Available online at www.sciencedirect.com
ScienceDirect
Journal of Research in Personality 42 (2008) 914–932
JOURNAL OF RESEARCH IN PERSONALITY
www.elsevier.com/locate/jrp

The good judge of personality: Characteristics, behaviors, and observer accuracy ☆

Tera D. Letzring *

Department of Psychology, Idaho State University, Pocatello, ID 83209, USA

Available online 25 December 2007

Abstract

Personality characteristics and behaviors related to judgmental accuracy following unstructured interactions among previously unacquainted triads were examined. Judgmental accuracy was related to social skill, agreeableness, and adjustment. Accuracy of observers of the interactions was positively related to the number of good judges in the interaction, which implies that the personality and behaviors of the judge are important for creating a situation in which targets will reveal relevant personality cues. Furthermore, the finding that observer accuracy was positively related to the number of good judge partners suggests that judgmental accuracy is based on more than detection and utilization skills of the judge.

Keywords: Good judge; Realistic accuracy; Personality; Judgmental accuracy; Person perception

1. Introduction

People frequently make important decisions based on their assessments of the personality of others. A decision regarding who should baby-sit your children will depend at least partly on how reliable and conscientious you judge your choices of baby-sitters to be; and

☆ This article is based on the doctoral dissertation by Tera D. Letzring titled "The Good Judge: Personality and Behavioral Correlates of Accurate Personality Judgment." Similar findings were also presented at the 2005 and 2006 annual meetings of the Society for Personality and Social Psychology. Data gathering was supported by National Institute of Mental Health Grant MH42427 to David C. Funder.
* Corresponding author. Fax: +1 208 282 4832.
E-mail address: letztera@isu.edu

0092-6566/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2007.12.003

who you decide to marry might depend on how trustworthy and loving you judge your dating partner to be. People will make better decisions when they accurately judge the personality of others, and knowledge about the process of accurate judgment may help people make more accurate judgments. One important aspect of the process of accurate judgment concerns the person making the judgment, or the judge. Theorizing about the good judge posits that a good judge should be knowledgeable about how personality relates to behavior, have high levels of cognitive ability and general intelligence, and be motivated to be accurate, among other characteristics (Funder, 1995, 1999). It is also possible that good judges are people who are able to obtain a relatively large amount of useful information about their targets.

There are two goals of the current study. The first goal is to examine the personality characteristics and behaviors of judges who make more accurate judgments of personality following a 50-min or 3-h interaction with partners with whom they were previously unacquainted. The second goal is to assess whether the accuracy of observers of videotaped interactions is related to the number of good judges in the interaction. The second goal is designed to assess whether targets which interact with good judges make more relevant cues to personality available, because if this is the case then someone observing the interaction should make more accurate judgments than someone observing an interaction without a good judge.

One conceptualization of the process of accurate judgment is the Realistic Accuracy Model (RAM), which is based on Brunswik's Lens Model (Brunswick, 1956; Funder 1995, 2001a, 2001b). According to RAM (Funder, 1995), four stages must be successfully completed for accuracy to be achieved: the target has to do something that is *relevant* to the personality characteristic being judged, the cue has to be *available* to and *detected* by the judge, and the judge has to appropriately *utilize* the cue to form a judgment. Therefore, good judges should have a relatively large amount of relevant information available to them and be able to detect and appropriately use these cues. Previous research suggests that the last two stages—detection and utilization—are the key stages pertaining to being a good judge (Funder, 1999). The current study examines whether the first two stages—relevance and availability—are also key to being a good judge.

There are many methods for measuring the accuracy of personality judgments. One common distinction is between person-centered and variable-centered approaches (Biesanz and West, 2000; Biesanz et al, 2007). In person-centered approaches, accuracy is computed across several traits of the target person, whereas in variable-centered approaches accuracy is computed across several targets for a single trait. The current project uses a person-centered approach, because this approach allows for the computation of an accuracy score for each judge or each target and it is necessary to compute an accuracy score for each judge in order to examine how accuracy of the judge is related to the personality and behavior of the judge.

However, with the person-centered approach, the issue raised by Cronbach (1955)—that accuracy scores are composed of multiple components—must be considered. There are several ways to deal with this issue. One way is to use a method other than difference scores to compute accuracy. The current study uses correlations, which alleviates some of the problems with difference scores but not all of them. A second way is to separate the components of accuracy scores and look only at differential accuracy. However, this practice may remove important information from the overall accuracy score (Colvin and Bundick, 2001). Another solution is to use the unpartitioned scores with the awareness that

they are composed of multiple components. The main argument for using the unpartitioned scores is that some components might be important aspects of accuracy (Colvin & Bundick). For example, stereotype accuracy—the ability to accurately judge the average of all targets—is likely to be an important skill for judging others. Furthermore, when one is concerned with comparing accuracy scores, rather than looking at the absolute level of accuracy, the unpartitioned scores do not present such a large problem. Even if elevation, differential elevation, and stereotype accuracy exist, they are likely to exist to a similar degree across conditions and therefore the component of interest (differential accuracy) is the one most likely to differ among conditions. The current study does not focus on the absolute level of accuracy scores but instead examines how accuracy is related to other variables (personality and behavior) and how accuracy varies as a function of the number of good judges in an interaction. For this reason, the unpartitioned score is an appropriate measure of accuracy in the current study.

Another important element of the measurement of accuracy is the criterion that is used to determine accuracy. In self-other agreement, the self-rating is used as the criterion and the judge is said to be accurate to the degree that his or her ratings corresponds with the self-ratings. The difficulty with self-other agreement is that it assumes that the self is an accurate judge of his or her own personality, which is only true when people are willing and able to provide accurate self-judgments. Consensus, or the agreement about a target among two or more judges, has been used as a proxy for accuracy. Hofstee (1994) proposed that the average judgment of a sufficient number of people who are knowledgeable about the target is an adequate criterion for accuracy. The benefit of combining several judgments is that errors are likely to cancel each other out, leaving a better estimate of the true personality of the target (Colvin and Bundick, 2001). The difficulty with consensus is that it may not necessarily reflect what the target is really like, but instead is an index of how well two or more other people agree about what the target is like. A potentially more fruitful way to assess accuracy is to use a criterion that includes judgments from several people who know the target well (Funder, 1995, 1999), which is referred to as *realistic accuracy*. The defining feature of realistic accuracy is the attempt to approach a description of what the target is really like by creating a criterion based on multiple sources of information (Webb et al., 1966). The current study assesses realistic accuracy.

Data analyses for the current project will use ratings provided by the self, real-life acquaintances, and clinician-interviewers as the criterion for determining realistic accuracy. Furthermore, the accuracy scores in the current analyses are based on profile correlations, so that a judge achieves a high level of accuracy by ordering the characteristics of a target in a way that is consistent with the way the sources for the accuracy criteria also ordered the characteristics of that target. For example, if a target is rated by the sources of accuracy criteria as more friendly than manipulative, then a judge who also describes the target as more friendly than manipulative would be accurate.

1.1. Previous findings

Findings from early research on the good judge suggest that judgmental accuracy is related to several characteristics including independence, trustfulness, sympathy, courage, a sense of humor, experience with human nature, maturity, similarity to the target, intelligence, and social skills (Adams, 1927; Allport, 1937; Vernon, 1933). Taft (1955) found that characteristics of good judges of others included gender (with a slight advantage

for women), intelligence, aesthetic ability and sensitivity, emotional stability, self-insight, social skills, and social detachment. Taft concluded that "the main attributes of the ability to judge others seem to lie in three areas: possessing appropriate judgmental norms, judging ability, and motivation" (p. 20). Largely as a result of Cronbach (1955) critique of the accuracy literature, research shifted to focus on a cognitive approach to personality judgment based on object perception (Swann, 1984). This research emphasized the utilization of cues to personality (which, according to RAM, is only one of four stages of personality judgment) and used hypothetical targets. In the 1980s, research on personality judgment began to move away from an exclusive focus on the utilization stage and back to looking at the entire process of accurate judgment using real targets.

Recent research suggests that a variety of personality characteristics of judges are related to judgmental accuracy. For example, Kolar (1995) found higher levels of self-other agreement among male judges who rated themselves positively, as interpersonally experienced, and as not anxious, defensive, or concerned with what others think; and among female judges who scored high on intelligence and rated themselves high in openness to experience. Some evidence suggests that judges who score high on sociability and measures of sensitivity to nonverbal cues make judgments of strangers that agree more with the strangers' self-ratings (Ambady et al., 1995), although other research has failed to confirm that relationship (Colvin and Bundick, 2001). Several characteristics (such as warm, outgoing, sympathetic, empathic, interpersonally oriented, and not hostile, manipulative, or autonomous) as judged by the self and parents have been found to be related to accuracy (Vogt and Colvin, 2003). Davis and Kraus (1997) conducted a meta-analysis and found that good judges were generally field independent, cognitively complex, psychologically adjusted, socially sensitive, and not rigid or dogmatic. Also, self-other agreement has been found to be positively related to intelligence but unrelated to openness to experience, emotional stability, extraversion, agreeableness, and conscientiousness (Christiansen et al., 2005). An important caveat to findings about the characteristics of the good judge is that no single characteristic or set of characteristics has emerged that is consistently associated with judgmental accuracy (Davis and Kraus, 1997). One possible explanation for the mixed results is that different researchers use different criteria for accuracy and measure accuracy in different ways, which may have important consequences for how accuracy is related to the characteristics and behaviors of good judges.

Additionally, some researchers have expressed doubts about whether judgmental accuracy is an individual difference, primarily based on low reliability of accuracy scores across targets (Kenny and Albright, 1987). Therefore, this project will also examine whether reliability across targets is high enough to support the idea that judgmental accuracy is an individual difference.

1.2. Expected relations between personality of judges and judgmental accuracy

As mentioned previously, the Realistic Accuracy Model (RAM) describes a four-stage process of accurate judgment (Funder, 1995, 1999). It is important to consider how the personality and behaviors of the judge are likely to affect success at all four stages in order to form a complete picture of how the judge affects the entire judgment process. However, previous work has focused on the last two stages of RAM and not considered how the characteristics and behaviors the judge may influence the first two stages (Funder, 1995).

Judges may influence the first two stages, relevance and availability, by making others feel comfortable and able to be themselves so that they will reveal information about themselves that is useful for judging a variety of personality traits. Judges around whom others feel comfortable might have many of the positive characteristics that have been found to be related to judgmental accuracy, such as being warm, genuine, and sympathetic, and not hostile, anxious, or defensive (Colvin and Bundick, 2001; Kolar, 1995; Vogt and Colvin, 2003), as well as good social skills and agreeableness. Furthermore, judges who encourage targets to talk about their thoughts and feelings will likely have more personality-relevant cues available to them (Anderson, 1984; Funder, 1995, 1999). Several characteristics of judges could be related to a target's willingness to talk about thoughts and feelings, some of which have already been found to be related to judgmental accuracy such as having an interpersonal orientation (Vogt and Colvin, 2003), being socially sensitive (Davis and Kraus, 1997), trustful and talkative (Adams, 1927), not anxious or self-defensive (Kolar, 1995), and sympathetic and empathic and not hostile or manipulative (Colvin and Bundick, 2001).

Judges will influence the detection stage by being observant and paying attention to their surroundings rather than being inwardly-focused (Adams, 1927). Engaging in eye contact and seeming interested in what the target has to say are behaviors that may indicate that the judge is attending to the target. Additionally, targets are more likely to reveal more information about themselves when it appears that their interaction partners are interested in them and paying attention to what they are saying.

Judges will influence the utilization stage by being able to correctly combine and interpret cues. Judges with high intellectual capacity, including general and social intelligence, are expected to be more successful at this stage because they should be able to remember and successfully manipulate more cues (Funder, 1999). Some research has found support for the idea that intelligence and cognitive complexity are related to judgmental accuracy (Adams, 1927; Christiansen et al., 2005; Davis and Kraus, 1997; Reimer et al., 2006; Kolar, 1995; Vernon, 1933).

Finally, a characteristic that is likely to affect judgmental accuracy at all stages of the judgment process is psychological adjustment (Davis and Kraus, 1997). Well-adjusted judges are likely to be involved in comfortable interactions in which their partners make relevant information available, to not be focused on themselves and therefore have the ability to detect information, and to have coherent thought patterns and an appropriate view of human nature so that they can successfully utilize cues. Therefore, characteristics and behaviors that indicate psychological adjustment are expected to be related to judgmental accuracy.

Expectations can also be based on other, more general, theories. For example, trait activation theory (Tett and Guterman, 2000) proposes that "personality traits are expressed as responses to trait-relevant situational cues" (Tett and Burnett, 2003, p. 502). A socially skilled and agreeable judge may create the type of situation that allows for the expression of multiple cues that are relevant to personality, and therefore the judge will make more accurate judgments of personality. Theory based on the interpersonal circumplex suggests that people respond in predictable ways to certain behaviors, such as responding to dominance with submission and to affiliation with affiliation (Markey and Kurtz, 2006; Tracey, 2004). It is possible that people may respond to socially skilled behavior with more relevant information about themselves.

1.3. Hypotheses

First, it is hypothesized that judgmental accuracy is related to personality characteristics and behaviors of the judge that influence the judgment process at one or more stages of RAM, and especially to social skills, agreeableness, and psychological well-being. These characteristics are especially important because they are likely to increase the amount of relevant cues that are available to the judge. Without success at the two first two stages of RAM, even judges who are very good at detecting and utilizing cues will not be accurate because they do not have appropriate and useful information to use in making their judgments. Second, it is hypothesized that judgments based on a video observation will be more accurate when the target interacts with good judges because more information relevant to personality will be available to the observer-judge.

2. Method

Data from the Riverside Accuracy Project—Phase 2 (RAP-2) were used for the analyses evaluating the hypothesis that judgmental accuracy is related to personality characteristics and behaviors of the judge that influence the judgment process at one or more stages of RAM. Results pertaining to correlations between accuracy and behavior were presented in a previous, unpublished presentation (Letzring et al., 2005). Other published papers from this data set include examinations of how information quantity and quality are related to judgmental accuracy (Letzring et al., 2006), the generalization of self-report scales of ego-control and ego-resiliency (Letzring et al., 2005), the relation between impulsivity and narcissism (Vazire and Funder, submitted for publication), and the relation between personality and academic achievement (Wagerman and Funder, 2007). Data from RAP-2 and data collected at Idaho State University were used in analyses evaluating the hypothesis that judgments based on a video observation will be more accurate when the target interacts with good judges.

2.1. Participants

A total of 506 undergraduates from the University of California, Riverside (UCR) participated in the original RAP-2 project were paid $10 per hour. A core group of 180 target participants were recruited from fliers advertising "Research on Personality." These participants helped with the recruitment of up to two close acquaintances (for a total of 326 acquaintance informants) to provide personality descriptions of themselves and the target participants. Of the 180 target participants, 142 are included in the first set of analyses that examine the correlations of personality and behavior with judgmental accuracy (72 female, 70 male; 22 Caucasians, 56 Asians, 26 Hispanics, 18 African Americans, and 20 with other or unknown ethnicities). These 142 participants participated in a 50-min or 3-h 3-person interaction.[1] Analyses for a few items have sample sizes of as low as 137 as a result of missing data. In the second set of analyses, 139 targets were each observed via videotaped interactions by three participants from UCR and 101 targets were each observed by

[1] The participants not included made judgments of their partners without being given the opportunity to engage in interaction first.

between 1 and 14 judges from Idaho State University (ISU). There were a total of 89 judges from ISU (45 female, 44 males; 82 Caucasian, 7 other ethnicity). All targets were from UCR and were part of the core group of participants who were videotaped during the initial 3-person interaction. The judges from UCR were also part of the core group of participants. The judges from ISU took part in a separate study for which they received course credit.

2.2. Procedures

2.2.1. Target participants

Participation in RAP-2 involved a pre-session, three separate laboratory sessions, and the completion of several take-home packets of self-report measures (only some of which are relevant to the current project). At the pre-session, participants were asked to provide names and contact information of two *acquaintances*, defined as the two locally available people who "knew them best" and were able to come to the laboratory to complete a set of questionnaires. During the first session, participants interacted in 3-person groups in one of five types of interactions. In all cases, the three interaction partners were not previously acquainted. In the Minimal Information condition, the participants were not given an opportunity to interact and were told that the researchers were interested in first impressions. In three experimental conditions, the participants interacted for 50 min under different instructions: to complete a packet of trivia-like questions as a group, to talk about whatever they would like, and to get to know each other as well as possible. In the final experimental condition, the participants interacted for three hours under the instructions to talk about whatever they would like (see Letzring et al., 2006 for a more detailed explanation). Following each interaction, the participants rated the personalities of both interaction partners. Judges are included in the first set of analyses if they interacted in a condition other than Minimal Information condition.

2.2.2. Video observations

Two data sets were used for the observation analyses. In both observations, the targets of judgment were participants from the UCR data. The first data set consisted of participants from UCR who watched a videotape of an interaction that differed from the type of interaction they participated in previously and did not include anyone whom they recognized. After viewing the entire interaction (50 min or 3 h), the participant rated the personality of all three interaction partners. Only the data from observations of conditions other than the Minimal Information condition are included in the second set of analyses. The second data set consisted of participants from ISU who observed the second 5-min segments of four interactions. Participants did not all observe the same four interactions, but were assigned to different subsets of interaction videos. Participants were asked to focus on one person in the interaction and rated the personality of only that person directly following each observation. Participants at ISU observed three types of interactions: the 50-min and 3-h interactions in which people were told to talk about whatever they would like and the 50-min interaction in which people were told to get to know each other as well as possible.

2.2.3. Clinician-interviewers

Each participant from UCR individually participated in an hour-long life history interview conducted by one of six professionally trained clinical psychologists. All of the clini-

cians had at minimum a Master's degree in clinical psychology and training and experience with clients of the same age as our college student population. The participants were told that they would be interviewed by a "professionally trained interviewer" to diminish demand characteristics. The clinicians conducted a semi-structured interview adapted from a protocol used by the Institute of Personality Assessment and Research (IPAR; Craik et al., 2002). The interview was adapted to better apply to college students and to capture a broad range of personality-relevant information without explicitly asking about sensitive topics and risky behaviors. Each interview started with the clinician asking the participant to "tell me something about yourself." The interview then covered a broad range of topics including college and academic experiences, future plans, interpersonal relationships, and childhood and family history. In conclusion, each participant was asked to describe "a defining event in your life that had a significant impact on or changed your life in some way."[2] Following the interview, the clinician described the personality of the target participant.

2.2.4. Acquaintances

Project staff contacted the acquaintances identified by each UCR participant and scheduled them to come to the laboratory to describe the personality of the target participant.

2.3. Measures

2.3.1. The California adult Q-set (CAQ; Block, 1961; as modified by Bem and Funder, 1978)

The CAQ includes 100 items that describe a broad range of important personality characteristics, such as "critical, skeptical, not easily impressed" and "sympathetic and considerate." In the *Q-sort*, items are printed on separate cards and ipsatively placed within a quasi-normal forced-choice distribution on a scale ranging from 1 (*not at all characteristic*) to 9 (*extremely characteristic*). The Q-sort rating format was used to obtain descriptions from the clinical psychologists. In the *Q-item rating*, each item is separately rated on a Likert-like scale with the same range and labels as the Q-sort. The CAQ-item rating format was used to obtain the self-descriptions, descriptions of the participants provided by the acquaintances, and the judges' descriptions of the targets following the interaction and observation.

2.3.2. Big Five inventory (BFI; John et al., 1991)

The BFI is a 44-item self-report measure of five broad personality traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. A third-person rating form was used by the acquaintances to describe the target participants.

2.3.3. Psychological well-being (PWB; Ryff, 1989)

This 84-item scale measures a person's level on six facets of psychological well-being; autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. The correlations of the dimensions of the published

[2] The full protocol for the clinical interview is available at http://www.faculty.ucr.edu/%7Efunder/lab/supplemental.htm.

100-item version and the unpublished 84-item version used in the current study were reported to range from .97 to .99 (C. Ryff, personal communication, September, 2001).

2.3.4. Narcissistic personality inventory (NPI; Raskin and Hall, 1979, 1981; Raskin and Terry, 1988)

This 40-item measure requires test-takers to choose which of two items they agree with more (i.e. "Modesty doesn't become me." or "I am essentially a modest person"). The NPI assesses 7 components of narcissism (authority, self-sufficiency, superiority, exhibitionism, exploitativeness, vanity, and entitlement), which all contribute to the full-scale score.

2.3.5. Inventory of interpersonal problems (IIP; Soldz et al., 1995)

This 32-item scale measures eight dimensions of interpersonal problems: domineering, vindictive, cold, socially avoidant, non-assertive, exploitive, overly nurturant, and intrusive.

2.3.6. Social skills inventory (SSI; Riggio, 1986)

This 90-item scale measures 7 dimensions of social skills: emotional expressivity, emotional sensitivity, emotional control, social expressivity, social sensitivity, social control, and social manipulation.

2.3.7. Adult romantic attachment (Brennan et al., 1998)

This 36-item scale measures two dimensions of romantic attachment, anxiety and avoidance. An example item for avoidance is, "Just when my partner starts to get close to me I find myself pulling away." An example item for anxiety is, "I worry a lot about my relationships".

2.3.8. Riverside behavioral Q-sort

The Riverside Behavioral Q-sort (RBQ; Funder et al., 2000) is a set of 64 descriptions of behaviors that are likely to occur during a social interaction, such as "tries to control the interaction" and "expresses agreement frequently." The RBQ is used to describe behavior by placing the items into a quasi-normal, forced-choice distribution ranging from 1 (*not at all or negatively characteristic of the behavior of the person*) to 9 (*highly characteristic of the behavior of the person*).

2.4. Behavioral coding

The three-person interactions described previously was used as the sample of behavior for the judges. The behavioral codings were completed by a set of trained undergraduate research assistants. Each judge was coded by four independent coders, and each coder only coded one judge from each interaction. To code a judge's behavior, a coder first watched the entire interaction while focusing on one participant/judge. Then, the coder used the RBQ to describe the behavior of that judge during the interaction. To arrive at the final description of the behaviors of a particular judge, the codings for each item of the RBQ were averaged across the four coders. To compute the reliability of this composite, Cronbach's alpha was calculated for each of the 64 items, across the 142 participants (average $\alpha = .61$).

2.5. Computations of accuracy scores

The current project measures realistic accuracy using a broad-based criterion that is composed of ratings provided by three types of knowledgeable informants for each target: the self, acquaintances,[3] and clinicians. In the first set of analyses, accuracy scores were computed using profile correlations, so that for every judge-target pair the 100 CAQ ratings of the judge were correlated with the 100 CAQ criterion ratings of the target, separately for each accuracy criterion. In this way, each accuracy score represents the level of agreement between judgments and three criteria across 100 personality characteristics, averaged across two targets. In the second set of analyses, accuracy scores were computed in the same way, but this time the scores were averaged across judges for each target, so that there was one accuracy score associated with each target.

3. Results

3.1. Descriptive statistics and gender comparison

The mean accuracy for all 142 judges from UCR was .33 ($SD = .13$, range $= -.12$ to .63). The mean accuracy for all 89 judges from ISU was .21 ($SD = .13$, range $= -.10$ to .45). In the UCR data there was not a significant difference between female judges ($M = .30$, $SD = .12$) and male judges ($M = .30$, $SD = .14$), $t(140) = .45$, $p = .65$, but in the ISU data female judges ($M = .23$, $SD = .11$) were more accurate than male judges ($M = .17$, $SD = .14$), $t(87) = 2.46$, $p = .02$.

3.2. Reliability of accuracy scores

The reliability of the accuracy scores reflects the consistency of a judge's level of accuracy across the two targets. Reliability can also be used to provide evidence concerning whether judgmental accuracy is an individual difference. To calculate this reliability, for each judge-target pair the level of agreement with the three sources of criterion ratings were first averaged, which resulted in two scores per judge, one for each target. Then, these two scores were used to compute an alpha reliability, which was $\alpha = .48$. This reliability seems low for a reliability score, but it was only based on two scores (which can be thought of as the reliability of a two-item test). In fact, the zero-order correlation between accuracy scores for the two targets of each judge was significant ($r = .32$, $p = .0001$), suggesting that judges who achieved a high level of accuracy for ratings of one target were also likely to achieve a high level of accuracy for ratings of the other target.[4] This level of reliability, which was based on only two targets, implies that judgmental accuracy is an important individual difference. The Spearman-Brown formula can be used to project the reliability for a larger number of targets using the correlation for two targets. For example, for 15

[3] When ratings from two acquaintances were available, the average rating on each item was used as the accuracy criterion. The average reliability of the ratings provided by the two acquaintances, in terms of a profile correlation between the 100 ratings of the acquaintances, was $\alpha = .49$.

[4] Differential accuracy scores were also computed by first standardizing item ratings across judges and targets and then computing profile correlations between sets of ratings. However, these scores were not sufficiently reliable ($\alpha = .11$) to be used in further analyses.

targets the reliability is estimated to be .88. Additionally, many other factors influence the level of accuracy, such as the characteristics of the target and the type of information that is available to the judge, so even a moderate correlation between the accuracy of two targets should be considered important.

3.3. Personality correlates of judgmental accuracy

The correlations between judgmental accuracy and personality are presented in Table 1. Reports by both the self and acquaintances generated positive correlations between accuracy and agreeableness and negative correlations between accuracy and power-orientation. Accuracy was also positively correlated with the PWB purpose in life subscale, self-ratings of behaving in a sympathetic or considerate manner and being genuinely submissive, and rating by acquaintances of having a clear-cut, internally consistent personality and feeling

Table 1
Personality correlates of judgmental accuracy based on the interaction

Personality trait	Profile accuracy
Positive correlations	
Acq BFI: Agreeableness	.24**
Self CAQ 17: Behaves in a sympathetic or considerate manner	.20*
Acq CAQ 75: Has a clear-cut, internally consistent personality	.20*
Self CAQ 14: Genuinely submissive	.19*
Acq CAQ 74: Is subjectively unaware of self-concern, feels satisfied with self	.19*
PWB: Purpose in life	.18*
Self BFI: Agreeableness	.17*
Negative correlations	
IIP: Domineering	−.30**
IIP: Vindictive	−.27**
Attachment avoidance	−.26**
Self CAQ 91: Is power-oriented; values power in self and others	−.25**
NPI: Full scale score	−.24**
IIP: Cold	−.23**
Self CAQ 53: Unable to delay gratification	−.22**
Acq CAQ 61: Creates and exploits dependency in people	−.21*
Acq CAQ 68: Is basically anxious	−.21*
IIP: Intrusive	−.20*
SSI: Social expression	−.20*
Self CAQ52: Behaves in an assertive fashion	−.19*
Acq CAQ 91: Is power-oriented; values power in self and others	−.19*
Acq CAQ 94: Expresses hostile feelings directly	−.19*
SSI: Emotional expression	−.18*
Self CAQ 23: Tends to transfer or project blame	−.18*
Self CAQ 27: Shows condescending behavior in relation with others	−.18*
Self CAQ 73: Tends to perceive many different context in sexual terms	−.18*
Acq CAQ 78: Feels cheated and victimized by life; self-pitying	−.18*
Acq CAQ 10: Anxiety and tension find outlet in bodily symptoms	−.17*
Acq CAQ 89: Compares self to others	−.17*

Note. Acq BFI, Big Five Inventory completed by acquaintances; Acq CAQ, California Adult Q-set rating completed by acquaintances; IIP, Inventory of Interpersonal Problems; NPI, Narcissistic Personality Inventory; PWB, Psychological Well-Being; Self BFI, self-rating of the Big Five Inventory; Self CAQ, self-rating of the California Adult Q-set; SSI, Social Skills Inventory. $N = 138–142$ (depending on missing items).

Table 2
Behavioral correlates of judgmental accuracy based on the interaction

#	Item	Profile accuracy
Positive correlations		
55	Emphasizes accomplishments of self, family or housemates	.32**
60	Engages in constant eye contact	.28**
15	Compares self to other(s)	.21*
33	Expresses warmth	.21*
43	Seems to enjoy the interaction	.21*
13	Seems to like partners	.20*
46	Displays ambition	.20*
4	Seems interested in what partners have to say	.19*
25	Expresses sympathy towards partners	.18*
57	Speaks in a loud voice	.18*
Negative correlations		
27	Seeks reassurance from partners	−.28**
34	Tries to undermine, sabotage, or obstruct	−.21*
64	Partners seek advice from subject	−.21*
1	Expresses awareness of being on camera or in an experiment	−.19*

Note. $N = 142$.
$^{**}p < .01$, $^{*}p < .05$.

satisfied with one's self. Accuracy was negatively related to several subscales of the IIP (domineering, vindictive, cold, and intrusive), the full scale score of the NPI, two subscales of the SSI (social expression and emotional expression), the avoidance dimension of the Adult Romantic Attachment scale, and several items from the self-rating and the acquaintance-rating of the CAQ.[5]

3.4. Behavioral correlates of judgmental accuracy

Fourteen out of 64 correlations between judgmental accuracy and the behaviors of the judge reached significance at the .05 level (see Table 2). The significant correlations include engages in constant eye contact, expresses warmth, seems to enjoy the interaction, seems to like partners, seems interested in what the partners have to say, expresses sympathy, and tries to undermine, sabotage, or obstruct (negative). These behaviors are consistent with the idea that good judges have good social skills and therefore have more relevant information available to them when making judgments. The behavior most consistent with good detection of cues is engages in constant eye contact.

3.5. Accuracy of observers as a function of number of good judge partners

If good judges have an advantage because they are able to elicit more useful information from their targets, then people who observe targets interacting with good judges

[5] Significant correlations between judgmental accuracy and some characteristics expected to be related were not found, including most notably intelligence as measured with the Wonderlic Personal Inventory (Wonderlic, Inc., 1999), $r = -.01$, $p = .89$, and attributional complexity as measured by the subscales of the Attributional Complexity Scale (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986), r's = −.11–.07, p's > .19.

should also have access to this information and make more accurate judgments than observers of targets without good judge partners. To test this prediction, participants in the three-person interactions who had average accuracy scores for the two interaction partners in the top quartile of scores ($r \geqslant .435$) were designated as good judges. Then, for each target it was determined whether they interacted with zero, one, or two good judges. Next, the accuracy with which each target was rated by the judges who observed the videotaped interaction was computed. Profile correlations were computed across the 100 items of the CAQ between the ratings of the target provided by the judges and the ratings provided by the self, acquaintances, and clinicians. Correlations were first averaged across the three criteria for each target ($\alpha = .85$ for RAP-2 data and $\alpha = .89$ for ISU data), and then across all judges for each target ($\alpha = .83$ for RAP-2 data and $\alpha = .93$ for ISU data). This procedure resulted in an accuracy score for each target. Finally, analysis of variance (ANOVA) and contrast analysis (Rosenthal et al., 2000) were used to examine the relationship between number of good judge partners and the accuracy with which targets were judged by observers. The specific prediction tested by contrast analysis was that accuracy would increase linearly with the number of good judge partners, and contrast weights of −1 (0 good judge partners), 0 (1 good judge partner), and +1 (2 good judge partners) were used. The convention for contrast analysis is to present one-tailed p-values because a directional hypothesis is being tested. In a second but similar analysis, designation of good judges from the interaction were based on a median split such that good judges represented those whose average accuracy scores of the two interaction partners were in the top half of accuracy scores ($r \geqslant .325$).

In the RAP-2 data, observers of targets with more good judge partners (based on the top quartile of scores) were more accurate than observers of targets with fewer good judge partners, $F(2, 132) = 5.13$, $p = .007$ (see Fig. 1). Based on post-hoc Tukey HSD tests using an alpha of .05, observers of targets with two good judge partners ($n = 16$, $M = .36$, $SD = .21$) were more accurate than observers of targets with zero good judge partners ($n = 89$, $M = .24$, $SD = .14$), but were not more accurate than observers of targets with one good judge partner ($n = 30$, $M = .29$, $SD = .13$). There was also not a difference between the accuracy of observers of targets with one or zero good judge partners. The contrast analysis was significant, $t_{\text{contrast}}(132)=2.98$, $p_{\text{1-tailed}} = .002$, $r = .25$, suggesting that accuracy scores increased in a linear fashion as the number of good judge partners increased. However, when good judges were designated as those who achieved accuracy scores above the median, there was a less pronounced difference among accuracy scores of targets as a function of number of good judge partners (two good judge partners: $n = 41$, $M = .31$, $SD = .17$; one good judge partner: $n = 50$, $M = .26$, $SD = .14$; zero good judge partners: $n = 44$, $M = .23$, $SD = .15$), $F(2, 132) = 2.57$, $p = .08$. Even though the ANOVA did not reach conventional levels of significance, this pattern was consistent with a linear increase in accuracy as number of good judge partners increased, $t_{\text{contrast}}(132) = 2.43$, $p_{\text{1-tailed}} = .008$, $r = .21$.

In the ISU data, there were not differences in accuracy when the top quartile of accuracy scores were used to designate good judges (two good judge partners: $n = 16$, $M = .24$, $SD = .15$; one good judge partner: $n = 26$, $M = .21$, $SD = .16$; zero good judge partners: $n = 56$, $M = .18$, $SD = .16$), $F(2, 95) = .96$, $p = .39$ (see Fig. 2). The contrast analysis showed that the data were consistent with the direction of the prediction, but not to a significant degree, $t_c(95) = 1.31$, $p_{\text{1-tailed}} = .10$, $r_c = .13$. However, when good judges were designated as those who achieved accuracy scores above the median, there was a difference

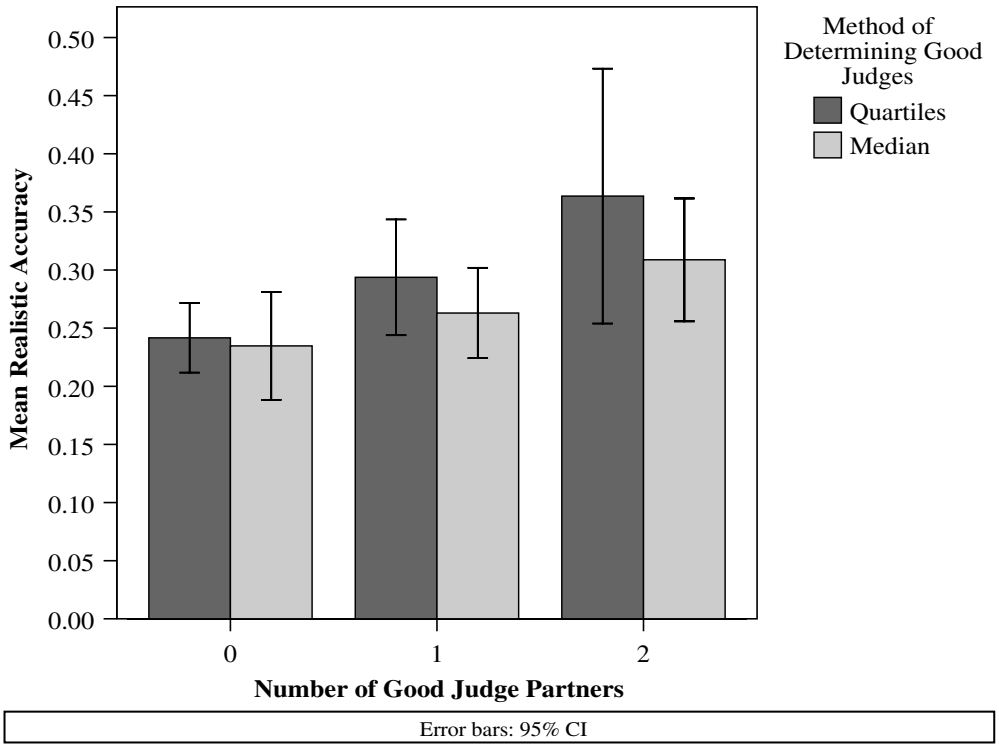

Fig. 1. Realistic accuracy scores of observers of videotaped interaction in which the targets interact with different numbers of good judge partners, using the RAP-2 data.

among accuracy scores as a function of number of good judge partners, $F(2, 95) = 7.73$, $p = .0008$. Post hoc analyses using the .05 significance level reveal that observers of targets with two good judge partners ($n = 35$, $M = .25$, $SD = .16$) or one good judge partner ($n = 37$, $M = .21$, $SD = .16$) were more accurate than observers of targets with zero good judge partners ($n = 26$, $M = .10$, $SD = .12$), but observers of targets with two good judge partners were not more accurate than observers of targets with one good judge partner. This pattern was also consistent with a linear increase ins accuracy as number of good judge partners increased, $t_{\text{contrast}}(95) = 3.91$, $p_{\text{1-tailed}} < .001$, $r = .37$.

4. Discussion

It is necessary to provide evidence that judgmental accuracy is an individual difference before considering the personality and behavioral correlates of this variable. The correlation between the accuracy of two targets was statistically significant, and the reliability was small but acceptable for an essentially two-item test, suggesting that judgmental accuracy is an important individual difference.

4.1. Personality characteristics and behaviors of good judges

A relatively large number of personality characteristics that are related to judgmental accuracy seem to play a role at the relevance and/or availability stages of the judgment

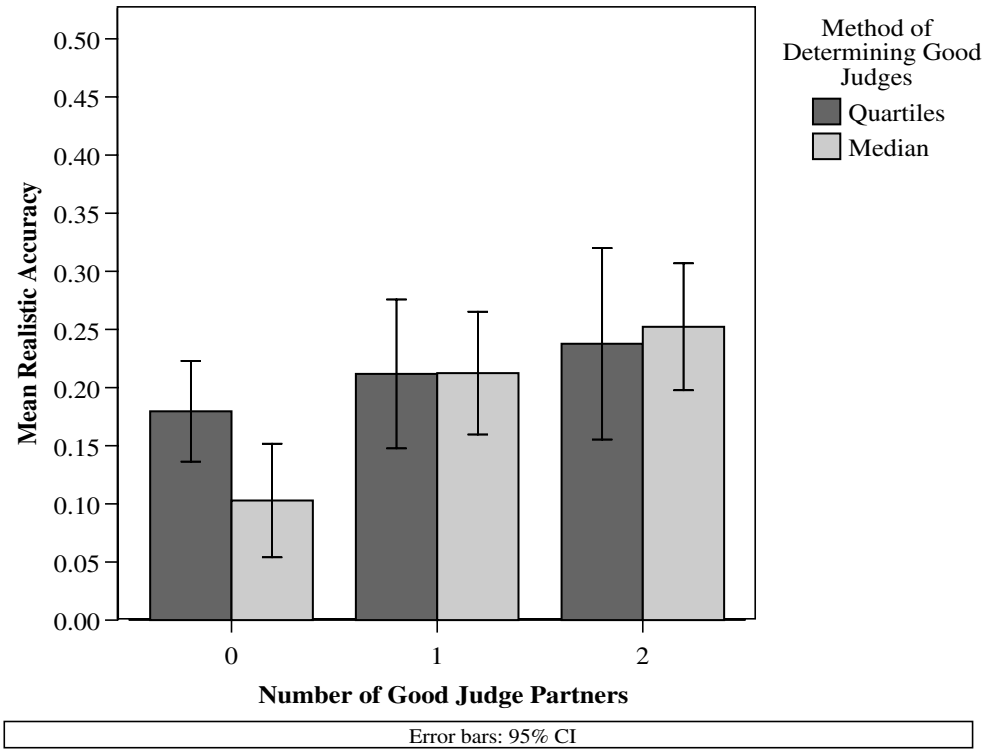

Fig. 2. Realistic accuracy scores of observers of videotaped interaction in which the targets interact with different numbers of good judge partners, using the ISU data.

process. The findings suggest that good judges are agreeable, have good social skills, are psychologically adjusted, and posses several other favorable characteristics. Therefore, the findings imply that good judges are people around whom others are likely to feel comfortable and willing to reveal information about their true personalities. The ability to obtain relevant cues to personality seems to be an especially important part of the judgment process, and a part of the process that is likely to be related to the personality of the judge.

The behavior of the judge during an interaction is also related to judgmental accuracy. Judges who exhibited social skills—in terms of expressing warmth, seeming to enjoy the interaction, seeming to like the interactions partners, and not undermining—achieved higher accuracy than judges who did not. Also, judges who exhibited behaviors related to the detection stage—engaging in eye contact and seeming interested in what the partners had to say—achieved higher accuracy. These behaviors may elicit more cues from the targets, because the targets are more likely to be comfortable and to reveal more information about themselves because it appears that the judge is paying attention to them.

4.2. Accuracy of observers

Observers of targets with more good judge partners were more accurate than observers of targets with fewer good judge partners. The finding held for observations of the entire 50-min or 3-h interactions and for observations of only 5 min when a median split in accu-

racy scores was used to designate good judges. The difference was also significant for observations of the entire 50-min or 3-h interaction when the top quartile was used to designate good judges. The difference was in the predicted direction but not significant when observations were of only a 5-min segment of the interaction when good judges were designated using the top quartile of scores. These findings support the idea that good judges create situations in which targets are more likely to reveal relevant cues to personality and that these cues are even available to observers and allow the observers to make more accurate judgments of these targets.

These findings add to the current knowledge about the good judge because it has been assumed that the personality and behaviors of the judge would be most likely to affect the detection and utilization stages of the judgments process (Funder, 1999). In fact, intelligence is one characteristic that has shown consistent relations with judgmental accuracy (Adams, 1927; Davis and Kraus, 1997; Taft, 1955; Vernon, 1933). However, the current findings highlight the importance of the first two stages of RAM—relevance and availability—for making accurate judgments of personality. Good judges appear to be able to elicit relevant information from targets that can then be detected and utilized in order to make accurate judgments. Furthermore, this information is available to observers of these interactions and can also be detected and utilized by them to make more accurate judgments.

4.3. Limitations

The first limitation of these data is that judges and targets interacted in three or four different situations. All situations were analyzed together, even though evidence has been found that levels of accuracy differ across these situations (Letzring et al., 2006). However, this only presents a minor concern for the present analyses and actually offers an advantage in terms of generalizability. Since not all judges interacted with and observed the targets in the same context, the current results do not apply to only one situation but to four slightly different unstructured social interactions that involve three previously unacquainted people.

Another limitation of the current study is that all ratings were based on first encounters between judges and targets. These data are useful for insight regarding the characteristics and behaviors related to the accuracy first impressions, but do not provide information regarding judgments of longer-term acquaintances. Early research on judgmental accuracy suggested that the ability to judge strangers differs from the ability to judge acquaintances (Vernon, 1933), so it is important to keep in mind that the current findings are based on first encounters between judges and targets and may not be directly generalizable to other relationships.

4.4. Future directions

4.4.1. Examine judgmental accuracy when targets are acquaintances

The current project has revealed much about characteristics and behaviors of good judges who have just met the targets. An important next step would be to examine how accuracy is related to characteristics and behaviors of judges who are well-acquainted with their targets. It is expected that characteristics related to the relevance and availability stages would still relate to judgmental accuracy when acquaintances are being judged, because it would still be important for the acquaintances to reveal information about

themselves, and this is most likely to occur when acquaintances trust the judge and feel comfortable around the judge. Another interesting question is whether observers would have higher accuracy when they observe good judges who are acquainted with the target vs. good judges who are unacquainted with the target. It is reasonable to assume that targets would reveal more relevant information when interacting with acquaintances than when interacting with strangers, and perhaps the "goodness" of the judge would be less of an issue among acquaintances than among strangers.

4.4.2. Examine the usefulness of the information available to judges

Support has been found for the idea that good judges have more relevant information available to them, and an additional question to ask concerns how types of information differ in relevance. In other words, what types of information are most useful for making accurate judgments of personality? It is possible to manipulate the type of information that is available to judges and compare the accuracy of the judgments that are based on different types of information. Results from this type of research will help researchers determine which types of information are most useful for making judgments of personality in general or of certain personality traits.

5. Conclusion

Good judges of personality possess personality traits related to social skills and agreeableness, and possess several other positive characteristics. Good judges also exhibit behaviors consistent with social skill and that indicate they are interested and paying attention to their interaction partners. These characteristics and behaviors are likely to make the interaction partners feel comfortable and therefore more likely to reveal information about themselves, which helps facilitate more accurate judgments. Furthermore, this information can also be detected by observers of the interaction and allow observers to also make more accurate judgments.

Acknowledgments

The author thanks David Funder for his guidance and feedback during all aspects of this project and Lisa Greve for her feedback on the manuscript. The author also thanks previous reviewers for their helpful comments.

References

Adams, H. F. (1927). The good judge of personality. *Journal of Abnormal and Social Psychology, 22*, 172–181.

Allport, G. W. (1937). *The ability to judge people. Personality: A Psychological Interpretation*. New York: Holt, pp 499–522.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*, 518–529.

Anderson, S. M. (1984). Self-knowledge and social inference: II The diagnosticity of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology, 46*, 294–307.

Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85*, 485–501.

Biesanz, J. C., & West, S. G. (2000). Personality coherence: Moderating self-other profile agreement and profile consensus. *Journal of Personality and Social Psychology, 79*, 425–437.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology, 92*, 119–135.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press.

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In W. S. Rholes (Ed.), *Attachment theory and close relationships*. New York: Guilford Press.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance, 18*, 123–149.

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement*. Associates, New Jersey: Lawrence Erlbaum.

Craik, K. H., Ware, A. P., Kamp, J., O'Reilly, C., III, Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. *Journal of Occupational and Organizational Psychology, 75*, 171–193.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin, 52*, 177–193.

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic Accuracy* (pp. 144–168). New York: Guilford Press.

Fletcher, G. J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual difference measure. *Journal of Personality and Social Psychology, 51*, 875–884.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C. (2001a). Accuracy in personality judgment: Research and theory concerning an obvious question. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace*. Washington, DC: American Psychological Association.

Funder, D. C. (2001b). *The Personality Puzzle* (2nd ed.). New York: Norton.

Funder, D. C., Furr, M., & Colvin, C. R. (2000). The Riverside behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality, 68*, 451–489.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8*, 149–162.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" Inventory: Versions 4a and 54*. Berkeley, CA: Institute of Personality Assessment and Research.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102*, 390–402.

Kolar, D.W. (1995). Individual differences in the ability to accurately judge the personality characteristics of others. Unpublished doctoral dissertation, University of California, Riverside.

Letzring, T. D., Block, J., & Funder, D. C. (2005). Ego-control and ego-resiliency: Generalization of self-report scales based on personality descriptions from self, acquaintances, and clinicians. *Journal of Research in Personality, 39*, 395–422.

Letzring, T.D., Greve, L.A., & Funder, D.C. (2005). Behaviors of Accurate Judges of Personality. Poster presented at the annual meeting of the Society for Personality and Social Psychology, New Orleans, LA.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Quantity and quality of available information affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91*, 111–123.

Markey, P. M., & Kurtz, J. E. (2006). Increasing acquaintanceship and complementarity of behavioral styles and personality traits among college roommates. *Personality and Social Psychology Bulletin, 32*, 907–916.

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*, 590.

Raskin, R. N., & Hall, C. S. (1981). The Narcissistic personality inventory: Alternate form reliability and further evidence of construct validity. *Journal of Personality Assessment, 45*, 159–162.

Raskin, R. N., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*, 890–902.

Reimer, H.M., Greve, L.A., & Funder, D.C. (2006). The social wisdom of the attributionally complex. Paper presented at the Western Psychological Association, Palm Springs, CA.

Riggio, R. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology, 51*, 649–660.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. New York, NY: Cambridge University Press.

Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*, 1069–1081.

Soldz, S., Budman, S., Demby, A., & Merry, J. (1995). A short form of the inventory of interpersonal problems circumplex scales. *Assessment, 2*, 53–63.

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91*, 457–477.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52*, 1–23.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500–517.

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34*, 397–423.

Tracey, T. J. G. (2004). Levels of interpersonal complementarity: A simplex representation. *Personality and Social Psychology Bulletin, 30*, 1211–1225.

Vazire, S., & Funder, D. C. (2006). Impulsivity and the self-defeating behavior of narcissists. *Personality and Social Psychology Review, 10*, 154–165.

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology, 4*, 42–57.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality, 35*, 238–246.

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*, 221–229.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobstrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.

Wonderlic, Inc. (1999). Wonderlic Personnel Test, form A. Wonderlic Personnel Test & Scholastic Level Exam User's Manual. Libertyville, IL.