondents often approach self-presentation on personality measures in complex ways that are partially inconsistent with traditional test-scoring methods. In this study, we are concerned with perceptions of social desirability, which we define next.

We define socially desirable responding as behaving in a manner that is consistent with what is perceived as desired by salient others. This group of salient others is likely to shift from situation to situation (work, school, marriage), although many behaviors are considered desirable in nearly all situations.

Some behaviors designed to make a favorable impression may have become routine and even automatic and unconscious but other behaviors may be consciously manipulative and deceptive (Zerbe & Paulhus, 1987). In either case we are dealing with conduct designed to effect some degree of deception with the goal of influencing the decisions or opinions of other people (earlier we defined faking as deliberately engaging in such behavior). The knowledge, skill, and a desire to behave in a socially desirable manner are said to be the grease that lubricates many social

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

gears (Hogan, 2005). It is not always nice to see everyone and not all presents are appreciated but it is typically best to act that way. Some people may automatically suppress a negative reaction but for others or under some circumstances this becomes a more conscious decision. Unlike deliberate grossly misleading dishonesty, more moderate levels of conscious deception and manipulation are likely commonplace and socially tolerated. An example would include the interviewee who states that his/her greatest flaw is taking on too many projects instead of saying "not getting work done." At another extreme is blatant and manipulative lying with no pretense of honesty or integrity.

The desirability of personality items has been evaluated to understand the direction responses will tend to take as a function of traits and situations. In the following sections, we discuss the two traditional methods for estimating the desirability of items, discuss their historical and current use in research and practice, call attention to the implicit assumptions these approaches make about the relation between social desirability and item responses, and outline several possible relations between social desirability and item responses.

Traditional Methods for Estimating Item Desirability

As noted by Marlowe and Crowne (1961) the social desirability of items (item desirability) has been traditionally assessed through two methods. The first method employs a group of judges who rate the desirability of an item on a quasi-continuous scale from low to high desirability or place the item into a socially desirable, neutral, or undesirable category (e.g., Edwards, 1953; Jackson, 1964; Rosen, 1956). The average of those ratings or extent of judge agreement is taken as the degree of social desirability of that item. The second method is to correlate responses to an item with a social desirability measure and use the magnitude of that correlation as an index of the desirability of that item (e.g., Hanley, 1957; Jackson, 1994).

For what follows, it is necessary to realize that, until now, desirability researchers using the rating method, though ostensibly obtaining desirability ratings of *items*, have always in effect obtained desirability ratings of item *responses*, specifically of affirmative item responses. Because of this apparent confusion of item desirability with the desirability of a particular response to those items, the desirabilities of alternative responses to the same item have never been specifically examined. Instead, certain simplifying implicit assumptions appear initially to have been made about these desirabilities and to our knowledge have never been subsequently articulated.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For example, when using the desirability rating method, researchers implicitly assume that the rating of a "true" response to a true–false item can also yield the desirability of a "false" response, namely, by simply reflecting the social desirability value for the "true" response. This assumption is unwarranted; even for true–false items the rating or sorting based on a single response option cannot be assumed to represent accurately the desirability values of any other available response options and its corresponding trait levels. The potential for drawing misleading conclusions about an item's desirability could be especially strong if there are many response options.

The distinction between the desirability of items and the desirability of item responses is most clearly visible with the identification of what were labeled "neutral" desirability items (Edwards, 1966). When authors designate an item as neutral, they make the implicit assumption that all available response options for that item are perceived as having a neutral desirability value. However, it seems intuitively unlikely that all "neutral" items would necessarily receive neutral desirability ratings across even two response levels (e.g., true/false) let alone a wider range of endorsement levels (e.g., 5-point agreement scale). For example, a number of items from the California Psychological Inventory receive neutral desirability ratings for endorsement (Gough, 1987). The items "It is difficult for me to begin a conversation with strangers" and "I sometimes pretend to know more than I really, in fact, do" are similar to those that receive neutral desirability ratings for endorsement. This is probably because they are very common and minor human failings. Yet, stating "false" to these items would not be neutral. Instead, they would indicate desirable and virtuous behavior (being unusually socially comfortable and forthright, respectively). Furthermore, if scaled on a typical 5-point Likert-type scale a rating of "strongly agree" would be quite undesirable for both items.

However, rating methods have been used to create scales of items believed to be absent of social desirability. In his classic monograph Block (1965) reports constructing an MMPI measure of Ego Resiliency based on items rated as "neutral" in desirability that was substantially correlated with Edwards's measure of social desirability. Implicit in this study is the assumption that all endorsement options are neutral with respect to desirability. If false responses are perceived as either desirable or undesirable or the underlying trait is nonlinearly related to desirability, then the demonstration based on neutral desirability items is damaged. The underlying logic of the analysis depends on this assumption.

In the correlational approach, items are correlated with a presumed measure of social desirability. The correlational model, often used to select items with a reduced desirability influence, assumes, implicitly, that there is a linear relationship between the desirability of item responses

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and scores on the social desirability measure. Furthermore, it is generally assumed that the latent variable underlying the item is linearly related to social desirability. Significant deviations from linearity at the item-response or latent-trait level are likely to result in misleading statistical results. In his treatise on scale development, Jackson (1970) specifically describes a three-stage process for eliminating item desirability from instruments. He states that the correlation between an item and a social desirability scale is an index of "the degree to which it is free from desirability variance. . ." (p. 76) and that "items improve as a inverse function of their correlation with desirability" (p. 76). This approach improves on the rating approach in that the correlation with a social desirability scale is a function of all response options. However, in the desirability literature only linear functions have been considered. Our argument is that nonlinear functions are necessary to accommodate the more complex links to be expected between alternative response options and perceived desirability. An item with a zero linear correlation with a social desirability scale could easily have response options that are perceived as very desirable or undesirable. As we demonstrate later, for some items the relation between endorsement level and desirability is decidedly nonlinear.

Item desirability methods have had an enduring presence in psychological research. They have been used for research and scale development in more recent research on topics ranging from mood to teacher behavior to cross-cultural differences in self-enhancement (Barrett, 1996; Chen, Dai, Spector, & Jex, 1997; Cuixia, Jian, & Zhongfang, 2003; Ferrando, 2008; Kearney & Plax, 1997). It is also incorporated into methods for developing forced-choice personality assessments (Jackson, Wroblewski, & Ashton, 2000).

All of this work correctly applies the methods from long-standing measurement traditions for quantifying item desirability. Unfortunately, the measurement traditions themselves have not explored the nature of the relationship between social desirability and trait level as captured by degree of endorsement. Instead, simplifying assumptions appear to have been made about the test-taking behavior of people who either intentionally or unintentionally distort their responses. For example, the assumption appears to have been that persons with a desirable response set as well as persons intentionally faking will always endorse "desirable" items (that is, items for which an affirmative response appears desirable) at the highest level of endorsement permitted by the item format, presumably on the assumption that the strongest endorsement option is also the most desirable response. This relation between trait level and social desirability is critical when social desirability is used in studies of desirable responding in the development of personality scales or, in applied settings, to evaluate the validity of self-report test results. To provide a framework for discussion

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in this area, we will consider several different possible relations between item endorsement level and social desirability level.

Modeling Item Desirability Across the Trait Continuum

If considered at all, the relation between item endorsement level and item desirability level has apparently been assumed to be linear. Stronger endorsement achieves greater desirability for desirable items and lower desirability for undesirable items. However, more complex relations are also possible. For our purposes, we distinguish four possible general functions. They are linear, non-linear monotonic, weakly non-linear monotonic, and nonmonotonic (nonlinear).

If we assume the relationship is linear, then any increase or decrease in trait level will coincide with a proportional increase or decrease in desirability. A limiting case in this category would be the unlikely ideal "neutral" item with neutral desirability ratings at each endorsement level.

If the relationship is monotonic but not linear, then any increase or decrease in trait level will also result in an increase or decrease in desirability. However, unlike the linear situation, the rate of change is not constant across the trait continuum. A familiar example is income, the marginal utility of which is generally believe to be nonlinear. That is, 40,000 per annum is far more desirable than 30,000 but 140,000 is not comparably more desirable (or useful) than 130,000.

In the case we call weakly monotonic, increases in trait level will generally be related to increases in desirability, but there are one or more regions of the function that are essentially flat. In other words, a little more or less of a trait does not invariably result in a notable change in perceived desirability, although reversals in direction of change would not occur, and for a desirable trait the overall trend would be upward. For example, we might feel that students who are above average to very high in openness are equally desirable and fun to teach.

Finally, with a nonmonotonic function the relation between the desirability and trait level will have at least one trend-reversing "bend" in the function where a positive slope turns into a negative slope or vice versa. The number of bends and overall shape could be fitted with a quadratic, cubic, or higher polynomial function. An example of this would be the case of "too much of a good thing" indicated by an upward trend reversing direction at the high extreme of a desirable trait. As we will demonstrate, talkativeness would be one example where some degree of the trait is desirable, but it is, quite simply, annoying to be very talkative.

In Study 1, we hypothesize that the relationship between trait level and desirability will be nonmonotonic or weakly monotonic for most items. The existence of nonmonotonic patterns holds up across settings (life

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in general and work). To the extent that there are differences, they will be for work-related traits (conscientiousness) such that higher levels are perceived as more desirable at work than in general. We expect linear relationships for items belonging to the two highly evaluative "lexical" traits examined in this study, Positive Valence and Negative Valence. Because these two traits deal with very powerful and very socially important characteristics (e.g., cruelty, wickedness, superiority, exceptionalness), we expect that raters will view, for example, less cruelty as increasingly more desirable. That is, more of Positive Valence characteristics is always a good thing but more of Negative Valence characteristics is always a bad thing.

Study 1: Methods

Subjects

The first sample that responded for desirability for life "in general" consisted of 63 undergraduates from a large public midwestern university. Subjects were recruited from their introductory psychology course to participate in this study in return for receiving extra credit points. Subjects were primarily Caucasian (89%) with a small majority being female (55%). The second sample of 129 undergraduates from a second large public midwestern university was recruited and rated the desirability of trait levels for work settings. These subjects were primarily Caucasian (79%) and female (63%) as well.

Measure

Data were collected on what was named the Trait Desirability Inventory (TDI). The 55 adjectives were taken from the Inventory of Personal Characteristics-7 (IPC-7; Tellegen, Grove, & Waller, 1987). The IPC-7 was developed from a factor-analyzed set of adjectival trait descriptors sampled from a dictionary. This instrument was selected for two reasons. First, in light of its lexical origin, the IPC-7 can be said to be an inventory of "folk concepts" (Tellegen, 1993). Because social desirability is based on lay evaluations of desirability, a folk concept instrument is arguably especially appropriate. Second, joint factor analytic research on the IPC-7 suggests that five of its seven dimensions represent the Big Five (Benet & Waller, 1995; Benet-Martinez & Waller, 1997; McCrae & Costa, 1995). The overlap with the Big Five increases the generalizability of this study's findings to many modern personality measures and theories. Third, in addition to the five Big-Five-like dimensions, the IPC-7 also includes two highly evaluative dimensions,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Positive Valence and Negative Valence. Jackson (1970) discussed the value of examining social desirability in scales that are highly evaluative. We agree, and given that this study was concerned with social desirability, the inclusion of these two evaluative traits seemed both interesting and appropriate.

Inclusion of all IPC-7 items would have resulted in an overwhelming rating task. With five rating items per adjectival descriptor, the resulting measure might have frustrated the subjects and diminished the quality of the data. Therefore, it was decided to reduce the item set. The IPC-7 was pruned to seven to eight items per dimension using factor analytic results. Findings from factor analyses of the original IPC-7 (Tellegen, Grove, & Waller, 1987) conducted with indigenous and cross-cultural samples (Almagor, Tellegen, & Waller, 1995; Benet & Waller, 1995; Benet-Martinez & Waller, 1997) were consulted. Those items that consistently had the strongest loadings on the IPC-7 factors, both in the U.S. samples and cross-culturally, were retained. Selecting items based on factor data is common practice in scale construction. The factor saturation of the items was not associated with the different desirability/trait functions we obtained.[1]

Under each adjectival trait descriptor, five items were listed asking the subject to rate how desirable he or she thought it was to be, respectively, extremely low (bottom 1%), below average (bottom 30%), average, above average (top 30%), and extremely high (top 1%) on the characteristic in question. Subjects provided a desirability rating for each trait phrase at each of the five trait levels on a 5-point scale with the following response options: very undesirable, undesirable, neutral, desirable, very desirable. A sample trait descriptor with its five-item list of questions is displayed in Figure 1 and the full descriptor list is shown in the Appendix.

Procedure

Subjects were instructed to rate the desirability of the personality characteristics at each of the five trait levels. They were encouraged to think of the behavior of other people who behaved in a way consistent with the particular trait and trait level, and rate its desirability. For the first sample, no specific job or other situation was specified and subjects were instructed to rate the desirability of the trait "in general." The second sample rated desirability for work. This created a between-subjects comparison of desirability for work versus in general. Finally, all subjects were encouraged

[1]Presenting full results would have been too lengthy. Full means and *SD*s are available from the first author.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Not Easily Upset

How desirable is it to be:

1. Extremely High in this characteristic (top 1%)

Very Undesirable	Undesirable	Neutral	Desirable	Very Desirable

2. Above Average in this characteristic (top 30%)

Very Undesirable	Undesirable	Neutral	Desirable	Very Desirable

3. Average in this characteristic

Very Undesirable	Undesirable	Neutral	Desirable	Very Desirable

4. Below Average in this characteristic (bottom 30%)

Very Undesirable	Undesirable	Neutral	Desirable	Very Desirable

5. Extremely Low in this characteristic (bottom 1%)

Very Undesirable	Undesirable	Neutral	Desirable	Very Desirable

Figure 1: **Sample Item Set from TDI.**

to ask for clarification of the instructions or any descriptors with which they felt unfamiliar.

Results

The desirability ratings were analyzed using a two-way repeated measures ANOVA. First- and second-order polynomial tests were also conducted. This allowed an initial test of whether data means differed from each other at all as well as whether the data fit a linear or quadratic pattern. The Huynh–Feldt p-value was also examined to test for compound symmetry. In analyzing these data, the spacing of the intervals was specified in accordance with the instructions to participants to provide desirability ratings at five points corresponding to the 1st, 30th, 50th, 80th, and 99th trait level percentile. Consistent with the design of the study, data were analyzed assuming these intervals. Across all characteristics, the null of no differences across means was rejected indicating that trait level must be considered when discussing desirability. Therefore, the remaining question was whether the data fit a linear or quadratic shape and the extent to which there were differences between the work setting and general life conditions. The polynomial tests provide additional structure and order for understanding the nature of the relationship between desirability and trait level.

The mean desirabilities of five different trait levels for each trait adjective were also plotted for both conditions to permit comparison. On the basis of the linear, monotonic, and nonmonotonic function characteristics considered earlier, the resulting curves could be loosely clustered by the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

authors into three broad categories representing three general patterns: one quasi-linear/monotonic pattern, and two (nonmonotone) inverted U-shaped patterns, one quasi-symmetric and the other asymmetric. Individual plots of each item would result in an unwieldy number of figures. Therefore, two figures representing each of the three general patterns of results are presented along with Tables 1–3, which list other items that fit the same overall pattern based on visual inspection for the in-general condition. In each table, we present the partial eta-squared (η^2_p) for the linear and quadratic tests as well as the η^2_p for the interaction between trait level and condition (i.e., between subjects differences for desirability in general or for work). These results help quantify the magnitude of the effect as well as display summary tests of significance.[2]

The first general pattern is monotonic, indicating the trait is roughly linearly related to desirability. For some traits in this category the trend appears nearly linear whereas others are weakly monotonic with one or more plateau regions. Interestingly, the general tendency for those traits with a strong linear or monotone relationship is for less of the trait to be perceived as more desirable (e.g., disgusting, cruel, horrible). The quadratic effect for these characteristics appears to be largely due to floor effects (i.e., even above average levels of cruelty are high undesirable). These characteristics also tended to have effectively identical ratings for both work and general conditions. It appears that Negative Valence characteristics tend to have less complex relationships between trait level and desirability. This may be important in the scale construction methods we discuss later. However, more complex patterns by trait level and condition were common. Figure 2 displays an example of a monotone relationship for work settings but a nonlinear relationship in general settings for "well-organized." The desirability of being more highly organized generally increases but plateaus for work settings but declines at the extreme high end in general life settings.

Table 1 lists the other traits that fit this pattern. Those functions that appear weakly monotonic often flatten at the 30% and 1% levels. The two subcategories (monotonic and weakly monotonic) are noted in the table.

The second is a category consisting of nonmonotonic curves that show a clearly inverted and quasi-symmetric U-shape, such that being average on the trait is considered most desirable and being above or below average in the trait is less desirable. Figure 3 displays an example of the inverted U-curve that is consistent across conditions for the item "talkative."

[2]Reviewers were understandably concerned with the influence selecting items based on a factor analysis might have had on the transparency of the items. We note that all of the IPC-7 items are highly transparent items (highest loading or otherwise). It is possible that different curves would be obtained with subtle items. We speculate that item complexity or ambiguity, if anything, would be associated with more radical, rather than less radical, curves.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

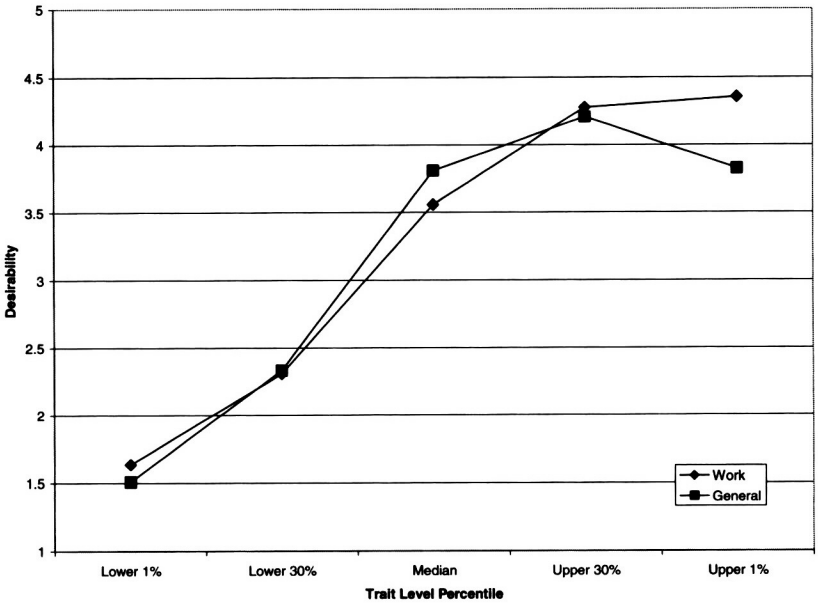

Figure 2: **Average Desirability of "Well Organized" at Five Trait Levels.**

These figures show that the most desirable level of the trait is average with decreases in desirability as a person deviates from average. Table 2 lists the other traits that fall in this general category based on visual inspection. In all cases, the quadratic contrasts fit the data best. The characteristic "strong, forceful" in Figure 4 also fits this pattern but only for the general condition. In work settings, the most desirable level of "strong, forceful" is above average with a decline for being extremely high in this characteristic.

Third, this group of asymmetric inverted U-curves indicates that more of the trait is considered desirable except at the high extreme, which is rated less desirable than the adjacent lower trait level. For example, "spontaneous, impulsive" in Figure 5 received the largest desirability rating for above average though only in the general life condition. At work, being average in spontaneity and impulsivity was rated as the most desirable. It is important to note that in all cases within this category, the quadratic contrasts were significant although sometimes obtaining smaller partial eta-squares than the linear contrasts. Table 3 lists the other traits that fit the nearly linear pattern with a drop-off at one extreme. Further research with additional power will, of course, help clarify the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 1
Characteristics That Fit the Linear, Monotonic, or Weakly Monotonic Pattern

Trait desirability inventory item	Linear η^2_p	Quadratic η^2_p	Condition η^2_p
Linear or monotonic TDI item			
Dangerous to others, harmful	.87***	.33***	.00
Others think I am quarrelsome...	.74***	.03*	.04***
Awful, terrible	.83***	.29***	.00
Treacherous, disloyal	.80***	.29***	.00
Not exceptional, not that special	.69***	.00	.01
Prompt, punctual, get things done...	.75***	.21***	.00
Often irritated by minor setbacks	.70***	.02	.01
Mentally disturbed, sick	.76***	.33***	.00
Wicked, evil	.83***	.26***	.00
Disgusting, horrible	.80***	.31***	.00
Deserve to be hated	.83***	.33***	.01
Cruel, mean	.84***	.14***	.00
Weakly monotonic TDI item			
Prefer to be a alone, a loner	.54***	.10***	.01
Reserved, distant	.58***	.13***	.01
Odd, peculiar	.37***	.14***	.00
Stubborn, obstinate	.57***	.16***	.00
Often jump and jittery	.68***	.09***	.12**
Get into arguments, argumentative	.70***	.06***	.01
Do not talk much uncommunicative	.48***	.17***	.00
Do not let things bother...me	.64***	.38***	.01
Strange	.50***	.10***	.00
Nervous, high-strung	.80***	.06***	.01

Note. η^2_p - = partial eta-squared, *p < .05, **p < .01, p < .001.

relationships between desirability and trait level for different personality items.

Study 2

The results from Study 1 suggested that higher levels of a generally desirable trait are not always perceived as more desirable. Study 2 was conducted to shed some light on reasons for such perceptions and provide an initial insight into the subject's response process. We believed it would be informative to place respondents in a hypothetical employment test-taking situation in which an extreme (low or high) level of endorsement would probably be a more advantageous option according to test scoring keys for relatively transparent measures of favorable personality characteristics than a less extreme endorsement level in the same direction. Should participants under these circumstances elect not to pick the more

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

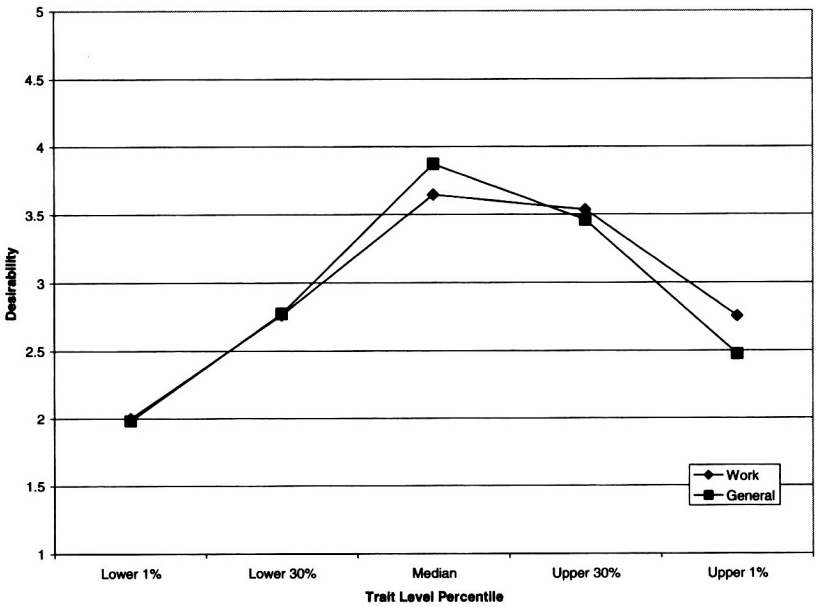

Figure 3: **Average Desirability of "Talkative" at Five Trait Levels.**

TABLE 2
Characteristics That Fit the Quasi-Symmetric Invented U-Curve Pattern

Trait desirability inventory item	Linear η^2_p	Quadratic η^2_p	Condition η^2_p
Unusual, unconventional	.07***	.27***	.01
Politically radical, hold. . .views	.28***	.22***	.01
An ordinary, everyday person	.02	.16***	.00
Easy on others, lenient	.09***	.50***	.04***
Conservative	.02	.36***	.00
Cautious, circumspect	.35***	.49***	.01
Talkative	.17***	.61***	.01

Note. η^2_p - = partial eta-squared, *p < .05, **p < .01, p < .001.

extreme endorsement level, we could ask them to explain their behavior. The goal was simply to collect information about why some more extreme responses are perceived as less desirable despite the fact that they could be expected to lead to a higher score. To accomplish this, we chose a directed faking design for two reasons. First, through this manipulation participants would be encouraged to select strong endorsements of positive items. Second, we would simultaneously gain additional information

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

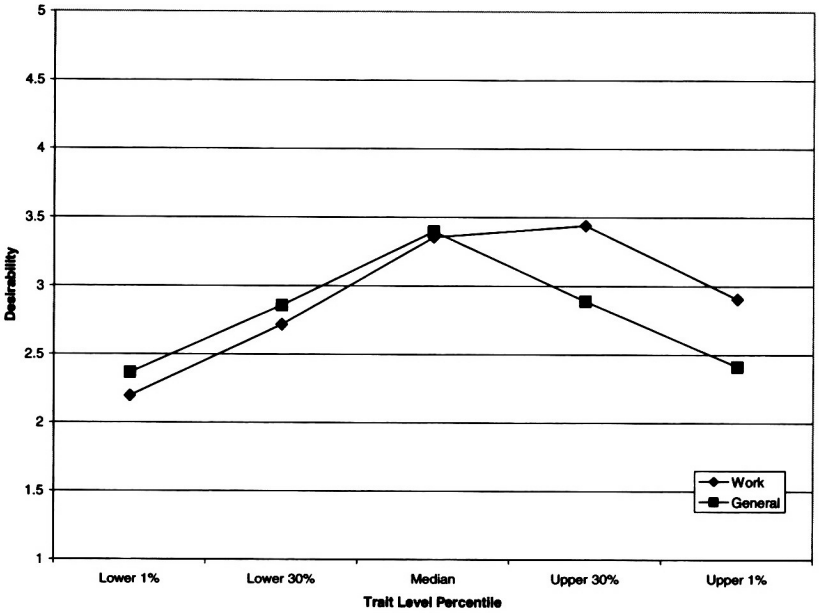

Figure 4: **Average Desirability of "Strong, Forceful" at Five Trait Levels.**

about the link between item desirability and intentional faking behavior, potentially supporting and expanding on Study 1.

Methods

Subjects

The sample consisted of 138 undergraduates from a large public midwestern university. This sample is independent from Samples 1 and 2 in Study 1. Participants were recruited from their introductory psychology course and were compensated for their participation with extra credit points. Participants were primarily Caucasian (87%) with a small majority being female (51%).

Procedure

Subjects were instructed to fake good in the context of applying for a desired job with the goal of receiving the best possible score even if it involved deception. Subjects were asked to fill out the IPC-7 and the Big-Five markers from Goldberg (1990), both of which have multiple response options. After completion of the personality measures, participants were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

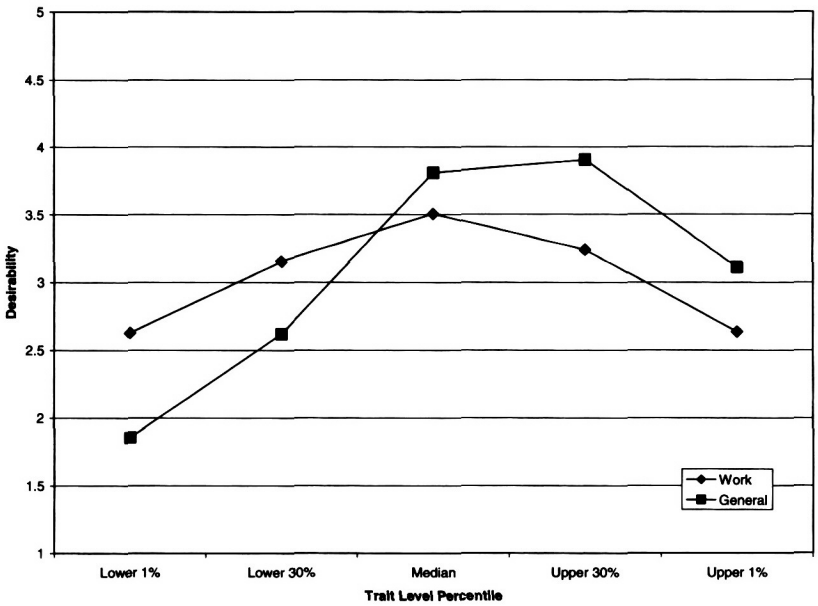

Figure 5: **Average Desirability of "Spontaneous, Impulsive" at Five Trait Levels.**

asked to identify one or more items in which they had not used the strongest (most extreme) response (e.g., endorsing "agree" rather than "strongly agree") and indicate in writing why they had elected to select a different response option. Subject responses were transcribed into a database and turned into sorting cards.

The two authors did an initial content sorting of the responses, which resulted in the nine categories used in this study. However, some of the responses provided more than one explanation, sometimes in reference to multiple items. A coding form was made that included all nine response categories, which were not treated as mutually exclusive, allowing the two authors to place the response of each participant, depending on its content, into one or more than one category. The initial agreement between the two coders was 94%. Disagreements, which mostly reflected the extent to which a response contained multiple explanations, were resolved through discussion.

Results

The percentages of responses sorted into each category out of the total number of responses are presented in Table 4. Overall 61% of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 3
Characteristics That Fit the Asymmetric Inverted U-Pattern

Trait desirability inventory item	Linear η^2_p	Quadratic η^2_p	Condition η^2_p
Feelings are easily hurt	.65***	.11***	.01
Keep belongings neat and tidy	.56***	.43***	.02
Lively, animated	.60***	.52***	.03*
Hold traditional values and beliefs	.33***	.41***	.02*
Quiet	.42***	.26***	.01
Like to be with people, sociable	.63***	.55***	.03**
Spontaneous, impulsive	.16***	.53***	.11***
Excellent, first rate	.68***	.37***	.01
Impressive, remarkable	.65***	.37***	.00
Strong, forceful	.05*	.42***	.03*
Believe in strict discipline at home	.12***	.37***	.05***
Deserve to be admired	.15***	.43***	.06***
Do not worry about the little things	.38***	.45***	.04***
Do things in an orderly manner	.64***	.47***	.03**
Tough, uncompromising	.29***	.23***	.01
Conventional	.21***	.30***	.01
Exceptional, special	.78***	.32***	.01
Playful	.20***	.47***	.12***
Can put worries out of mind	.49***	.38***	.01
Not easily upset	.74***	.45***	.04***
Like to have a place for everything	.30***	.44***	.00
Outstanding, superior	.60***	.41***	.02*
Like to improvise. . .play by ear	.07***	.33***	.04***
Dislike arguments and conflict	.38***	.47***	.00
Well-organized	.74***	.46***	.03**
Headstrong, willful	.39***	.40***	.00

Note. η^2_p - = partial eta-squared, *p < .05, **p < .01, p < .001.

responses indicated that the most extreme option was undesirable for a reason related to interactions or relationships with others. The next most common comment (25%) was that the response option was sufficiently discrepant from the subject's self-perceptions to make that option unacceptable (more on this category later). The overall category can be broken down into its components. The most common response (21% of all responses) was that the more extreme response was undesirable but not specific (e.g., "It would look bad."). Comments in this category clearly indicated that the more extreme response was undesirable but provided no specific information as to why the extreme response is undesirable.

The next most frequent response (14%) was that behavior associated with the more extreme response would be maladaptive. For example, a number of subjects indicated that high levels of tidiness would be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 4
Free Response Sort Results for Participants Providing Explanations for Not Using the Most Extreme Endorsement Option Under Faking Conditions

Percent	Explanation
21	Extreme response is undesirable (unspecified)
14	Behavior associated with extreme response would be maladaptive
10	Extreme response is improbable or not believable
8	Extreme response is too much of a good thing
8	Extreme response would be bragging or arrogant
7	Made exclusive use of extreme response
4	Characteristic was neutral in content
3	Nonresponse/illegible response
25	Extreme response would be too inaccurate

counterproductive and a waste of time. Next were responses (10%) that indicated that the response would be in some way improbable or not credible. In other words, the strongest endorsement would not be believed by others. Some comments (8%) indicated directly that the extreme response reflected too much of a good thing. This category is similar to the maladaptive category except that subjects made explicit that above-average levels of the trait are desirable but at the highest levels are somewhat less desirable. A like number (8%) of comments indicated that that they felt that endorsing a very high trait level would be perceived as bragging or arrogant. A few responders nonetheless made exclusive use of extreme responses (7%). Two small groups indicated that they considered the attribute in question to be generally neutral in desirability (4%) or provided either no response or else one that was unreadable (3%).

In sum, whereas those personality tests that measure attributes that are considered desirable are typically scored such that stronger levels of endorsement of the items (as keyed) yield higher scale scores, 61% of the comments indicated that for various reasons an extreme endorsement level is not as desirable as a more moderate level of endorsement. When more specific reasons were supplied, subjects indicated that the more extreme response would generally indicate maladaptive behavior, be perceived as arrogant by others, or be perceived as improbable by others leading to an unfavorable evaluation of the item response.

Discussion

Although based on a very different methodology, we believe the results from the second study support and expand on the results from Study 1. Participants quite consistently reported that for certain items, moderate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

endorsements are the most desirable, and more extreme endorsements, for a variety of reasons, are less desirable. Furthermore, the free response data on these same items often explicitly indicated that participants viewed the relation between item endorsement level and trait level as nonlinear, with an extreme endorsement level seen as less desirable than a more moderate endorsement level. This is most evident in those comments indicating that a stronger response than the one endorsed would be "too much of a good thing" or that the behavior associated with the extreme response would be maladaptive because it was excessive. These data provide several causal explanations for why subjects consistently fail to attain maximal scores on personality measures even when provided a financial incentive to maximally distort responses (e.g., Schmit, Ryan, Stierwalt, & Powell, 1995).

What we did not expect was the relatively large number of responses indicating that the more extreme response levels simply did not reflect the individual's actual personality characteristics, in view of actual trait level, variability over time, or situational influences. Subjects consistently indicated that they understood the faking task but felt that the response was too extreme given their actual personality. It appears that they were unwilling to present themselves in a manner too inconsistent with their actual conduct. Additional research is needed to better understand the cause of this phenomenon. One possible explanation is that participants may have felt that to claim an extremely high level of a favorable trait would cause problems in the workplace because they could not sustain that level of behavior and would ultimately be found out. That is, the gap between their actual self and the self presented on the test would suggest dishonesty or naively uncritical self-appraisal to others. Another approach would be to consider dissonance theory (Festinger, 1957) and its variants, including self-consistency theory (Aronson, 1999) in which dissonance is caused by inconsistent thoughts or actions that are not consistent with one's self-concept. Here participants may be unwilling to select the most extreme response because it would cause dissonance.[3]

Also of interest were the comments of three participants indicating an approach that took the total test battery into account. These participants noted that they had endorsed some extreme responses on one or more other like items and felt that to retain credibility they needed to respond more conservatively on this particular item. It appears then that at least some test takers attempted to implement a total test strategy of desirable responding. This approach is simultaneously highly sophisticated and uninformed. It is sophisticated in that test takers in completing individual

[3]We thank one reviewer for suggesting this set of theoretical explanations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

items take into account their responses related to other items in the test, in an effort to optimize the impact of their overall test response. It is uninformed about (and arguably too sophisticated for) common methods of scoring and interpreting the relatively transparent additive measures of favorable personality attributes that are often used in personnel selection settings. Similar strategies have been reported in other verbal protocol studies (Kim, 2006). Under these circumstances, failure to make excessive claims of positive characteristics would actually reduce the likelihood of achieving the desired outcome (i.e., a job offer).

Overall Discussion

We suggest this study makes two contributions. First, it makes explicit that the desirability of alternative responses and corresponding trait levels are the proper focus of desirability studies of items and traits. Consequently, the desirability of all available response options needs to be determined. The association between social desirability and responding is more complex than traditional item scoring methods have allowed for. Second, our empirical findings illustrate the different kinds of response options/desirability functions that can emerge.

The results suggest that the relation between social desirability and traits is more complex than has been typically contemplated. For example, the present results indicate that we cannot assume that for a generally adaptive and desirable trait an increase in trait level means an increase in perceived desirability. The relation between the two depends on the trait in question, the item response, and the context. It also appears that some individuals, when asked to respond in a socially desirable manner, consider the desirability implications of their responses to individual items in the context of their other item responses.

Assessments of item desirability based on simple ratings of the desirability of the content of an item, which are implicit ratings of an affirmative response to that item, are not likely to provide accurate assessments of the differing specific desirabilities of alternative responses. Measurements of item desirability should assess the desirability of all response options because desirability is jointly defined by item content and each specific response option. The overall desirability of an item could be conceptualized as the average desirability across all trait levels as estimated by evaluations of a range of response levels, but the meaning and usefulness of such a general item parameter remains to be determined and has never, to our knowledge, been estimated. More relevant may be the maximum desirability difference between any two available item response options (i.e., the range rather than the central tendency of response desirabilities).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Although the causes of the various functions have not been fully explored, several alternative explanations can be inferred from the free-response data. For those items for which there was a generally positive relation between endorsement strength and degree of desirability, and a downward trend only at the high extreme, very high levels of the characteristic in question may suggest excessiveness to many participants. Excessive talkativeness, neatness, or cautiousness may be perceived as interfering with effective overall functioning. On the other hand, it does not appear that this excessiveness was perceived as actually maladjusted. Extreme levels of many characteristics were still rated as desirable if less so than a lower but still above average level. Raters may feel that extreme levels of some traits suggest only an element of maladaptive behavioral rigidity in the expression of an otherwise adaptive trait. Note that even some very desirable characteristics from the Positive Valence scale (e.g., excellent, first-rate) received lower desirability ratings at very high levels than at the lower but still well above average level. It is possible that the attribution to oneself of extreme levels of Positive Valence characteristics is assumed to elicit negative reactions from others (e.g., disbelief, ridicule, resentment).

On the other hand, Negative Valence items (e.g., treacherous, disloyal) received nearly linear or monotonic ratings across the trait continuum (though often with what appear to be floor effects). These functions do not contradict explanations stressing the avoidance of unwanted reactions from others, as it is not unreasonable to assume that being extremely low on these traits is not going to be met with negative reactions from others. That is, people generally do not resent, and are quite happy to be around, others who are not remotely treacherous.

These results also suggest that deliberate misrepresentation or faking on tests may be more complicated than previously thought. For measures of personality traits judged "desirable" it is often assumed that higher scores are better. Validation research findings have been interpreted as supporting this assumption. Based on this way of thinking, there is a tendency to assume that individuals who fake good will endorse the most extreme response possible, as higher scores are "better." Our results indicate that this may not be true for most people. If a test taker's goal is to present herself in the best possible light, these results suggest that people will not necessarily select the most extreme (low or high) response option possible. The most extreme response may not be the one perceived as most desirable. Test takers may attempt to achieve their goals in ways that are logically inconsistent with commonly used scoring schemes. Kim (2006) provides a taxonomy of strategies and initial study that can be used as a basis for understanding the disconnect between respondent behavior and traditional scale scoring methods. An exception observed in this study

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

seems to be for the items that received fairly linear desirability ratings across the trait continuum. These characteristics tended to be traits where possessing any amount of the trait was seen as undesirable. A desirable response approach to these items is likely to follow a simpler pattern.

Note that these results do not argue either for or against the idea that predictive validity is reduced when test takers intentionally distort their responses. Nor do they speak to the amount of distortion that occurs in personnel selection or clinical settings. Additional research is needed to determine the implications of our findings for test validity. Finally, the nonlinear relationships suggest that raters believe that extreme levels of many traits may be associated with less effective behaviors and, therefore, poorer performance. Although some research in the work domain does not support this inference (Robie & Ryan, 1999), others have suggested that there is some degree of nonlinearity in personality performance relations at work (LaHuis, Martin, & Avis, 2005) and at school (Cucina & Vasilopoulos, 2005). What is most important for responding behavior is that subjects appeared to think this might be true. The results do suggest that the self-presentation motives and behaviors of test takers may not lead to straightforward and easy predictions of specific item responses. The complexity of desirable responding may account for the inconsistent findings on the use of social desirability or unlikely virtues scales as corrections to personality tests (Hough, 1998). That is, desirability scales may simply be too crude to be effective.

An additional layer of complexity may be introduced when differences between jobs or job categories enter the picture. It is not unreasonable to speculate that different jobs would tend to produce different curves. For example, high levels of dominance may be perceived as more desirable for management positions. Low levels of talkativeness may be perceived as particularly undesirable for sales positions. In addition, new workers may have different patterns (due to a naive understanding of the job) than seasoned incumbents.

Limitations

Some important limitations of the current studies warrant discussion and suggest additional directions for future research. A common criticism of using college student samples would be that we might obtain different functions from a noncollege student sample. We think it is unlikely that different members of the working population will have dramatically different perceptions of the desirability functions underlying these characteristics for life in general. However, work experience in a specific job may have an influence. A fruitful direction would be to examine the extent to which our findings generalize across other situations. In other words,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

we examine the desirability of trait levels in general and for work settings, but other more specific situations are certainly possible. Focusing the situation on work increased agreement as indexed by rating variability. Variability was consistently higher for the life in general condition (SD = .94) versus ratings at work (SD = .85). Therefore, a more narrowly delineated work situation (say, a sales job or for a middle manager position) may yield stronger agreement particularly given evidence that people can generate stereotyped patterns for specific jobs in a lab (Furnham, 1990). Different patterns for different jobs could have useful applications we discuss shortly.

This study also focused on desirability functions for adjectives and short phrases. Although it seems reasonable to believe that these findings will generalize to more traditional personality items this question will need to be specifically examined. Finally, future research will need to examine alternative trait level anchors and response options than the sets examined here. For example, our labeling of the high ends of the trait levels "extreme" may have influenced the nonlinear relationships. However, two observations temper this interpretation. First, the free response data indicate that a "strongly agree" response option can also be perceived as reflecting a potentially maladaptive level of a trait. This suggests our anchors did not create a spurious effect. Second, we obtained a number of nonlinear relationships within the "above average" to "below average" range for some traits (e.g., talkative). That is, the effect was not limited to the extreme anchors. Nonetheless, anchor labels may exaggerate the shape of the curves and additional work is needed to corroborate the effects demonstrated here.

These findings are consistent with the concept of ideal point items in which the probability of endorsing an item does not necessarily increase with theta (the individual's underlying trait level; Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001). Our findings suggest multiple potential causes of the psychological process underlying ideal point judgments. The combination of findings in this domain with our own findings suggests the possibility that individuals with very high theta may actually view extreme responses as desirable. This would add further complexity to how the perceived desirability of the item endorsement options is affected by both individual difference and situational effects. Finally, given individual differences in self-presentation style (e.g., McFarland & Ryan, 2000), such individual differences may moderate the use of specific self-presentation strategies or item responding patterns.

The present findings may provide some useful leads for scale development and the detection and control of tendencies to overreport desirable characteristics. One could create new measures of those tendencies based on the desirability curves. A probabilistic approach could be adopted such

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that different desirability weights are given to each response option depending on the desirability function. A person who is overreporting may consistently give the most desirable response, which may not be the most extreme rating. In other words, test takers who consistently fail to endorse the most extreme response option may be attempting to represent themselves in a desirable manner. The probability of such a set of responses would be assessed with a normative sample under conditions of low incentive for such representations. To implement this idea properly, it may be necessary to increase the number of possible item endorsement levels. It may also be necessary to include items that display an inverted-U relation between trait level and desirability even if these items are not central to the primary focus of the measure. In contrast, the items that displayed nearly linear relationships would probably not be optimal for detecting desirable-response distortion.

Although it is likely that an individual who is overreporting desirable characteristics is more likely to give a more extreme response, it will also be the case that employers may be most interested in hiring people with extreme responses on those items (e.g., not at all disloyal or argumentative). This raises the interesting question of the perceived desirability of extreme levels of these characteristics versus the actual adaptiveness of extreme levels of these characteristics. Being not at all argumentative is close to being a push over. Similarly, absolute loyalty may be something desired by organizations but it may not be adaptive for the person in all cases. Finally, a more complicated direction would be to consider whole test behavior. Some subjects reported adjusting responding to control appearing boastful or not credible. Inconsistent responses on parallel items may be indicative of intentional overreporting of desirable characteristics.

One additional implication for applied practice and future research that we find particularly intriguing is that the endorsement-desirability function could differ across both jobs and experience levels, which could have implications for evaluating fit. Misperceptions of optimal trait levels may reflect a naïve understanding of the job (e.g., corporate accounting is for the quiet and cautious). Associations with turnover, performance, and job satisfaction could be very useful. Therefore, we speculate that actually having applicants rate the desirability of personality trait levels in addition to their own trait level could prove useful.

Second, efforts to control overreporting of desirable characteristics, including faking, through forced choice instruments may need to take the current findings into account. If, as our findings suggest, perceptions of item desirability are not well captured by simple desirability ratings of an endorsement, the appropriate matching of items for a forced choice format may be more complex than currently assumed. At a minimum we speculate that item desirability would be the desirability of making one

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

choice versus the other. For example, desirability would not be controlled if items with similar desirability ratings are matched. Instead, desirability could only begin to be controlled for an item if the choice of selecting one stimulus is perceived as equally desirable to selecting the other. If paired comparisons are made across different traits, it also seems likely, given the current findings, that the desirability of different response patterns across traits might need to be considered. Emphasizing one trait over another (dutifulness at work, kindness as a parent) could be seen as a desirable strategy.

A final direction would be to attempt to identify items for which there is disagreement as to what is the most desirable response option. In the context of assessing psychopathology, Morey (1997) noted that at a trait level "there is likely to be little consensus among people as to whether being talkative is a desirable or undesirable characteristic..." (p. 939). Interestingly, our data suggest that, with respect to talkativeness, there is some consensus at least for specific adjective markers but that the desirability is not at the extremes. However, the pattern is not so clear for other characteristics such as "unusual" for which being average was, on average, the highest rated trait level. However, the high and low extremes also received strong endorsement from some raters. If there is genuine disagreement between respondents about the optimal response for multiple items or even multiple traits, responses to some items may show idiosyncratic patterns when test takers are motivated to overreport desirable characteristics. This makes it possible to detect desirable response patterns, such as faking by identifying these items and scoring them for idiosyncratic responses as demonstrated by Kuncel and Borneman (2007).

In conclusion, we have argued that the "desirability of an item" must be explicated and evaluated in terms of the desirability of each alternative response to that item. Taking this approach, we have demonstrated that endorsement levels and corresponding trait levels have complicated relations to social desirability. No linear or monotonic function will accurately represent these relations across all traits and trait indicators (items). Endorsement levels and corresponding trait levels have complicated relationships with social desirability both in general and for work settings. To capture these relations the "desirability of an item" is best explicated and measured as the desirability of each alternative response to that item.

REFERENCES

Almagor M, Tellegen A, Waller NG. (1995). The big seven model: A cross-cultural replication and further exploration of the basic dimensions of natural language trait descriptors. *Journal of Personality & Social Psychology, 69*, 300–307.

Aronson E. (1999). Dissonance, hypocrisy, and the self-concept. In Harmon-Jones E, Mills J (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology*. Washington, DC: American Psychological Association.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Barrett LF. (1996). Hedonic tone, perceived arousal, and item desirability: Three components of self-reported mood. *Cognition and Emotion, 10*, 47–68.

Benet V, Waller NG. (1995). The big seven factor model of personality description: Evidence for its cross-cultural generality in a Spanish sample. *Journal of Personality & Social Psychology, 69*, 701–718.

Benet-Martinez V, Waller NG. (1997). Further evidence for the cross-cultural generality of the big seven factor model: Indigenous and imported Spanish personality constructs. *Journal of Personality, 65*, 567–598.

Block J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century-Crofts.

Chen PY, Dai T, Spector P, Jex S. (1997). Relation between negative affectivity and positive affectivity: Effects of judged desirability of scale items and respondents' social desirability. *Journal of Personality Assessment, 69*, 183–198.

Chernyshenko OS, Stark S, Chan, K-Y, Drasgow F, Williams B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*, 523–562.

Cucina JM, Vasilopoulos NL. (2005). Nonlinear personality-performance relationships and the spurious moderating effects of traitedness. *Journal of Personality, 73*, 227–260.

Cuixia L, Jian, X, Zhongfang, Y. (2003). A compromise between self-enhancement and honesty: Chinese self-evaluations on social desirability scales. *Psychological Reports, 92*, 291–298.

Edwards AL. (1953). The relationship between judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*, 90–93.

Edwards AL. (1959). Social desirability and personality test construction. In Bass BM, Berg IA (Eds.), *Objective approaches to personality*. Princeton, NJ: Van Nostrand.

Edwards AL. (1966). Relationship between probability of endorsement and social desirability scale value for a set of 2,824 personality statements. *Journal of Applied Psychology, 50*, 238–239.

Ellingson JE, Sackett PR, Hough LM. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155–166.

Ferrando PJ. (2008). The impact of social desirability on the EPQ-R item scores: An item response theory analysis. *Personality and Individual Differences, 44*, 1784–1794.

Festinger L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Furnham A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology, 9*, 46–55.

Goldberg LR. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*, 1216–1229.

Gough H. (1987). *California psychological inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Hanley C. (1957). Deriving a measure of test-taking defensiveness. *Journal of Consulting Psychology, 21*, 391–397.

Hogan R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance, 18*, 331–341.

Hough LM. (1998). The millennium for personality psychology: New horizons or good old daze. *Applied Psychology: An International Review, 47*, 233–261.

Jackson DN. (1964). Desirability judgments as a method of personality assessment. *Educational and Psychological Measurement, 24*, 223–238.

Jackson DN. (1970). Sequential system for personality scale development. In Spielberger CD (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61–96), New York: Academic Press.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Jackson DN. (1974). *Personality research form manual.* Goshen, NY: Research Psychologists Press.

Jackson DN. (1994). *Jackson personality inventory–revised manual.* London, ON: Sigma Assessment Systems.

Jackson DN, Wroblewski VR, Ashton MC. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13,* 371–388.

Kearney D, Plax TG. (1997). Item desirability and the BAT checklist: A reply to Waltman and Burleson. *Communication Education, 46,* 95–99.

Kim BH. (2006). *Building a theory of strategic deception.* Unpublished doctoral dissertation, East Lansing, Michigan State University.

Kuncel NR, Borneman MJ. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment, 15,* 220–231.

LaHuis DM, Martin NR, Avis JM. (2005). Investigating nonlinear conscientiousness-job performance relations for clerical employees. *Human Performance, 18,* 199–212.

Marlowe D, Crowne DP. (1961). Social desirability and response to perceived situational demands. *Journal of Consulting Psychology, 25,* 109–115.

McCrae RR, Costa PT. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

McCrae RR, Costa PT. (1995). Positive and negative valence within the five factor model. *Journal of Research in Personality, 29,* 443–460.

McFarland LA, Ryan AM. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85,* 812–821.

Mills JF, Kroner DG. (2005). An investigation into the relationship between socially desirable respondings and offender self-report. *Psychological Services, 2,* 70–80.

Morey, LC. (1997). Personality diagnosis and personality disorders. In R. Hogan, JA. Johnson, SR. Briggs, Handbook of Personality Psychology. (pp. 919–643). San Diego, CA: Academic Press

Ones DS, Viswesvaran C, Reiss AD. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660–679.

Pervin LA, Lilly RS. (1967). Social desirability and self-ideal self on the semantic differential. *Educational and Psychological Measurement, 27,* 845–853.

Robie C, Ryan AM. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment, 7,* 157–169.

Rosen E. (1956). Self-appraisal, personal desirability, and perceived social desirability of personality traits. *Journal of Abnormal and Social Psychology, 52,* 151–158.

Schmit MJ, Ryan AM, Stierwalt SL, Powell AB. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80,* 607–620.

Stein LAR, Graham JR. (1999). Detecting fake-good MMPI-A profiles in a correctional facility. *Psychological Assessment, 11,* 386–395.

Tellegen A. (1993). Folk concepts and psychological concepts of personality and personality disorder. *Psychological Inquiry, 4,* 122–130.

Tellegen A, Grove WM, Waller NG. (1987). *Inventory of personal characteristics #7 (IPC-7).* Unpublished materials, University of Minnesota.

Wiggins JS. (1968). Personality structure. In Fransworth PR (Ed.), *Annual review of psychology* (Vol. 19). Palo Alto, CA: Annual Reviews.

Zerbe WJ, Paulhus DL. (1987). Socially desirable responding in organizational behavior: A reconceptulization. *Academy of Management Review, 12,* 250–264.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX

Items From the IPC-7 Included in the Trait Desirability Inventory (TDI)

Positive Valence
- Exceptional, special
- An ordinary, everyday person
- Deserve to be admired
- Impressive, remarkable
- Excellent, first rate
- Outstanding, superior
- Not exceptional, not that special

Negative Valence
- Dangerous to others, harmful
- Disgusting, horrible
- Awful, terrible
- Wicked, evil
- Odd, peculiar
- Deserve to be hated
- Cruel, mean
- Mentally disturbed, sick
- Treacherous, disloyal

Positive Emotionality (Extroversion)
- Do not talk much, uncommunicative
- Quiet
- Reserved, distant
- Lively, animated
- Talkative
- Prefer to be alone, a loner
- Playful
- Like to be with people, sociable

Negative Emotionality (Neuroticism)
- Not easily upset
- Often irritated by minor setbacks
- Can put worries out of mind
- Often jump and jittery
- Do not worry about little things
- Feelings are easily hurt
- Do not let many things bother or frustrate me
- Nervous, high strung

Agreeableness
- Strong, forceful
- Others think I am quarrelsome and contentious
- Easy on others, lenient
- Tough, uncompromising
- Stubborn, obstinate
- Get into arguments, argumentative
- Headstrong, willful
- Dislike arguments and conflict

continued

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX (continued)

Conscientiousness
Do things in an orderly and systematic manner
Keep belongings neat and tidy
Prompt, punctual, get things done on time
Cautious, circumspect
Well-organized
Like to improvise, play things by ear
Spontaneous, impulsive
Like to have a place for everything and everything in its place
Conventionality (Openness)
Conventional
Hold traditional values and beliefs
Politically radical, hold revolutionary views
Believe that strict discipline at home
Unusual, unconventional
Conservative
Strange

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.