The Employment Interview via the Telephone: Are We Sacrificing Accurate Personality Judgments for Cost Efficiency?

Melinda C. Blackman

California State University, Fullerton

Published online March 26, 2002

The goal of the study was to determine whether impoverished personality judgments of job candidates would be yielded by interviewers who conduct telephone interviews in comparison to face-to-face interviews due to the telephone interviews' inherent lack of crucial nonverbal communication. Participants in this study conducted mock job interviews using either the face-to-face format or the telephone format, while their behavior was coded by independent raters. For each participant, self-ratings of job-related personality traits were obtained from the applicant, while ratings of the applicant's personality were obtained from the interviewer and a peer of the applicant. It was found that the average self-interviewer and peer-interviewer agreement correlations, which were used as the criteria for accuracy, were significantly greater when the interviewer implemented the face-to-face interview method, thus supporting the hypothesis. An item analysis further supported the hypothesis in that face-to-face interviewers were found to rate the candidates significantly higher/more favorably on those traits that are typically conveyed via nonverbal communication.   © 2002 Elsevier Science (USA)

Numerous organizations conduct interviews with job candidates via the telephone due to the cost-efficient nature of this personnel-screening device. But can the reliance on this method be hazardous to the health and growth of an organization? The goal of this research was to determine whether interviewers who used the telephone conference call interview were sacrificing accurate personality judgments of job candidates' work-related traits for cost efficiency.

Employers see many benefits to using the telephone interview as a screening device. For instance, the most prominent benefit is that organizations that use the conference call method no longer need to compensate job candidates and their recruiters for travel and time-related costs. Some organizations may prefer to use this interview format because it helps to eliminate

Address correspondence and reprint requests to Melinda Blackman, Ph.D., Department of Psychology, California State University, Fullerton, CA 92834. E-mail: mblackman@fullerton.edu.

the use of superficial information in the judgment process that can frequently lead to incorrect judgments about the candidate. The faceless telephone conference call interview may allow the interview to maintain it sense of structuredness, more so than the face-to-face format, because visual cues such as physical appearance will not bias the judgment process. An additional advantage to implementing this format is that the use of rating forms can be efficiently completed by the interviewer during the conversation because the candidate will be unaware.

Organizations should be cautioned that there could be potential drawbacks to this interview format. Behavioral differences may arise from the interviewer and the interviewee that differentiate this format from the face-to-face format. For instance, Fletcher (1997) found that candidates report being ill-prepared for the experience and not being in ''interview mode,'' as Fletcher put it. In addition, Silvester, Anderson, Haddleton, Cunningham, and Gibb (2000) found that ratings of candidates made in telephone interviews were harsher than ratings made during face-to-face interviews. The authors attributed this to several possibilities such as the candidates not being able to sufficiently ''psych'' themselves up for the interviews while sitting at home wearing blue jeans and T-shirts. Another potential cause for these low ratings could be due to the many interruptions or intrusions that the candidates might experience during the interviews at their homes (e.g., call waiting, children walking through the room).

Another potential disadvantage is that the telephone interview may be significantly briefer in length than a face-to-face interview. Lack of personal contact may cause the telephone interview to be briefer than a face-to-face interview. Each party during a telephone interview gets less feedback on the other's reaction to what has been said, and this may ultimately result in a briefer interview. Briefer interviews would undoubtedly result in less information about the job candidate being revealed and used in the judgment process, resulting in a less accurate personality judgment (Blackman, in press; Funder, 1995).

An additional concern, of primary interest to this article, is that an employer that relies on the telephone conference call might not be in the best position to accurately judge the job candidate's work-related personality traits as compared to conducting a face-to-face interview. The conference call's inherent lack of revealing nonverbal feedback may put the interviewer at a significant disadvantage, for instance, in accurately determining whether the job candidate has personality attributes similar to those of the successful incumbent or in determining whether the individual is mentally well-adjusted and not prone to counterproductive or volatile behavior (Blackman & Funder, in press). There is no doubt that personality characteristics play an important part in job performance for some occupations (Barrick & Mount, 1991; Hough, Dunnette, Eaton, & Kamp, 1990; Tett, Jackson & Rothstein,

1991). For instance, the work of a child care provider entails a patient and nurturing personality style. On the other hand, an accountant who must achieve numeric accuracy can do so only through being meticulous and conscientious. Because detecting and accurately judging many personality traits involves decoding the target's nonverbal behavior to some degree, the lack of nonverbal behavior in the telephone interview may produce impoverished personality judgments of the job candidate. In a similar vein, the quality and quantity of information that a judge of personality has available to him or her about the target person has been shown to moderate the accuracy of personality judgments (Blackman & Funder, 1998; Funder, 1995). Funder's (1995) research would suggest that this lack of nonverbal communication would yield a lower quality of information and fewer behavioral clues about the job candidate to the judge of personality, thus obstructing the path to making an accurate personality judgment. An investigation is needed to determine whether these inferences are true and interviewers are sacrificing accurate personality judgments of job candidates for cost efficiency.

Previous Research Findings

A literature review revealed that few studies have *directly* attempted to explore the validity of the telephone conference call in the arena of employment interviews. Studies in the past have touched on this issue in an indirect manner. Some studies, for instance, have examined the effect of presentation mode (live vs videotape vs audiotape) on judgments of personality or job performance predictions; however, the judges were instructed to listen to or view the interviews passively instead of conducting them (Motowidlo & Burnett, 1995; Washburn & Hakel, 1973; Wenker, Wegener, & Hart, 1996). This passive observation method undoubtedly removes an element of realism to the study because it is rare that a passive observer to a job interview would be responsible for rating the candidate's future job performance and personality traits. Passive observation also eliminates the ''immediacy'' dimension that develops when an interviewer and job candidate interact directly. This dimension involves greater physical proximity and/or greater perceptual availability of the two persons and significantly affects interview impressions and subsequent decisions (Imada & Hakel, 1977). Due to the immediacy dimension, Imada and Hakel (1977) warned researchers to not assume an equivalence in live and videotaped interviews. Blackman and Funder's (1995) research supports this assertion. They found that, although judges who passively view an interview achieve levels of accuracy similar to those of judges who actively participate in the interview, the traits that each set of judges assess accurately are very different. Blackman and Funder attributed this difference to the differing vantage points of the judges.

Much of the existing literature that examines interviewers' judgments when exposed to an audiotaped interview versus a videotaped interview does

not focus on assessing the accuracy of these judgments. The majority of these studies are focused more on the favorability and range of the ratings that are given by the interviewers. For instance, it has been found that more favorable personality traits are assigned to the interviewee under live conditions or when visual cues are present. This effect is especially strong under conditions of increased use of gestures, eye contact, and smiling (Washburn & Hakel, 1973; Wenker et al., 1996). However, much of the literature on effects of visual cues on interviewers' judgments is strongly laced with the assumption that such cues are superficial indicators that serve only to fool interviewers into evaluating applicants more favorably and injecting error into the interviewers' judgments.

After reviewing the relevant literature, it was decided that a study that implemented interviews via the telephone, in which the judge of personality was also the interviewer instead of a passive observer, needed to be conducted. It was also deemed necessary that this study must approximate as close as possible the realism of the telephone conference call interview and the face-to-face format while being in a controlled environment.

The current study used simulated job interviews to determine which interview format, the face-to-face or the telephone conference call, enhanced the chances of the interviewer making an accurate judgment concerning the job candidate's job-related personality attributes. Both interview formats consisted of structured interviews. For instance, all applicants were asked a set of standardized questions that were based on a job analysis for an actual position within a university's psychology department. The interviewers were not allowed to ask the applicants follow-up or probe questions. The lengths of the interviews were made to be as equivalent as possible. The trained interviewers also used a standardized rating form to rate the applicants.

For the current study, two different criteria were used to assess the accuracy of the personality judgment made by the interviewer concerning the job candidate's personality. These criteria were self-interviewer and peer-interviewer agreement. The behavior of the interviewer and the job candidate was also coded to detect potential variables that might mediate the effect of interview format on the accuracy of personality judgments.

METHOD

Participants

A total of 120 (66 females and 54 males) undergraduate students at a 4-year university participated in the study. Each student received course credit for participating in the study.

Procedure

The Face-to-Face Format

On arriving at the study site, the participants were assigned to same-sex dyads. The experimenter then confirmed that the members of each dyad were not previously acquainted. Those

participants who were previously acquainted were assigned to new partners. At this time, the experimenter randomly assigned one of the participants in each dyad to take on the role of ''interviewer'' and the other participant to be the ''job candidate.'' The participants were randomly assigned to their roles by using an alternating strategy based on their arrival time to the study site. For instance, for the first dyad, the participant who arrived first to the study was assigned the role of interviewer and the participant who arrived last was assigned the role of job candidate. For the next dyad, the assignments were altered; the participant arriving first to the study was assigned the role of job candidate, while the other participant was assigned the role of interviewer. At this point in the study, the interviewers and job candidates were escorted to different rooms in which the experimenter explained the nature of the study to these individuals separately.

The interviewer. Those participants who were assigned to the role of interviewer were told that they would be conducting a mock job interview. They were informed that they would be interviewing the other participant for a hypothetical student-clerical position in the psychology department office. A student-clerical position was selected for the mock interview because this position was viewed as a job for which all participants could knowledgeably partake in the interview process. The interviewers, as well as the job applicants, were then given a job description. The interviewers received examples of critical incidents of the desired behavior and personality traits that the future employee should possess. They were then told that the ideal job candidate would be responsible for photocopying exams, answering telephones, assisting students at the counter with paperwork, answering questions with regard to department and university policies, and doing some light word processing. It was emphasized that the ideal candidate would be an individual who was reliable and dependable and who possessed integrity and good interpersonal skills. The interviewers were also informed that they would be judging the job applicants' work-related personality characteristics after the interviews while a copy of the rating form was placed in front of them. The experimenter emphasized to each interviewer that obtaining relevant information with regard to the candidate's personality was the focus of the interview.

The interviewers were then given a sheet of paper with 11 preset questions (e.g., ''How would your peers in your current or past workplace describe you?'') that they were required to ask the candidates (see Appendix). The interview questions were developed from a task-based job analysis that was conducted for a similar position within the university's psychology department. The questions were structured to elicit the candidate's work-related personality attributes relevant to the position. The interviewers were also told that they could ask only the questions given to them and that they were not allowed to ask any follow-up or probe questions. The interviewers were also given a short training session to familiarize them with the inventory traits and the rating method. These measures were taken to ensure that the interview format adhered to the strict operationalization of a structured interview.

The job candidate. Before the nature of the study was revealed to the participants, the experimenter asked the job candidates to complete the California Q-set ratings inventory on their own personality characteristics (Bem & Funder, 1978; Block, 1978).[1] The California Q-set ratings consist of 100 personality characteristics (e.g., ''is assertive,'' ''is a genuinely dependable person'') that the respondent must rate on a 5-point Likert scale as to how characteristic or uncharacteristic the trait is of the judged person. The five categories ranged from *highly characteristic* to *not at all characteristic.* The California Q-set is well-suited for use

---

[1] A slight variation, however, was made to the California Q-set. The item content of the Q-set was retained, while a 5-point Likert scale replaced the traditional forced-choice distribution method used to assess the target's personality. Furr and Funder (1999) showed that there is not a clear significant advantage to the forced-choice Q-sort method in comparison to the Likert rating method.

by lay judges of personality because it represents the categorical breadth of personality attributes that lay judges use naturally when judging others in day-to-day interactions.

Precautionary measures were taken to minimize the pressure that some participants might feel to ''fake good'' on their self-report measures. For instance, the participants were not told until after completing their self-report personality inventory that they would be participating in a mock job interview so as to help reduce a potential social desirability effect. The participants were also assured that their responses would remain anonymous and locked in a file cabinet to which only the principal investigator would have access.

After completing the personality inventory, the job applicants were told that they would be interviewing for a hypothetical position as a student clerk in the psychology department office. The job applicants then received a copy of the job description to look over.

The interviewer and job applicant for each dyad were then reunited in the same room and introduced to each other. Both participants were told that a research assistant would be sitting in on the interview to record the questions that were asked. The interviewer then proceeded with the interview for as close to 10 min as possible. The interviewer then completed the California Q-set inventory on the applicant's work-related personality traits. After completing the inventory, the interviewer was also debriefed and informed as to the purpose of the study. All participants were told that they could contact the experimenter in the future for a briefing on the results of the study.

The Telephone Interview

The telephone interview format condition was conducted similarly to the face-to-face format except that the interviewer and job applicant who were randomly assigned to this condition were told to arrive at two separate locations within the university. The separate locations ensured that the interviewer and job applicant did not meet face-to-face and present a confound to the study. The research assistants communicated via telephone prior to the interview to inform each other which participant arrived first for the random assignment of the roles. The interviewer was told that he or she would be conducting an interview via the telephone. The interviewer was instructed to dial a specific extension and that the job applicant would be waiting for the call.

Peer Report

After the interview, the job applicant was separately debriefed. At this time, the experimenter handed the job applicant (in each condition) an additional copy of the California Q-set and asked that he or she give the inventory to a friend, who was well acquainted with the job applicant, to complete on the job applicant's personality. The experimenter then explained to the job applicant that, on returning the peer report, he or she would receive course credit. A peer report was solicited from each applicant to provide an additional validation measure of the accuracy of the job applicant's self-report. The applicant was instructed to return the peer report in a sealed envelope with the peer's signature across the seal of the envelope to ensure the confidentiality of the ratings.

The Personality Inventory

The 100 Q-set traits were individually rated, with regard to the student-clerical position, as traits that were ''job related'' or ''not job related.''[2] Three independent judges were asked

---

[2] The practice of using only the 47 relevant Q-set traits for this study was established to remain consistent with the author's previous study (Blackman, in press), which used a similar methodology.

to read the job analysis for the student-clerical position. The judges were then told to determine whether the trait was ''job related'' or ''not job related'' for each Q-set trait. It should be noted that the judges who were chosen to make the ratings had extensive experience in the field of interviewing and human resources. If judges deemed the trait as job related, then they coded the trait as a ''+''. Traits deemed as not job related were coded as ''−'', while traits deemed as neutral to the position were coded as ''0''. The judges first rated the items independently. At this point, the interrater reliability, assessed using Cohen's kappa, was $\kappa = .79$. The three lists of ratings were then compared by the experimenter for discrepancies. The judges were then told to reconcile any discrepancies in the ratings by discussing them until all three judges were in agreement about the coding of a particular trait. After the discrepancies were reconciled, 47 of the 100 Q-set traits remained and were used in the analyses (see Table 1).

Coded Behavior

To detect potential mediating variables that might be inherent with respect to the different interview formats, student raters were trained to code the behavior of the job applicant and interviewer during the interview. The following variables were coded by the raters:

1. The length of the interview in minutes
2. The number of questions asked by the interviewer
3. The number of follow-up questions asked. (By measuring this variable, it helped to ensure that the interviewer did not ask follow-up or probe questions as instructed.)
4. A word-for-word transcription of the questions asked
5. The tone of the interviewer and job candidate on a 5-point scale ranging from *formal* to *informal*
6. An estimated percentage of the interview time that the job candidate spent talking
7. The expressiveness (facially and gesturally) of the job candidate on a 5-point scale ranging from *very expressive* to *not at all expressive*

Analyses

Data analyses focused on the way in which the independent variable, interview format, affected the dependent variable, accuracy (self–other agreement) in personality judgment. That is, to the degree a lay judge's assessment of a target's personality attributes matches that target person's own self-assessment, the lay judge is assumed to be accurate. The analyses were conducted using the 47 job-related Q-set traits.

Self–other agreement was calculated by correlating each interviewer's Q-ratings with the self-description of his or her target applicant. The agreement between each perceiver and his or her target was characterized by a profile correlation coefficient calculated across the 47 items in the Q-set. These correlations were then treated as agreement scores and averaged across all interviewer–applicant pairs for each interview format type. These correlations were calculated again, but this time using the peer-report in place of the applicant's self-report Q-ratings.

RESULTS

An analysis of variance (ANOVA) was run to determine whether there was a significant interaction or main effect for interview format and gender. A significant main effect was found for interview format, $F(1, 56) = 7.83$, $p = .01$, $\omega^2 = .25$. The average self–other agreement score for the face-to-face interview format was .51 ($n = 33$), while the average agreement score

TABLE 1
Job-Related Personality Q-Items ($N = 47$)

Number	Q-Item
Q91	Is power oriented
Q62	Tends to be rebellious
Q68	Is basically anxious
Q65	Stretches the limits
Q53	Has little self-control
Q45	Has a brittle ego-defense system
Q23	Is extrapunitive
Q26	Is productive
Q70	Behaves in an ethically consistent way
Q98	Is verbally fluent
Q77	Appears forthright in dealing with others
Q6	Is fastidious
Q38	Has hostility toward others
Q84	Is cheerful
Q94	Expresses hostile feelings directly
Q78	Is self-pitying
Q39	Has unconventional thought processes
Q13	Is thin-skinned
Q41	Is moralistic
Q22	Feels a lack of meaning in life
Q19	Seeks reassurance from others
Q86	Tends to deny unpleasant thoughts
Q87	Interprets situations in complicated ways
Q48	Keeps people at a distance
Q71	Has high aspiration level for self
Q25	Tends toward overcontrol of needs
Q92	Appears socially at ease
Q49	Is distrustful
Q2	Is dependable
Q34	Is irritable
Q17	Behaves in a considerate way
Q37	Is deceitful
Q12	Is self-defensive
Q97	Is emotionally bland
Q14	Is submissive
Q35	Has warmth
Q36	Is subtly negativistic
Q61	Exploits dependency in others
Q50	Is unpredictable
Q40	Is fearful
Q8	Appears to have a high degree of intellectual capacity
Q30	Withdraws in the face of adversity
Q28	Tends to arouse liking
Q27	Shows condescending behavior
Q47	Feels guilty
Q83	Able to see to the heart of important matters
Q74	Feels satisfied with self

Note. Item content is abbreviated.

for the telephone format was significantly less at .32 ($n = 27$). The face-to-face interview format appeared to yield significantly more accurate personality judgments than did the telephone format.

A significant main effect did appear for the gender of the dyad, $F(1, 56) = 5.43$, $p = .02$, $\omega^2 = .25$, with female dyads manifesting higher levels of interjudge agreement. The average agreement score for male dyads was .32 ($n = 27$), while the average agreement score for female dyads was .48, ($n = 33$). No significant interaction occurred for gender and interview format, $F(1, 56) = .00$, *ns.*

Peer Reports

With regard to peer–interviewer agreement, a significant difference was also found between the two interview formats, $t(33) = -2.05$, $p = .05$, $\omega^2 = .14$. The average peer–interviewer agreement score for the face-to-face interview format was .56, while the average agreement score for the telephone format was significantly less at .45. The face-to-face interview format again appears to yield significantly more accurate personality judgments than the telephone format when peer reports are used in place of self-reports. It should be noted that the peer completing the report was not necessarily of the same sex as the applicant and that not all peers of the job applicants completed the inventory. A breakdown of the average agreement scores by interview format can be found in Table 2.

Item Analysis

To provide support for the hypothesis that a lack of nonverbal feedback during a telephone interview contributes to impoverished personality judgments, an analysis at the item level was conducted. Three trained raters were asked to rank the top 10% of the Q-items (or roughly 4 of the 47 work-related traits) that they perceived to be job desirable interpersonal skills traits that, during an interaction, have the highest likelihood of being conveyed with nonverbal communication. The top 10% of traits was chosen so that the raters would be forced to make large discretions in their ratings. The rating process was conducted in a fashion similar to that when the 47 of 100

TABLE 2
Average Agreement Scores by Interview Format

Format	Face-to-face format	Telephone format	Male dyads	Female dyads
Self–interviewer	.51$_a$ (33)	.32$_b$ (27)	.32$_e$ (27)	.48$_f$ (33)
Peer–Interviewer	.56^c (18)	.45$_d$ (17)		

Note. Between adjacent columns, means without a common subscript differ significantly ($p < .05$). n's are in parentheses.

work-related Q-traits were chosen. The interrater reliability, prior to reconciling rating discrepancies, that was assessed using Cohen's kappa, was $\kappa = .69$. The following 4 Q-items were judged as desirable interpersonal skills traits: Q#92 (has social poise and presence/appears socially at ease), Q#84 (is cheerful), Q#35 (has warmth/the capacity for close relationships/is compassionate), and Q#17 (behaves in a sympathetic or considerate manner). It was found that the mean interviewer rating (on the 5-point scale) for these combined items, where each item's mean rating for each interview format was based on an n of 33 and 27 for the face-to-face and telephone interview formats, respectively, was significantly higher when the face-to-face interview was conducted, $t(3) = 6.6$, $p = .00$, $\omega^2 = .18$ (face-to-face $x = 3.9$, telephone $x = 3.3$). Judges in the face-to-face interview perceived their job applicants to be significantly more cheerful, warm, considerate, and socially at ease than did interviewers in the telephone condition. These ratings by the interviewers of the applicants support others' previous research findings as well (Silvester et al., 2000; Washburn & Hakel, 1973; Wenker et al., 1996). Not only did these interviewers perceive their job candidates to be more cheerful, warm, and at ease, but they reached higher levels of interjudge agreement in making these assessments than did the interviewers in the telephone condition. The average self–other agreement score with regard to the four desirable traits for the face-to-face condition was .20, whereas in the telephone interview format it was only .07, although the difference between these scores did not quite reach significance ($p = .07$). It can be inferred from these results that personality traits that are typically conveyed via nonverbal behavior will be judged with lower levels of accuracy when a telephone interview is implemented. Because the assessment of many personality traits involves decoding some degree of nonverbal behavior, one can see that the telephone interview puts judges of personality at a disadvantage.

Coded Behavior

With respect to the coded behavior of the job applicants and interviewers, a notable difference was found between the conditions: Significantly more follow-up questions were asked in the face-to-face condition than in the telephone condition, $t(55) = 2.5$, $p = .03$, $\omega^2 = .64$.[3] This finding is unusual in that the interviewers were specifically instructed not to ask probe or follow-up questions in either condition. The number of follow-up questions was then correlated with the interviewer's degree of accuracy (self–interviewer agreement), while the interview format was held constant. A significant relationship was found between the number of follow-up questions asked and the interviewer's degree of accuracy ($r = .25$, $p = .05$, $n = 60$). These results

---

[3] Of the 60 interviews, 3 were not coded for behavioral differences.

suggest that follow-up questions play a role in the degree of accuracy attained by the interviewer.

Another prominent finding was that job candidates in the face-to-face condition were rated equally behaviorally and gesturally expressive as candidates in the telephone condition, $t(55) = .39, p = .70$. One would guess that during a telephone interview, job candidates would be less likely to gesture and use facial expressions without the interviewers present and that this might influence their voice tone; however, this was not the case. The tone of both the interviewers and the job applicants was also not found to be significantly more informal/formal when the telephone format was used, $t(55) = .56, p = .58$, and $t(55) = 1.4, p = .17$, respectively. This finding directly contradicts Fletcher's (1997) assertion that job candidates who interview via the telephone are more informal with their responses because they have trouble psyching themselves up for the interviews.

The average lengths of the face-to-face and telephone interviews were found to be relatively similar at 6.8 and 6.0 min, respectively, $t(55) = -1.7, p = .09$. Job candidates in the face-to-face and telephone formats talked for relatively equal portions of the interviews, $t(55) = -.39, p = .70$.

DISCUSSION

This study examined the telephone conference call interview versus the face-to-face interview to determine which format yielded more accurate personality judgments concerning the job candidate. It was found that when conducting simulated job interviews, the face-to-face format produced significantly higher levels of self–interviewer and peer–interviewer agreement about the job applicant's job-related personality traits than did the telephone conference call interview. There are two viable explanations for this finding. The first hypothesis suggests that the nonverbal feedback that is inherent to the face-to-face format produces a larger quantity and a richer quality of revealing behaviors about the job candidate than does the telephone format. This rich quality of data would then undoubtedly produce more accurate personality ratings. An item analysis of Q-set items helped to add evidence to this assertion. It was found that judges in the face-to-face format rated applicants as more warm, friendly, compassionate, and socially at ease than did judges in the telephone condition. These are traits that are typically assessed by analyzing the target's nonverbal behavior, so it follows logically with the study's finding that the judges in the face-to-face condition obtained higher levels of accuracy about these traits than did the interviewers in the telephone condition.

It is also believed that telephone interviewers, who have the tendency to inaccurately rate candidates as less warm and cheerful, may be more likely to make incorrect assumptions about the candidates' other related traits than are interviewers who implement a face-to-face format. The personality char-

acteristics of warm and cold are considered to be central traits and have the power to color the way in which interviewers perceive candidates' other related traits. It just may be that, at times, candidates who are inaccurately perceived to be less warm or cold are also incorrectly assumed to be "emotionally bland" or to be individuals who "keep others at a distance," ultimately biasing the overall personality judgments of the candidates. A study conducted by Rinella, Ferguson, and Sager (1970), in accord with Solomon Asch's hypothesis on the centrality of the cold–warm construct, found evidence that supports this assertion by using blindfolded interviewers. A future follow-up study is recommended to test the likelihood of this strong assertion. Because the assessment of many personality traits involves decoding some degree of nonverbal behavior, one can see that the telephone interview puts judges of personality at a consistent disadvantage.

The second hypothesis that could account for the study's main finding deals with specific behavioral differences found between the two interview formats. The behavior of the interviewer and the job candidate was coded to examine behavioral differences between the distinct interview formats so as to uncover variables that may have the potential to mediate the effect of interview format on the accuracy of personality judgments. The examination of the coded behavioral variables revealed that the interviewers reacted differently to the two formats, whereas the job candidates did not behave differently. The prominent behavioral differences between the telephone and face-to-face formats was that the interviewers asked significantly more follow-up or probe questions in the face-to-face format, even though they were instructed not to ask such questions. (It should also be noted, however, that there was not a significant difference between the lengths/times of the interview formats.) It may be the close proximity or "immediacy" of the interviewers to the applicants that made the interviewers feel more comfortable to ask follow-up questions in the face-to-face format (Imada & Hakel, 1977). The use of probe or follow-up questioning may naturally increase the quantity and quality of information that the candidates reveal about themselves. This assertion is supported by a study conducted by Blackman (in press). The study found that interviewers who used an unstructured interview format and were encouraged to ask the job applicants follow-up questions were significantly more accurate in assessing the job-related personality characteristics of the applicants than were interviewers who implemented a structured interview with no follow-up or probe questions.

It is also theorized that use of follow-up questioning elicits a higher/clearer quality of response from the applicant. This added information would undoubtedly help the interviewer garner a higher quality of clues about the candidate's personality so as to make a more informed and accurate judgment. Another possible reason as to why interviewers in the face-to-face format asked follow-up questions might be due to viewing simultaneous dis-

crepant verbal and nonverbal behavior from the candidates. To make sense of this discrepancy, the interviewers might have felt the need for clarification that would warrant follow-up questions.

As for the other coded variables measured—interview length, behavioral expressiveness of the candidate, and tone of voice—no prominent differences were found. One would suspect that interviews conducted via the telephone, with lack of personal contact, would be significantly briefer and evoke more formal tones and less expressiveness from the job candidates; however, this was not the case. The brevity of the interviews in general, 6 to 7 min on average, can probably be attributed to the structured format that interviewers were required to adhere to with no follow-up questions or small talk. This structured format controlled for uniformity in the interview procedures across the interview formats.

A robust gender difference was also found in this study. Female dyads manifested higher levels of interjudge agreement, as compared to male dyads, regardless of the interview format. To more fully understand this finding, a follow-up study that systematically examines self–other agreement in both same-sex and opposite-sex dyads should be conducted. At this point, it is unclear as to whether it is female judges in general who are professing superior judgment skills or whether female targets are easier to judge than male targets.

In summary, the findings of this study suggest that the face-to-face format produces significantly more accurate personality judgments than does the telephone format. An analysis of the coded behavior suggests that part of this effect can be attributed to the interviewers' use of significantly more follow-up questions when the face-to-face method is implemented, which may be prompted by the specific interview format that is used. More follow-up questions would undoubtedly elicit more information and behavioral cues from the job candidates from which to garner and make accurate personality judgments (Funder, 1995). The item analysis from the study also provides further evidence that the face-to-face format's superiority with regard to personality judgments is probably due to the presence of crucial nonverbal communication that is given off by the job candidates.

It should be acknowledged that by no means should self–other and peer–other agreement serve as the only criteria for accuracy. Obtaining strong self–other and peer–other agreement correlations is only the first step in determining that the face-to-face interview might be a more optimal method than the telephone format for accurately assessing a target person's personality attributes. Other forms of criteria need to be used as well to establish the assertions that this study is purporting. For instance, personality judgments made by the job applicant's supervisor and peers after the applicant has performed the job for several months could serve as a form of predictive validity. Also, behavioral measures such as the number of customer grievances

filed about the recently hired employee could be used as an indication of interpersonal skills and conscientiousness. As one can see, there seems to be no practical way in which to demonstrate directly whether the basis of self–other agreement is accurate judgment. Rather, it is only possible to gather multiple forms of criteria (e.g., critical incidents, peer/supervisor ratings) that support the findings yielded by the self–other and peer–other agreement correlations to establish that accurate judgment is being achieved.

It is recommended that a field study be the next logical extension to this study; several forms of predictive criteria can be gathered in this environment, and interviewers with extensive training can be used. Due to the fact that the current study's interviewers did not have extensive training in the art of conducting interviews, one should probably generalize the results of this study to only small businesses. Interviewers in these organizations probably have not been given formal training on how to conduct interviews, and they do not conduct numerous interviews on a daily basis.

It is important to keep in mind that there may be other processes or variables that were not coded in this study and that may potentially mediate the effect of interview format on personality judgment (e.g., vocal attractiveness [Zuckerman, Miyake, & Elkin, 1995]). The coded variables were chosen for the study based on suggestions from colleagues familiar with the literature. It is quite possible that, after conducting this study in the field, other mediating variables might become more apparent.

For organizations that are forced to continue to rely on telephone interviewing, the following advice should be taken into consideration. It is recommended that the candidate be briefed properly prior to the interview. The candidate should be told to find an appropriate place to take the call and what to expect in terms of the format of the interview. For example, the candidate should be informed if the interviewer will be writing notes during the interview. To help promote a state of equivalence between the telephone and face-to-face formats, interviewers also should be encouraged to treat the telephone interview with all the seriousness of a face-to-face interview and be made aware of how their behavior can produce a state of inequality between the two formats.

APPENDIX

Structured Interview Questions

1. If you were hired for this position, what would I see after a month of working with you as your strongest work-related trait?
2. If you were hired for this position, what would I see after a month of working with you as your weakest work-related trait?
3. How would your current or last employer describe you?
4. How would your peers describe you?

5. Think of your least effective manager/supervisor; tell me why [he or she was] ineffective.

6. Describe one strength of this manager.

7. Suppose that an individual came to the office and claimed to be a representative of the Disabled Students Office without presenting any identification. This individual then requests to pick up an exam for a disabled student to take at the Disabled Students Center and refuses to leave until you present him/her with the test. At this point, you notice that the office manager has taken a lunch break. How would you react to this individual? What course of action would you take?

8. Suppose that you were employed in the psychology department office, and you were asked by the office manager to [photocopy] the exams for the course in which you were currently enrolled; what would you do?

9. If the office manager was to return a job that you completed in this office for you to redo, which of the following activities from the job description would it most likely be and why?

10. With regard to your previous job performance evaluations, what personality characteristic were your previous employers most likely to compliment?

11. Describe yourself to me.

REFERENCES

Barrick, M., & Mount, M. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology,* **44,** 1–26.

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

Blackman, M. C. (in press). Personality judgment and the utility of the unstructured employment interview. *Basic and Applied Social Psychology.*

Blackman, M. C., & Funder, D. C. (1995). *The effects of observation vs. participation on the accuracy of personality judgments.* Paper presented at the annual conference of the Western Psychological Association, Los Angeles.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on the accuracy and consensus of personality judgments. *Journal of Experimental Social Psychology,* **34,** *164–181.*

Blackman, M. C., & Funder, D. C. (in press). Effective interview practices for accurately predicting job candidates' counterproductive traits. *International Journal of Selection and Assessment.*

Block, J. (1978). *The Q-sort method in personality assessment and psychiatric research.* Palo Alto, CA: Consulting Psychologists Press.

Fletcher, C. (1997). Just how effective is a telephone interview? *People Management,* **3**(13), 2–3.

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

Furr, R. M., & Funder, D. C. (1999). *A comparison of Q-sort and Likert rating methods of*

personality assessment. Paper presented at the annual conference of the Western Psychological Association, Irvine, CA.

Hough, L. M., Dunnette, M. D., Eaton, N. K., & Kamp, J. D. (1990). Criterion-related validities of personality constructs and the effect of response distortions on those validities. *Journal of Applied Psychology,* **75,** 581–595.

Imada, A., & Hakel, M. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology,* **62,** 295–300.

Motowidlo, S., & Burnett, J. (1995). Aural and visual sources of validity in structured employment interviews. *Organizational Behavior and Human Decision Processes,* **61,** 239–249.

Rinella, S., Ferguson, L., & Sager, E. (1970). Personality-impression formation as a function of visual cues and set. *Perceptual & Motor Skills,* **31,** 427–430.

Silvester, J., Anderson, N., Haddleton, E., Cunningham, S., & Gibb, A. (2000). A cross-modal comparison of telephone and face-to-face selection interviews in graduate recruitment. *International Journal of Selection and Assessment,* **8**(1), 16–21.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performances: A meta-analytic review. *Personnel Psychology,* **44,** 703–741.

Washburn, P., & Hakel, M. (1973). Visual cues and verbal content as influences on impressions formed after simulated employment interviews. *Journal of Applied Psychology,* **58,** 137–141.

Wenker, R., Wegener, J., & Hart, K. (1996). The impact of presentation mode and disfluency on judgments about speakers. *Journal of Fluency Disorders,* **21,** 147–159.

Zuckerman, M., Miyake, K., & Elkin, C. (1995). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality,* **29,** 253–272.