9-2005

The importance of traits and abilities in supervisors' hirability decisions as a function of method of assessment

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Scott HIGHHOUSE
Bowling Green State University

Wilfried DE CORTE
Ghent University

DOI: https://doi.org/10.1348/096317905X26093

Published in Journal of Occupational and
Organizational Psychology, 2005 September,
Volume 78, Pages 453-470
http://doi.org/10.1348/096317905X26093

The importance of traits and abilities in supervisors' hirability decisions as a function of method of assessment

Filip Lievens[1]*, Scott Highhouse[2] and Wilfried De Corte[1]
[1] Ghent University, Belgium
[2] Bowling Green State University, Ohio, USA

Past research on the importance of traits and abilities in supervisors' hirability decisions has ignored the influence of the selection method used to derive information about these traits and abilities. In this study, experienced retail store supervisors ($N = 163$) rated job applicant profiles that were described on the Big Five and General Mental Ability (GMA) personality dimensions. Contrary to past studies, the supervisors were also informed about the method of assessment used (paper-and-pencil test vs. unstructured interview). Hierarchical linear modelling analyses showed that the importance attached to extraversion and GMA was significantly moderated by the selection method, with extraversion and GMA decreasing in importance when store supervisors knew that scores on extraversion and GMA were derived from a paper-and-pencil test as opposed to from an unstructured interview. Store supervisors with more selection-related experience also attached more importance to GMA. Results are discussed in relation to the practice–science gap and the extant literature on perceptions of selection procedures.

Over the last decade, substantial advancements have been made to understand the constructs underlying personality inventories and cognitive ability tests. To date, there is relative widespread support for the five-factor model as a unifying theoretical framework to study personality. There also exists relative consensus that cognitive abilities have a hierarchical structure and that the highest-order factor, also known as *g* or general mental ability (GMA), typically accounts for more variance than all specific factors together (Jensen, 1998; Ree, Carretta, & Steindl, 2001). These construct-oriented approaches have provided a firm theoretical basis for developing hypotheses about the predictive validity of personality and cognitive ability in work-related settings. In fact, several meta-analyses (Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Hurtz & Donovan, 2000; Salgado, 1997, 2003; Tett, Jackson,

* *Correspondence should be addressed to Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium (e-mail: filip.lievens@ugent.be).*

& Rothstein, 1991) have demonstrated the validity of personality traits as predictors of industrial and occupational outcome criteria, using the five-factor model as a framework to sort traits. Similarly, GMA has been found to be a consistent predictor of job performance across various occupations in both the USA (Schmidt & Hunter, 1998; Schmidt, 2002) and Europe (Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003).

Although important advancements with regard to the underlying structure and predictive validity of personality and cognitive-based predictors have been made, we have less insight into the role that the Big Five dimensions and GMA play in practitioners' hirability decisions because only a limited number of studies have addressed such issues (Dunn, Mount, Barrick, & Ones, 1995; Ones & Viswesvaran, 1999). In these studies, practitioners typically rated the hirability of a series of hypothetical candidate profiles that varied on the Big Five personality traits and GMA. Generally, managers valued these dimensions that previous meta-analyses had identified as being predictive of job performance. For example, Dunn *et al.* (1995) found that across six occupations, GMA and conscientiousness were viewed as the most important attributes related to applicants' hirability decisions.

A significant limitation of these prior studies was that participants were not informed about the selection methods that were used to derive the trait scores. There are several reasons why it is problematic that participants did not receive this contextual information. First, from a conceptual point of view, these prior studies did not acknowledge the distinction between the constructs (e.g. extraversion) and the methods used to measure these constructs (e.g. an unstructured interview and a personality test are both methods that might assess extraversion; Arthur, Day, McNelly, & Edens, 2003; Schmitt & Chan, 1998). Second, from a practical point of view, it is unusual that hiring managers do not know which selection methods were used when they evaluate applicant qualifications and make hiring decisions. Third, as attested by a large body of research, people's reactions and beliefs with respect to selection procedures matter (Ryan & Ployhart, 2000). Therefore, these reactions and beliefs might well affect the importance attached to the trait and ability information generated by these procedures.

Taken together, the aim of this study was to extend prior research by examining the moderating effect of the selection method used on the relative importance of the Big Five and GMA in supervisors' judgments of applicants' overall qualifications. To operationalize selection method, two methods of assessment were chosen, namely a paper-and-pencil test versus an unstructured interview. There were two reasons why we chose these two specific methods. First, both of these methods represent key methods that are widely used (Ryan, McFarland, Baron, & Page, 1999). Second, this study's sample consisted of retail store supervisors. Pilot interviews with these retail store supervisors indicated that they were familiar with these basic assessment methods. Conversely, they were less knowledgeable about more sophisticated selection methods such as structured behaviour description interviews (see also Glode, Truxillo, & Bauer, 2003).

Background

The relative importance of GMA and the Big Five in hirability decisions

As mentioned above, the role that GMA and the Big Five personality dimensions play in the hirability decisions of managers has received a limited amount of research attention (Dunn *et al.*, 1995; Ones & Viswesvaran, 1999). The general aim of prior studies

consisted of examining whether people value those attributes that meta-analytic research has identified as being predictive of job performance. In particular, Dunn *et al.* presented 84 first-line managers with 39 carefully constructed profiles of hypothetical job applicants that were described on GMA and on the Big Five personality factors (emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness). After each profile, they were asked to make a hirability decision. Policy-capturing analysis was used to examine how first-line managers used and weighed GMA and personality characteristics in their hirability decisions. There were two key findings. First, GMA and conscientiousness were viewed as the most important attributes related to applicants' hirability. In other words, first-line managers valued these attributes that previous meta-analyses had identified as being predictive of job performance. Another important finding was that the relative importance attached to the personal attributes was consistent across six occupations, although some minor differences were found. A subsequent study with expatriates generally mirrored these original findings (Ones & Viswesvaran, 1999).

Williams, Munick, Saiz, and FormyDuval (1995) conducted a series of experiments that examined the role that the Big Five dimensions played in evaluations of people in general. In one experiment, agreeableness and extraversion emerged as the most important attributes in judgments about people. Yet, in another experiment, conscientiousness and openness adjectives were the most important Big Five dimensions. A third experiment resolved these conflicting findings. Williams *et al.* (1995) found that the relative importance attached to the Big Five dimensions varied across the context in which these judgments were obtained. People described with conscientiousness and openness adjectives were viewed more positively in the context of working relationships, whereas agreeableness and extraversion were most important in the context of personal relationships.

The Williams *et al.* study makes salient the importance of considering the context in which trait judgments are made. We believe, however, that one must go beyond a recognition of the setting in which the traits are to be employed, to considering how that trait information was derived in the first place. As argued below, any inferences about how people judge applicant information and qualifications must be based on research that considers the source of that information (i.e. the selection method used to derive that information) as an important contextual variable.

The relative importance of GMA and the Big Five as a function of the selection method

Recently, organizational and social psychologists have called for more attention to context variability in research design (Rousseau & Fried, 2001; Wells & Windschitl, 1999), and decision researchers have also criticized the idea that decision makers possess a context-independent utility function (Goldstein & Weber, 1995; Rettinger & Hastie, 2001). Much earlier, Brunswik (1955) argued for greater attention to representative design in psychological experiments. Brunswik lamented the fact that although considerable attention was given to sampling theory as it relates to generalizing across subjects, little attention was given to it as it relates to generalizing across environmental conditions. With regard to managerial judgments of the relative importance of applicant traits, such judgments are likely to occur in an environment where information about the source of that information is available. However, previous

research on managerial judgments of trait importance (Dunn *et al.*, 1995; Ones & Viswesvaran, 1999) has not explicitly included this critical contextual variable in their design.

The present research was motivated by the assumption that managerial judgments of the importance of traits and abilities in applicant profiles would depend greatly on how supervisors perceive the source of the trait and ability information. Research in attitude development and change has long established the importance of source credibility (see Petty & Wegener, 1998, for a review). Applied to this situation, source credibility pertains to people's beliefs of the selection procedure used to gather the trait and ability information. Studies in the field of applicant perceptions (Kravitz, Stinson, & Chavez, 1996; Lievens, De Corte, & Brysse, 2003; Rosse, Miller, & Stecher, 1994; Rynes & Connerley, 1993; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993; Steiner & Gilliland, 1996, 2001) have confirmed that some selection procedures are viewed as more valid than others. Generally, these comparative studies revealed that applicants have great faith in unstructured interviews as procedures to gather trait and ability information. These studies also indicated that applicant perceptions are significantly less favourable when a paper-and-pencil test (i.e. a cognitive ability test and a personality test) is used to garner trait and ability information than when an unstructured interview is used. This preference for selection procedures that favour subjective judgments, intuitive thinking, and personal evaluations (e.g. unstructured interviews) has also been observed among human resource management practitioners and hiring managers (Dipboye, 1994; Johns, 1993; Miller & Rosenbaum, 1997; Shrivastava & Mitroff, 1984; Terpstra & Rozell, 1997; Van der Zee, Bakker, & Bakker, 2002).

In understanding why hiring managers prefer intuitive assessment to more mechanical approaches, it is useful to examine the resistance to mechanical prediction over global judgments of experts observed in the judgment and decision-making literature. The considerable body of literature on the superiority of actuarial judgments over expert judgments has not deterred people from clinging to the belief that humans can become intuitive experts in predicting future behaviour (Kleinmuntz, 1990). Hastie and Dawes (2001) noted that, 'People have great misplaced confidence in their own [and others'] global judgments, a confidence that is strong enough to dismiss an impressive body of research findings and to dominate predictions' (p. 65). Hastie and Dawes suggest that one reason for this overconfidence in global or intuitive judgment is that relying on experts provides a justifiable reason for making an otherwise difficult choice. In other words, relying on the counsel of so-called experts is seen as more socially acceptable than relying on test scores or formulas. Support for this justifiability hypothesis is found in research by Simonson (1989) showing that people choose options for which they can provide the best reasons, even when such decision-making results in highly inconsistent responses.

Drawing on this research base about perceptions of selection procedures and the need to provide justifiable choices, our general expectation was that the selection method used (paper-and-pencil test versus unstructured interview) would moderate the relative importance attached to the trait and ability information. Specifically, regarding information about GMA, we hypothesized that GMA would play a significantly more important role (i.e. relative to other traits) in supervisors' hirability decisions if they knew that applicants' GMA scores were derived from an unstructured interview than from a paper-and-pencil test (Hypothesis 1).

In a similar vein, we expected that information about an applicant's personality would play a more important role in supervisors' hirability decisions when this information was derived from an unstructured interview than from a paper-and-pencil test (i.e. a personality inventory). However, we did not expect this to be the case for all Big Five traits. Specifically, our hypotheses about which personality traits might be susceptible to the effects of information about the selection method were based on theoretical and empirical research on the easiness of trait judgment in social psychology (Funder, 1999; Funder & Colvin, 1988; Funder & Dobroth, 1987; John & Robbins, 1993; Trope, 1986). These studies have consistently shown that not all personality traits are equally observable and detectable in short social interactions (which are comparable to unstructured interviews). In particular, a meta-analysis of Connolly and Viswesvaran (1998) revealed that there was much higher convergence among stranger and self-ratings on traits such as extraversion ($r = .29$) and conscientiousness ($r = .23$) as opposed to traits such as openness to experience ($r = .14$), emotional stability ($r = .05$) or agreeableness ($r = -.01$), indicating that even strangers (e.g. interviewers) who had limited opportunity to observe a target person were able to make relatively accurate judgments about that person's level of extraversion and conscientiousness.

Based on this research on the easiness of trait judgment, we expected that information about the selection method used would affect the relative importance attached to extraversion and conscientiousness. In particular, Hypothesis 2 stated that extraversion would play a significantly more important role (i.e. relative to other traits and abilities) in supervisors' hirability decisions if supervisors knew that applicants' extraversion scores were derived from an unstructured interview instead of from a paper-and-pencil test. Hypothesis 3 stated that conscientiousness would play a significantly more important role (i.e. relative to other traits and abilities) in supervisors' hirability decisions if supervisors knew that applicants' conscientiousness scores were derived from an unstructured interview instead of from a paper-and-pencil test.

Method

Design

This study had three conditions, using a policy-capturing design. The first condition was a control condition as it was a replication of the design used by Dunn *et al.* (1995). In this condition, participants did not receive any information about the selection methods used to derive the scores on the Big Five traits and GMA. In the second condition, participants were told that Big Five scores were derived from an unstructured interview and that GMA scores were derived from a paper-and-pencil test. In the third condition, participants were told that Big Five scores were derived from a paper-and-pencil test and that GMA scores were derived from an unstructured interview.

The mix of selection methods within condition was necessary because of the bootstrapping procedure used to derive importance weights that is characteristic of policy-capturing studies. If we were to hold selection methods constant within condition, this design would not inform us about rank-ordering differences of the constructs across selection methods. In other words, you cannot test for the effects of context when you hold context constant. This is because importance weights reflect relative importance within person (i.e. they are positive). Thus, for example, the relative

importance of constructs all assessed with the unstructured interview could be the same as relative importance of constructs all assessed with a paper-and-pencil test even when supervisors believed that one source (selection method) is substantially more credible than another.

Sample and procedure

Research assistants visited retail stores and distributed questionnaires in person to a total of 227 store supervisors. Only store supervisors who screened and evaluated sales job applicants on a regular basis were asked to voluntarily complete the questionnaire. Participants were randomly assigned to a particular questionnaire (see below). After 1 or 2 weeks, researchers visited the stores again and received usable and completed questionnaires from 163 store supervisors (51 men, 112 women, mean age = 37.4 years, $SD = 8.3$ years), yielding a response rate of about 72%. Mean working experience of the participants was 15.9 years ($SD = 9$ years) and their mean managerial experience was 10.5 years ($SD = 7.7$ years). They had previously been involved in hiring an average of 23.2 people ($SD = 59.6$; Median = 7 people). Supervisors were mostly working in retail stores selling clothes (50%), food (18%), and electrical devices (8%).

Performance profiles

We constructed job applicants' profiles that were described on six personal characteristics: the Big Five personality factors (emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness) and GMA. We began by determining the number of performance profiles. We chose a cue-profile ratio of 1–7, yielding 42 (6×7) profiles. This ratio was chosen to balance concern over sampling error with concern for participant fatigue (Cooksey, 1996). Because one repeat profile was included, there were 43 profiles in total.[1]

Next, the scores (i.e. the performance levels) of each profile were developed. To this end, we used a method similar to Dunn *et al.* (1995). In particular, we employed a random number generator for assigning digits from 1 to 5 to the six personal characteristics. These digits represented the five performance levels, with 1 representing *very low performance*, 2 representing *low performance*, 3 representing *average performance*, 4 representing *high performance*, and 5 representing *very high performance*.

We kept three considerations in mind when assigning digits to the six personal characteristics. First, we tried to match the percentage of times a rating category was used in the profiles to the percentage of times it would be expected to occur in a normal distribution of people. In the words of Dunn *et al.* (1995), we tried to 'simulate those ratings that would be expected if the applicant profiles were for real people' (p. 508). Across all profiles, this yielded the following proportions of the very low, low, average, high, and very high performance categories, respectively: 13%, 23%, 28%, 23%, and 13%. Second, we ensured that the mean performance levels were similar across the six constructs. Across all profiles, this produced the following mean performance levels: extraversion ($M = 2.88$; $SD = 1.31$), agreeableness ($M = 2.95$, $SD = 1.25$), conscientiousness ($M = 2.98$; $SD = 1.32$), emotional stability ($M = 2.79$; $SD = 1.26$), openness

[1] *Because one repeat profile was included in the questionnaire, it was possible to compute an overall reliability coefficient across raters. This reliability coefficient was .65. This value is in the range of values found by Dunn et al. (1995).*

($M = 3.14$; $SD = 1.22$), and GMA ($M = 3.29$; $SD = 0.99$). Third, we ensured that the correlation between the trait-based predictors was relatively low in order to obtain stable and accurate regression estimates in our policy-capturing analyses (Cooksey, 1996). The final set of profile scores yielded a matrix of correlations in which all correlations had absolute values less than .09.

Questionnaire and measures

The first section of the questionnaire differed in each condition because it contained the instructions given to the respondents. We ensured that these instructions were closely aligned with the different conditions of our design. In Condition 1, the instructions given were similar to Dunn *et al.* (1995) because Condition 1 was a replication of their study. Participants were instructed to study each of the following 43 personal profiles of hypothetical job applicants and to indicate after each profile whether they would hire the job applicant. It was mentioned that job applicants differed only with respect to the information provided in the profiles and that they met the minimum education, experience, and knowledge requirements for the job. In addition to this information, Condition 2 mentioned that applicants' scores on the Big Five personality factors were derived from an unstructured interview and that applicants' scores on GMA were derived from a paper-and-pencil test (i.e. cognitive ability measure). Conversely, in Condition 3, it was mentioned that applicants' scores on the Big Five personality factors were derived from a paper-and-pencil test (i.e. personality inventory) and that applicants' scores on GMA were derived from an unstructured interview.

The second section of the questionnaire was the same in all conditions and presented the personal profiles of 43 hypothetical job applicants for a sales job (see above). The Appendix provides a sample profile. The anchors for the Big Five traits and GMA were the same as the ones used by Dunn *et al.* (1995).

After each applicant profile, participants rated two items to assess the overall qualifications (hirability) of the applicant. These items were taken from Dunn *et al.* (1995). These items were, 'I would recommend that this person be hired', and, 'If hired, I believe that this person would perform well on a job as a sales person'. Ratings were made after reading each profile on a five-point scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*. The generalizability coefficient of this scale was computed across profiles, respondents, and items and equalled .70.

Analyses

Policy-capturing

As in previous research on this topic (e.g. Dunn *et al.*, 1995) we conducted policy-capturing analyses to capture the decision policy that supervisors used to determine job applicants' overall qualifications. To this end, each manager's judgments of the overall qualifications were regressed on the six personal characteristics. When non-zero intercorrelations exist, standardized regression weights do not provide unambiguous measures of the relative weights of the cues. Therefore, we also computed relative weights using the method outlined by Johnson (2000, 2001). This method has been found to provide unambiguous measures of the relative weights of the cues. In our study, these relative weights correlated highly with the standardized regression weights, with correlations varying from .82 to .97. Because our cues were virtually uncorrelated (see above), this high similarity is not surprising. Hence, we could use the more

common standardized regression weights as measures of the relative weight that supervisors attached to the six personal characteristics. Note also that our results (presented below) are exactly the same for both types of weights.

Hierarchical linear modelling

Whereas the policy-capturing analyses focused on within-subjects factors (i.e. the relative weights attached to the constructs), our next set of analyses examined whether the weights attached to the constructs were moderated by the selection method, which was a between-subjects factor (see Hypotheses 1, 2, and 3). To this end, HLM 5 (Raudenbush, Bryk, Cheong, & Congdon, 2001) was used to simultaneously estimate two models. The first one modelled the relationships between an individual manager's hirability decision and the six constructs at the individual level. This model paralleled the aforementioned policy-capturing model. The second model examined how the relationships estimated at the individual level (i.e. relationships between the six constructs and the hirability decision) varied as a function of the group-level variable, namely the experimental condition. An important advantage of these hierarchical linear modelling analyses over other analyses is that the use of hierarchical linear modelling was consistent with this study's mixed experimental design that contained both within-subjects (i.e. the Big Five traits and GMA) and between-subjects (i.e. the selection method) factors (Rotundo & Sackett, 2002). In addition, these analyses enabled to test our hypotheses.

At Level-1, the regression coefficients of the Big Five and GMA constructs were modelled as random coefficients. At Level-2, these coefficients (i.e. standardized slopes associated with the six constructs) served as dependent variables as they were regressed on the between-subjects selection method factor. As selection method was operationalized with two conditions (Condition 2 and 3), one dummy variable was created.

Results

Preliminary analyses

Prior to testing our hypotheses, we examined the results of Condition 1. As already mentioned, Condition 1 was a replication of prior relative importance studies (Dunn *et al.*, 1995; Ones & Viswesvaran, 1999) because no contextual information about the selection method used was provided to the supervisors. Table 1 shows the means and standard deviations of the standardized (policy-capturing) regression coefficients for the six personal characteristics in Condition 1. Agreeableness emerged as the most important criterion ($\beta = 0.52$) in the decision policy of store supervisors, followed by conscientiousness ($\beta = 0.33$) and GMA ($\beta = 0.25$). Table 1 further shows that the average R between the six trait-based predictors and hirability decision was .76, indicating that store supervisors consistently paid attention to the information provided in the profiles when making their judgments.

These relative importance judgments among retail store supervisors are relatively consistent with meta-analytic findings revealing that agreeableness was the best predictor of customer-service performance (Frei & McDaniel, 1998; Ones & Viswesvaran, 2001) and GMA, conscientiousness, and extraversion were good predictors of sales performance (Vinchur, Schippmann, Switzer, & Roth, 1998).

Table 1. Means and standard deviations of standardized regression coefficients and multiple correlation values in Condition 1

	M	*SD*
Extraversion	0.16[d]	0.14
Agreeableness	0.52[a]	0.19
Conscientiousness	0.33[b]	0.11
Emotional stability	0.16[d]	0.13
Openness to experience	0.14[d]	0.11
GMA	0.25[c]	0.12
R	0.76	0.12

Note. $N = 33$. Although questionnaires were distributed randomly, fewer questionnaires regarding Condition 1 (the control condition) were distributed because it was a replication of the design used by Dunn *et al.* (1995). Letter indices indicate mean differences in importance attached to the attributes. These differences were assessed with Tukey's honestly significant differences test, $p < .01$. Means with the same letter indices are not significantly different.

This result of supervisors' rank ordering being in line with meta-analyses replicates the findings of prior relative importance studies (Dunn *et al.*, 1995; Ones & Viswesvaran, 1999) in another occupation (retail store supervisors).

Test of hypotheses

Table 2 presents the means and standard deviations of the standardized regression (policy-capturing) coefficients for each of the six constructs in Conditions 2 and 3. Hierarchical linear modelling was then used to test our hypotheses about possible selection method effects. As noted above, hierarchical linear modelling allows for the simultaneous estimation of two models. The level-1 analysis estimated parameters describing the relationships between the six constructs and the hirability decision at the individual level. The parameters describing these individual relationships were then

Table 2. Means and standard deviations of standardized regression coefficients and multiple correlation values in Conditions 2 and 3

	Condition 2 ($N = 66$) Big Five scores: interview GMA score: test		Condition 3 ($N = 64$) Big Five scores: test GMA score: interview		
	M	*SD*	*M*	*SD*	*d*
Extraversion	0.26	0.13	0.20	0.13	.46
Agreeableness	0.48	0.16	0.48	0.16	.00
Conscientiousness	0.37	0.14	0.36	0.15	.07
Emotional stability	0.17	0.15	0.17	0.12	.00
Openness to experience	0.17	0.13	0.16	0.13	.08
GMA	0.23	0.11	0.31	0.18	−.55
R	0.79	0.10	0.80	0.07	–

used as dependent variables in the level-2 analyses wherein the role of the group variable (condition) was assessed.[2] Table 3 presents the results of these level-2 analyses. The values in the body of this table are the coefficients (γs) related to the effect of condition as estimated in the level-2 analyses.

Table 3. Summary of Level-2 analyses predicting Level-1 coefficients from condition

Dependent variables (Level-1)	Level-2 coefficients (γ's) associated with independent variable condition
Extraversion	.07**
Agreeableness	.00
Conscientiousness	.02
Emotional stability	.00
Openness to experience	.01
GMA	−.08**

Note. $N = 130$, $^{**}p < .01$.

Hypothesis 1 stated that the importance attached to GMA would vary by condition (selection method used). Specifically, we expected that GMA would play a significantly more important role in supervisors' judgments of applicant qualifications if they knew that applicants' GMA scores were derived from an employment interview (Condition 3) than if they knew that GMA scores came from a paper-and-pencil test (Condition 2). As shown in Table 3, condition significantly predicted the relationship between the GMA weight and the hirability decision ($\gamma = -.08$, $t = -3.01$, $p < .01$). Figure 1 graphically displays how the weight that supervisors attached to GMA varies as a function of information about the selection methods. Note that these GMA weights on the vertical axis are not weights from the individual regression (policy-capturing) analyses averaged by condition. Instead, the weights in Fig. 1 are obtained from the HLM analyses. As shown in Fig. 1, the relationship between GMA and hirability decision was stronger in Condition 3 than in Condition 2. All of this supports Hypothesis 1 because Condition 3 stated that applicants' GMA scores were derived from an unstructured interview, whereas Condition 2 stated that applicants' GMA scores were derived from a paper-and-pencil test.

We expected with Hypothesis 2 that the relative importance attached to extraversion would vary across Conditions 2 and 3. Table 3 shows that condition significantly predicted the relationship between the extraversion weight and hirability decision ($\gamma = .07$, $t = 3.05$, $p < .01$). As shown by Fig. 1, the relationship between extraversion

[2] *Before testing our hypotheses using hierarchical linear modelling, we examined whether the conditions associated with hierarchical linear modelling were met (see Hofmann, 1997). First, we explored whether there was systematic within-subjects and between-subjects variance in hirability decision. The* χ^2 *test on the null model* ($\chi^2 = 796.53$, $df = 128$, $p < .000$) *was significant, indicating that the hirability decision varied significantly by the different contexts. In addition, the intra-class correlation showed that 11.06% of the variance in hirability decision is accounted for by the context. The second and third conditions were also met because there was significant variance in both intercepts and slopes to justify level-2 analyses for each of our six constructs. Because these conditions involved estimating level-1 intercepts, we examined whether these conditions were established by using unstandardized regression coefficients (with group mean centring). For reasons of consistency with our individual-level results (Tables 1 and 2), the hierarchical linear modelling results presented (Table 3) were based on standardized regression coefficients.*

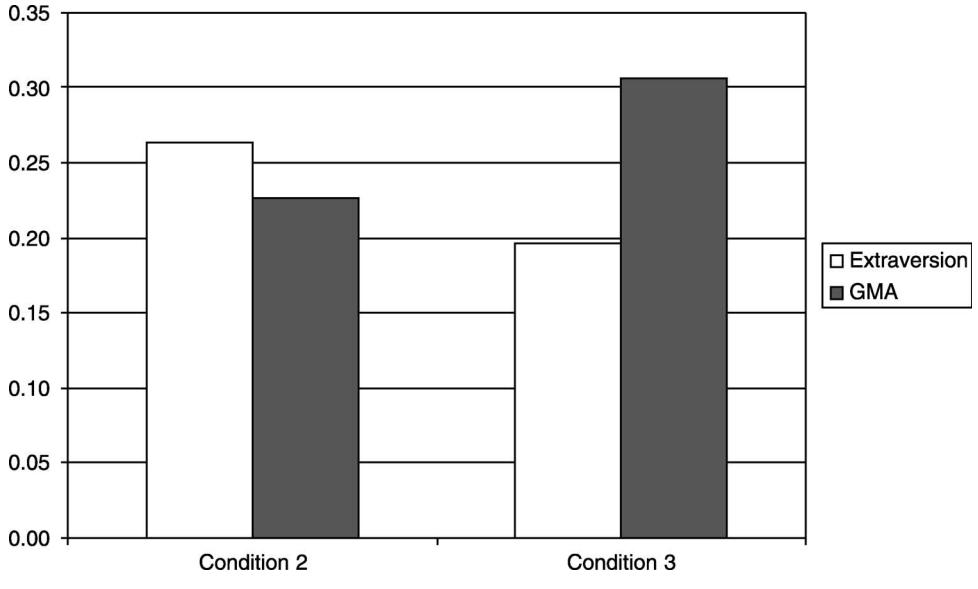

Figure 1. Bar graph of relative importance of GMA and extraversion in Conditions 2 and 3.

and hirability decision was stronger in Condition 2 than in Condition 3. This supports Hypothesis 2 because Condition 2 meant that applicants' scores on extraversion were derived from an unstructured interview, whereas Condition 3 implied that applicants' extraversion scores were derived from a paper-and-pencil test.

Furthermore, Fig. 1 shows that extraversion was more important than GMA in Condition 2, whereas GMA was more important than extraversion in Condition 3. In other words, extraversion became more important than GMA when it was measured via an unstructured interview (and GMA was measured via a paper-and-pencil test) and GMA became more important than extraversion when it was measured via an unstructured interview (and extraversion was measured via a paper-and-pencil test). This demonstrates that the rank ordering of these trait-based predictors (GMA and extraversion) changed as a function of the information provided about the selection method.

Hypothesis 3 posited that the relative importance attached to conscientiousness would vary as a function of the selection method used. However, Table 3 shows that the selection method did not significantly affect the relationship between the conscientiousness weight and hirability decision. Therefore, there was no support for Hypothesis 3.

Additional analyses

We also examined whether other variables besides condition might moderate the importance attached to the six constructs. Specifically, an anonymous reviewer suggested that the amount of selection-related experience (i.e. the number of people store managers had hired in the past) might be an important additional moderator. Therefore, we ran an HLM analysis with amount of selection experience as an additional variable included in the level-2 analysis. There were two important conclusions from this additional analysis. First, our original HLM results remained exactly the same when the selection experience variable was added, attesting to the robustness of our results.

Second, the amount of selection experience moderated the weight attached to two constructs, namely openness and GMA. Specifically, store supervisors with more selection experience attached more importance to GMA and openness.

Discussion

In previous studies on the relative importance of trait and ability judgments (Dunn *et al.*, 1995; Ones & Viswesvaran, 1999), managers evaluated applicant information and made hiring decisions without knowing the selection methods that were used to derive the construct-related information. This is at odds with operational selection practice wherein hiring managers typically know which selection methods were used. Therefore, this study provided supervisors with information about the source of the trait and ability information (paper-and-pencil test versus unstructured interview). In addition, we relied on decision-making research on preference for justifiable information and social psychological theories about the easiness of trait judgment to formulate hypotheses about the moderating effects of the selection method on the relative importance of GMA and the Big Five in supervisors' hirability decisions.

Two of our three hypotheses were supported. We found that the relative importance of extraversion and GMA in hirability decisions was significantly moderated by the mode of assessment of those predictors. These selection method effects on extraversion and GMA were substantial because the rank ordering of these traits was not the same within each condition. Specifically, the rank ordering of extraversion and GMA changed depending on the selection method used to measure these constructs: Extraversion became more important than GMA when it was measured via an unstructured interview (and GMA was measured via a paper-and-pencil test). Yet, GMA became more important than extraversion when it was measured via an unstructured interview (and extraversion was measured via a paper-and-pencil test).

Our results illustrated that the source of information to derive GMA scores plays an important role. Supervisors seem to be more sceptical about the paper-and-pencil assessment method than about the unstructured interview format for measuring GMA. As noted above, practitioners' preference for subjective procedures, experiential data, intuitive thinking, and personal evaluations is not new (Dipboye, 1994; Highhouse, 2002; Johns, 1993; Shrivastava & Mitroff, 1984; Terpstra & Rozell, 1997). However, no prior studies had established that knowledge of which selection methods were being used also affected the importance attached to the construct-related information derived from these methods. This link might have important effects on the validity of supervisors' hiring decisions. For instance, given the superior predictive validity of GMA (Schmidt & Hunter, 1998), the fact that retail store supervisors attached less importance to GMA as a function of the selection method might seriously detract from the validity of their hirability decisions. These practical ramifications of our study contribute to the field of perceptions of selection procedures because prior studies in that field have been criticized for their lack of linking people's perceptions and beliefs to practical consequences (Ryan & Ployhart, 2000, p. 592). Hence, future research should further integrate the literature on the relative importance of trait judgments with the literature on the perceptions of selection procedures. Equally important, future studies are needed to scrutinize why supervisors have such a prejudice against cognitive ability tests. We hypothesized that the preference for intuitive assessment over more mechanical approaches stems in part from the need to provide a justifiable reason for making an otherwise difficult choice, a phenomenon that is well documented in the

decision-making literature (Hastie & Dawes, 2001). Another explanation is that this study's store supervisors are simply less familiar with tests than with interviews. Our finding that the amount of hiring experience moderated the weight given to GMA (with more experienced store managers attaching more importance to GMA) supports this explanation. However, another explanation is that unstructured interviews are more under the control of store supervisors (see Dipboye, 1994, 1997), whereas most tests and structured interviews are developed by consultants (or HR departments). Therefore, participants might place more weight on unstructured interviews.

The source of information also influenced the relative importance attached to extraversion. Our finding that the weight attached to extraversion increased when supervisors knew that this trait was measured in an unstructured interview conforms to theories on easiness of trait judgment (e.g. Funder, 1999). It also fits well into recent interview studies (Barrick, Patton, & Haugland, 2000; Caldwell & Burger, 1998) showing that interviewers were able to accurately assess extraversion because it provided them with ample detectable and visible cues about extraversion (e.g. being talkative, being active).

An unexpected finding was that the importance attached to conscientiousness did not change when it was measured in an unstructured interview. Thus, contrary to our hypothesis based on theories on easiness of trait judgment, supervisors did not believe the unstructured interview to be a good procedure for obtaining observable cues related to conscientiousness. Along these lines, two recent studies (Barrick *et al.*, 2000; Caldwell & Burger, 1998) also obtained disappointing results for conscientiousness in an interview context as these studies demonstrated that interviewers were not able to make valid and accurate judgments of conscientiousness. Another explanation for the insignificant result for conscientiousness in our study is that supervisors may have viewed paper-and-pencil tests to be just as effective as employment interviews for assessing this construct.

This study issues a mixed message regarding the existence of the so-called practice-science gap. On the positive side, supervisors seem to attach most importance to agreeableness, followed by conscientiousness, GMA, and extraversion. These relative importance judgments among retail store supervisors parallel conclusions of meta-analyses about the predictors of sales and customer-service performance (Frei & McDaniel, 1998; Ones & Viswesvaran, 2001; Vinchur *et al.*, 1998). On the negative side, there seems to be an inverse relationship between the selection procedures that supervisors believe to be valid and the psychometric evidence regarding these procedures. In particular, we found that these practitioners were sceptical about paper-and-pencil tests. Conversely, supervisors seemed to prefer that traits/abilities be assessed using the unstructured interview, even though this method is least likely to be able to reliably and validly assess traits/abilities. Hence, contrary to supervisors' importance judgments of the constructs, supervisors' perceptions of the selection procedures are not in line with meta-analytical evidence.

With respect to implications for practice, this study's results inform industrial and organizational psychologists' understanding of what interventions are needed for improving hiring practitioners' selection decision practices. Industrial and organizational psychologists might do well to concentrate on practitioners' beliefs of the selection procedures used to derive the construct-related information. Specifically, training might provide people with feedback about their decision strategies so that they are better aware of their inaccuracies. Training might also provide them with guidance about how to weigh applicant information. For instance, this might be done by giving

anecdotal or scientific information on the reliability and validity of these selection procedures (Lievens *et al.*, 2003; Truxillo, Bauer, Campion, & Paronto, 2002). However, another approach may be to encourage decision makers to make their attribute importance decisions explicit and public. Encouraging supervisors to think about and articulate the importance they place on, say, extraversion relative to agreeableness in a given selection procedure may result in increased use of and commitment to pre-established standards (Highhouse, 1997). Future studies should examine whether these strategies enable industrial and organizational psychologists to change managerial beliefs in the scientific value of selection procedures.

Clearly, the contextual information that might influence the weight decision makers place on predictors is more complex than informing them whether the scores were derived from a paper-and-pencil test or an unstructured interview. More research is needed to determine the generalizability of our results across other interview formats (e.g. structured interviews such as behavioural description interviews or situational interviews), other test formats (e.g. on-line test, cognitive ability test with business-related content), other selection methods (e.g. assessment centres), and other decision makers (industrial and organizational psychologists, personnel managers, etc.). This study also sets the stage for future research about other contextual influences such as the type of job or organizational practices (see Wilk & Cappelli, 2003). For example, people are likely to perceive tests differently if they are used in the first hurdle of the selection process than in later stages.

Some limitations of this study should be acknowledged. Our study was an experimental policy-capturing study. This design typically maximizes internal validity at the expense of external validity. Therefore, we tried to increase the external validity of our study by using a field sample and by building realistic performance levels into the applicant profiles. Despite these efforts, other aspects of this study were less realistic. For example, supervisors rated written candidate profiles. Although the presentation format of these profiles resembled candidate reports, supervisors do not make hiring decisions solely based on such written reports. In operational selection practice, they typically have met the candidates. The policy-capturing design also required that the intercorrelation among the trait-based constructs was virtually orthogonal. Although this requirement is realistic for the intercorrelations between GMA and the Big Five (Ackerman & Heggestad, 1997), it seems less realistic for the intercorrelations among the Big Five traits. In empirical studies, the Big Five traits often correlate in the .30s (Block, 1995). However, Saucier (2002) recently demonstrated that intercorrelations among the Big Five are not a property of the Big Five traits themselves which are posited to be theoretically orthogonal (Digman, 1997). Instead, nonorthogonality seems to be an unintended outcome of scale construction procedures.

In conclusion, this field experiment showed that supervisors' judgments about the importance of applicant traits (extraversion) and abilities (GMA) can be strongly influenced by the way in which inferences about these traits and abilities are derived. This study's supervisors valued these constructs more when they were assessed using an unstructured interview, even though this method is known to produce inadequate reliabilities and validities. Conversely, supervisors valued these constructs less when they were measured via paper-and-pencil tests. It appears that paper-and-pencil testing continues to suffer from negative stereotypes dating back to William F. Whyte's (1956) classic condemnation in *The Organization Man*. The next step seems to be to gain a better understanding of the origin and nature of this alternative kind of test bias.

Acknowledgements

We would like to thank Stefanie Lefevere and Olivia De Poorter for their help in collecting the data. We also would like to acknowledge Neil Christiansen for his insightful comments on a previous version of this manuscript. A previous version of this manuscript was presented at the Annual Conference of the Society of Industrial and Organizational Psychology, Orlando, Florida (2003).

References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.

Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*, 125–154.

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

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology*, *53*, 925–951.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187–215.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Caldwell, D. F., & Burger, J. M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology*, *51*, 119–136.

Connolly, J. J., & Viswesvaran, C. (1998, April). *The convergent validity between self- and observer ratings of personality*. Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX, USA.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*, San Diego, CA: Academic Press.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*, 1246–1256.

Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. In G. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 12, pp. 79–123). Greenwich, CT: JAI Press.

Dipboye, R. L. (1997). Structured selection interviews: Why do they work? Why are they underutilized? In N. Anderson & P. Herriot (Eds.), *International handbook of selection and assessment* (pp. 455–473). New York: Wiley.

Dunn, W. S., Mount, M. K., Barrick, M. R., & Ones, D. S. (1995). Relative importance of personality and general mental ability in managers' judgments of applicant qualifications. *Journal of Applied Psychology*, *80*, 500–509.

Frei, R. L., & McDaniel, M. A. (1998). Validity of customer service measures in personnel selection: A review of criterion and construct evidence. *Human Performance*, *11*, 1–27.

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

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, *55*, 149–158.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with inter-judge agreement. *Journal of Personality and Social Psychology*, *51*, 409–418.

Glode, J., Truxillo, D. M., & Bauer, T. N. (2003, April). *Manager reactions to personnel selection methods*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL, USA .

Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. Busmeyer, D. L. Medin, & R. Hastie

(Eds.), *Decision making from a cognitive perspective* (Vol. 32, pp. 83–126). San Diego, CA: Academic Press.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world*. Thousand Oaks, CA: Sage.

Highhouse, S. (1997). Understanding and improving job-finalist choice: The relevance of behavioral decision research. *Human Resource Management Review*, *7*, 449–470.

Highhouse, S. (2002). Assessing the candidate as a whole: An historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology*, *55*, 363–396.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*, 723–744.

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

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, *85*, 869–879.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

John, O. P., & Robbins, R. W. (1993). Determinants of inter-judge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*, 521–551.

Johns, G. (1993). Constraints on the adoption of psychology-based personnel practices: Lessons from organizational innovation. *Personnel Psychology*, *46*, 569–592.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1–19.

Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology*, *86*, 984–996.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*, 296–310.

Kravitz, D. A., Stinson, V., & Chavez, T. L. (1996). Evaluations of tests used for making selection and promotion decisions. *International Journal of Selection and Assessment*, *4*, 24–34.

Lievens, F., De Corte, W., & Brysse, K. (2003). Applicant perceptions of selection procedures: The role of selection information, belief in tests, and comparative anxiety. *International Journal of Selection and Assessment*, *11*, 67–77.

Miller, S. R., & Rosenbaum, J. E. (1997). Hiring in a Hobbesian world: Social infrastructure and employers' use of information. *Work and Occupations*, *24*, 498–523.

Ones, D. S., & Viswesvaran, C. (1999). Relative importance of personality dimensions for expatriate selection: A policy capturing study. *Human Performance*, *12*, 275–294.

Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-related focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, *9*, 31–51.

Petty, 'R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 323–390). New York: McGraw-Hill.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2001). *HLM 5: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Ree, M. J., Carretta, T. R., & Steindl, J. R. (2001). Cognitive ability. In N. Anderson, D. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology: Personnel psychology*, (Vol. 1, pp. 219–232). London/New York: Sage.

Rettinger, D. A., & Hastie, R. (2001). Content effects on decision making. *Organizational Behavior and Human Decision Processes*, *85*, 336–359.

Rosse, J. G., Miller, J. L., & Stecher, M. D. (1994). A field study of job applicants' reactions to personality and cognitive ability testing. *Journal of Applied Psychology*, *79*, 987–992.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*, *87*, 66–80.

Rousseau, D. M., & Fried, Y. (2001). Location, location, location: Contextualizing organizational research. *Journal of Organizational Behavior*, *22*, 1–13.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, *52*, 359–391.

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*, 565–606.

Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology*, *7*, 261–277.

Salgado, J. F. (1997). The Five-factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, *82*, 30–43.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, *76*, 323–346.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*, 573–605.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the big five. *Journal of Research in Personality*, *36*, 1–31.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187–211.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.

Shrivastava, P., & Mitroff, I. I. (1984). Enhancing organizational research utilization: The role of decision makers' assumptions. *Academy of Management Review*, *9*, 18–26.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*, 49–76.

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, *81*, 134–141.

Steiner, D. D., & Gilliland, S. W. (2001). Procedural justice in personnel selection: International and cross-cultural perspectives. *International Journal of Selection and Assessment*, *9*, 124–137.

Terpstra, D. E., & Rozell, E. J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management*, *26*, 483–495.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703–742.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239–257.

Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2002). Selection fairness information and applicant reactions: A longitudinal field study. *Journal of Applied Psychology*, *87*, 1020–1031.

Van der Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology*, *87*, 176–184.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., III, & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, *83*, 586–597.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling in social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125.

Whyte, W. H., Jr. (1956). *The organization man*. New York: Simon and Schuster.
Wilk, S. L., & Cappelli, P. (2003). Understanding the determinants of employer use of selection methods. *Personnel Psychology*, *56*, 103–124.
Williams, J. E., Munick, M. L., Saiz, J. L., & FormyDuval, D. L. (1995). Psychological importance of the Big Five: Impression formation and context effects. *Personality and Social Psychology Bulletin*, *21*, 818–826.

Received 15 September 2003; revised version received 30 April 2004

Appendix

Example of applicant profile

Introverted, quiet, shy	×					Extraverted, assertive, talkative
Cold, unkind, unsympathetic	×					Warm, kind, sympathetic
Careless, unsystematic, inefficient				×		Careful, systematic, efficient
Tense, nervous, touchy		×				Relaxed, calm, unexcitable
Unimaginative, unintellectual, uncreative			×			Imaginative, intellectual, creative
Dull, slow to solve problems, slow to learn new skills			×			Bright, quick to solve problems, quick to learn new skills