How Accurate are Recruiters' First Impressions of Applicants in Employment Interviews?

Marianne Schmid Mast, Adrian Bangerter, Céline Bulliard and Gaëlle Aerni

Institute of Work and Organizational Psychology, University of Neuchatel, Rue Emile-Argand 11, CH-2000 Neuchatel, Switzerland. marianne.schmid@unine.ch

The ability of recruiters and laypersons (students) to detect applicant personality traits and deception was studied. Participants viewed mock videotapes of target applicants answering interview questions. They subsequently judged the applicants' personality on the Big Five dimensions. Then, they viewed another videotape with other applicants presenting themselves either truthfully or not, and subsequently guessed which version was truthful. Personality judgments were compared with targets' self-assessments and peer assessments to create an accuracy score. Both recruiters and students accurately detected applicants' global personality profile. Recruiters were better at this than students. However, students were better at judging the specific traits of openness, extraversion, and conscientiousness, whereas recruiters only accurately detected openness. Recruiters detected lies above chance whereas students did not.

1. Introduction

When undergoing recruitment training, recruiters are regularly discouraged to trust their instincts and first impressions about applicants (Woehr & Huffcutt, 1994) and indeed there is a plethora of research demonstrating that recruiters often are biased in their judgments through expectations or stereotypes they harbor (Dougherty, Turban, & Callender, 1994; Macan & Dipboye, 1990b). But what if recruiters' gut feelings, hunches, intuition, or first impressions are correct? Research on first impression formation suggests that even nonexperts in interpersonal assessment (e.g., laypersons who are not recruiters) are capable of correctly judging other's personality traits (Borkenau & Liebler, 1992; Funder & Colvin, 1997; Funder, Kolar, & Blackman, 1995; Funder & West, 1993), their intelligence (Borkenau & Liebler, 1992; Murphy, Hall, & Colvin, 2003), their emotions and motivations (Ickes, 2003, 1997; Nowicki & Duke, 1994; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), and the type of social relationships they are in (Barnes & Sternberg, 1989; Bernieri & Gillis, 1995, 2001; Schmid Mast & Hall, 2004) at better than chance level (note that deception detection is an exception in that research shows that people are generally not able to detect others' lies) (Malone & DePaulo, 2001).

So if people are able to judge others' traits and states accurately even without prior knowledge about their partner, why should recruiters be so biased? Indeed many recruiters seem to trust their instincts, gut feelings, and first impressions of applicants more than objective tests (Dipboye, 1994). Moreover, there is some evidence to suggest that such first impressions recruiters form about a job applicant can be correct. For instance, recruiters were accurate in assessing the personality characteristics of a job applicant based on the applicant's CV (Cole, Feild, & Giles, 2003).

Nevertheless, when reviewing the personnel psychology literature on how recruiters form impressions of applicants, one gets the impression that recruiters are – as a general rule – severely biased and constantly misled in their judgments (Dougherty *et al.*, 1994; Posthuma, Morgeson, & Campion, 2002). There thus has been much research activity showing that recruiters form first impressions of applicants (Dipboye, 1982) and how these impressions can be affected by expectations turning into self-fulfilling prophecies (Macan & Dipboye, 1990a) and by stereotypes affecting recruiter judgment (Bardack & McAndrew, 1985; Gravesa & Powell, 1988; Snyder & Swann, 1978). However, when focusing on accuracy instead of on bias and stereotypes, there is relatively scarce evidence available from the field of applicant

International Journal of Selection and Assessment

assessment showing that recruiters are able to correctly judge certain aspects of a applicant's personality (Barrick, Patton, & Haugland, 2000; Blackman, 2002). Moreover, these studies typically did not use a first impression approach.

Studies addressing whether and how people form accurate impressions about others typically investigate situations in which strangers meet for the first time. Such zero-acquaintance situations (Ambady, Hallahan, & Rosenthal, 1995; Kenny, Horner, Kashy, & Chu, 1992) have often been studied with the thin-slice paradigm (Ambady, Hallahan, & Conner, 1999; Ambady, LaPlante, & Johnson, 2001; Ambady & Rosenthal, 1992) where judgments are made based on a brief observation (<5 min) of the social interaction partner. In the present paper, we understand first impression judgments as assessments about a stranger based on a thin slice of observable behavior.

Note that recruiter interactions with applicants in an employment interview typically last longer (e.g., 30 min; Barrick *et al*., 2000) and in such situations, recruiters seem to be able to correctly assess the applicant's personality although there are differences depending on the specific traits assessed, as detailed below (Barrick *et al*., 2000). The question we investigated is whether recruiters are still good judges of job applicants when they have to rely on their first impressions, that is, a thin slice of observed applicant behavior. Given the research showing the good performance of laypersons in interpersonal assessment we would expect recruiters to perform well. But would they actually perform better than an average layperson? Moreover, we investigated whether recruiters were aware of their individual level of assessment performance and whether job experience and/or recruitment training were associated with better interpersonal judgments in recruiters.

1.1. Personality assessment via tests and interviews

Barrick and Mount (1991) showed that conscientiousness was the most important personality trait for predicting job performance, regardless of how performance was measured (i.e., training success, personnel data) and regardless of type of occupation (i.e., police, managers, sales). Other personality aspects differed in their predictive validity according to the type of performance measure or the occupational domain. For example, extraversion was a good predictor of any kind of job performance in occupational domains involving social interactions (e.g., sales, management).

Applicant personality characteristics not only predict job performance but also an array of other important outcomes such as job satisfaction (Judge, Heller, & Mount, 2002), counter-productive work behavior (Blackman & Funder, 2002) and they are important for person–organization fit (P–O fit) (Cable & Judge, 1997; OReilly, Chatman, & Caldwell, 1991). Lack of P–O fit is related to negative job outcomes such as withdrawal behaviors, lowered performance, negative attitudes, strain, and increased turnover (Kristof-Brown, Zimmerman, & Johnson, 2005). In the same vein, conscientiousness, agreeableness, and neuroticism explain counter-productive work behavior (Blackman & Funder, 2002).

Thus, personality is a predictor of many work outcomes. However, recent discussions (e.g., Morgeson *et al*., 2007) have cast doubt on the association between personality traits and performance, as well as on the appropriateness of personality tests as measures of personality in selection settings. Personality characteristics can also be assessed by observers, for instance by recruiters during an interview. The interview is widely used in personnel selection and possesses predictive validity for future job performance (Robertson & Smith, 2001; Schmidt & Hunter, 1998) and incremental validity when used together with a general mental ability test (Schmidt & Hunter, 1998). Observer-assessed personality traits are the most frequent construct assessed in interviews (35%), followed by social competencies, mental abilities, knowledge, and skills among others (Huffcutt, Conway, Roth, & Stone, 2001). The fact that personality is the most frequently assessed construct in interviews makes sense because it predicts future job performance (Barrick & Mount, 1991) and seems to be assessable by observers: ratings of conscientiousness and extraversion and to some extent also agreeableness predicted job performance for sales representatives (Mount, Barrick, & Strauss, 1994).

Remains the question as to whether recruiters are actually able to assess applicants' traits accurately. One study (Barrick *et al*., 2000) showed that recruiters were able to judge applicants' extraversion, openness, and agreeableness correctly (but not their conscientiousness or emotional stability) in a 30-min job interview. There is also evidence showing that in unstructured employment interviews, personality traits of applicants can be assessed more accurately than in structured interviews (Blackman, 2002). But are recruiters good judges of personality when basing their judgment on thinner slices of behavior in interview situations? It remains unclear how much behavioral information is necessary for accurate judgments.

1.2. Personality assessments by laypersons

Not only recruitment experts can accurately assess applicants' personality characteristics. Lay judges were able to accurately assess applicants' integrity in simulated employment interviews (Townsend, Bacigalupi, & Blackman, 2007). It has long been demonstrated that even untrained observers are able to judge the personality of others. As an example, strangers are able to correctly detect others' extraversion, agreeableness, conscientiousness, and openness, but not their neuroticism

(Borkenau & Liebler, 1992) based on videotapes of behavioral excerpts. Gifford (1994) found that ambitious-dominant, gregarious-extraverted, and aloof-introverted could be assessed accurately by strangers observing others interacting during 15 min while lazy-submissive and cold-quarrelsome could not. Moreover, Funder and Colvin (1988) showed that based on 5-min interactions, extraversion could be assessed accurately by strangers, whereas openness, agreeableness, and conscientiousness could not and neuroticism showed even a significant negative correlation between self- and stranger-ratings. Also, Watson (1989) showed that extraversion and conscientiousness could be assessed accurately by strangers whereas agreeableness, openness, and neuroticism could not.

These results clearly show that not all personality factors can be assessed equally accurately. The Realistic Accuracy Model (RAM) (Funder, 1995; Funder *et al*., 1995) describes how accurate judgments come about. One important aspect in the model is the 'good trait': some traits are more easily expressed in behavior than others and therefore more easily judgable. As an example, extraversion is readily expressed by talking a lot and behaving in a sociable way by nodding, smiling, and maintaining eye contact, which explains why extraversion seems to be relatively easily judgable.

1.3. Are recruiters better assessors than laypersons – or what is the effect of training and experience?

Barrick *et al*. (2000) compared personality assessments by experienced interviewers and strangers. While recruiters were able to correctly assess applicants' extraversion, openness, and agreeableness, strangers could only assess extraversion and agreeableness to some extent (approaching significance). This suggests that recruiters have an advantage when judging an applicant's personality. In the same vein, managers in an executive MBA program (who have experience in conducting selection interviews) were more accurate in assessing applicants for sales positions in an assessment centre than were I/O psychology students (Lievens, 2001). Some studies show that recruiters benefit from their experience and that training increases judgment accuracy, while others do not find an effect of training or expertise (Posthuma *et al*., 2002).

1.4. Deception detection

To the extent that applicants try to appear in the best light possible and use different impression management strategies to do so, deception detection becomes an important topic in personnel selection. Research shows that such impression management strategies are in fact extensively used by applicants (Ellis, West, Ryan, & DeShon, 2002; Levashina & Campion, 2007). Applicants might not often be blatantly lying but they may embellish certain facts about themselves.

The recruiter on the other end of the equation is motivated to detect these deceptions in order to form a more accurate impression of the applicant. Faking has been studied mostly from the applicants' perspective with regard to how truthfully they fill in personality questionnaires and present themselves during the interview. Detection of faking by recruiters has gained only moderate research attention (Lievens & Peeters, 2008; Van Iddekinge, Raymark, & Roth, 2005).

Research on deception detection suggests that people perform poorly when trying to detect others' lies. Mean accuracy resulting from a meta-analysis is around 54% and thus only slightly better than chance (Malone & DePaulo, 2001). However, there are groups of people who perform well on lie detection tasks. Professionals from the US Secret Service did much better than other groups tested by Ekman on a standardized lie detection test (Ekman, 2001). Judges, trial attorneys, policemen, polygraphers, and psychiatrists all performed at chance level. To date, we do not know whether recruiters are better lie detectors than other groups. One could argue that recruiters have much more experience in assessing others and detecting their attempts to deceive. We were thus interested in knowing whether recruiters are able to detect deception in job applicants, whether they are better at it than laypersons and whether their deception detection performance is related to their interpersonal personality assessment skills and their job experience.

1.5. Self-evaluation of interpersonal assessment skills

Davis and Kraus (1997) point out that people in general do not have an accurate perception about how good a judge of others' personality they are. Moreover, Funder (1995) argues that while a good judge of personality most likely knows that he or she is good, a poor judge might lack this insight, explaining why self-evaluation of one's assessment skills might not correlate with one's actual performance in inferring other's traits. Indeed, there was no relation between assessors' self-ratings of ability to accurately judge others with their actual ability to judge another's personality (Colvin & Bundick, 2001). Moreover, in lie detection, how accurately people can detect others' lies is unrelated to their confidence in their judgments (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). To the best of our knowledge, it has never been tested whether recruiters might have more accurate insight in their abilities to judge others than laypersons. Recruiters might outperform laypersons in deception detection because they might get feedback (i.e., hearing whether a person really showed the predicted personality traits at work) and thus be able to

learn and calibrate their judgments. In the present research, we investigate whether recruiters and laypersons are accurate in judging their ability to judge others.

1.6. The present study

There are a number of methodological challenges inherent in determining whether a person is a good judge of another's personality. One issue concerns the criterion to which the assessor's judgment is compared. In many instances, the self-reported scores on a personality questionnaire serve as the criterion (Funder & Colvin, 1997; Gifford, 1994; Watson, 1989). Because these judgments can be affected by social desirability, because applicants might be motivated to fake, and because we do not know to what extent a single measure is reliable, a common practice is to increase reliability of the self-report by combining it with the reports of acquaintances; these seem to be correlated with each other and with the self-report (Colvin, Vogt, & Ickes, 1997; Funder & Colvin, 1988). Thus, in addition to targets' self-reports, we also asked two acquaintances to fill out the same questionnaire about the targets.

We defined first impressions as judgments based on the observation of a thin slice of behavior (<5 min). Note that almost all research investigating whether recruiters are able to correctly assess applicants used longer slices of behavior: 30 min (Barrick *et al*., 2000) or 10 min (Blackman & Funder, 2002). Research on personality assessment by strangers outside the recruitment literature has also often relied on thicker slices of behavior: 15 min (Gifford, 1994; Watson, 1989) or 5 min (Funder & Colvin, 1988). Meta-analytic research shows that thin slices of behavior permit accurate judgments irrespective of slice duration (Ambady & Rosenthal, 1992). We therefore have good reason to believe that recruiters as well as laypersons are able to accurately assess personality traits of others from thin slices of interpersonal behavior (2 min in the present study). Also, we expect the nature of the trait assessed to affect accuracy, with extraversion being assessed accurately and neuroticism not being assessed accurately.

In sum, we tested several questions. First, are recruiters and students able to (a) assess the personality of applicants and (b) detect whether they tell the truth or not at better than chance level? Second, relying on the five-factor model of personality (Costa & McCrae, 1992), how does the nature of the trait to be assessed affect accuracy? We predict that extraversion can be assessed accurately by both recruiters and students and that neuroticism cannot. Third, are recruiters better judges of applicant personality and better at detecting whether the applicant tells the truth or not than are students? Fourth, are recruiters and students accurate in judging their actual performance regarding their interpersonal assessment skills and how do these different skills relate to each other? Fifth, are recruiters' job experience and training related to their performance in personality assessment or lie detection? Finally, we also added participant gender as a factor in our analyses because research shows that women outperform men in general when it comes to assessing interaction partners correctly (Hall, 1984, 1998).

2. Method

2.1. Participants

Participants were 51 recruiters (27 women, 22 men, and two recruiters of unknown gender) and 80 university students (55 women, 25 men, undergraduate and graduate level, various fields of study). Recruiters were contacted through the human resource department of several large companies via phone or email ($N = 118$ initially contacted). Students were contacted in classes. Recruiters performed the study individually – they received the instructions, questionnaires, and the video to watch by mail and returned everything completed via mail to the researchers, whereas students were tested in groups (five to 25 participants per group). On average, recruiters were 36 and students were 22 years old. Recruiters had on average 6 years of experience as recruiters and had on average undergone recruitment training for one and a half years.

2.2. Procedure

Participants first completed a questionnaire measuring their self-evaluated assessment skills, then watched six target applicants on videotape. After having watched an applicant, participants filled in a personality questionnaire assessing the personality traits of the target applicant. Participants also reported their gender and age. Recruiters also indicated how many years of recruitment training and how many years of job experience related to conducting job interviews they had.

2.3. Material and measures

2.3.1. Target applicants on videotape

Participants watched six different target applicants, all filmed while presenting themselves for a mock job interview in their respective domains and answering three questions: what motivates them in their job, what asset they bring to the job compared with other people, and describing one of their strengths and one of their weaknesses. Video recordings of each target applicant lasted approximately 2 min each. There were three female and three male target applicants (age range approximately between 23 and 55 years), put in random order to produce the video clip shown to participants. All participants saw the same video clip, thus assessed target applicants always in the same order.

2.3.2. Personality traits

We used an abbreviated form of the NEO Big Five personality trait measure (Schallberger & Venetz, 1999) for the target applicants' self-reported personality traits, for the questionnaire two of their friends filled in about them, and for the participants who judged each target applicant's personality traits. Each dimension of the NEO was assessed with six items in semantic differential form on six-point scales, for example, on the *calm-irritable* continuum anchor points were 1 = *very calm*, 2 = *pretty calm*, 3 = *rather calm*, 4 = *rather irritable*, 5 = *pretty irritable*, or 6 = *very irritable*. Scores on the items were averaged by dimension; larger values indicate stronger presence of a given trait.

2.3.3. Target applicants' personality self-assessment

Target applicants were asked to fill in the NEO (described above) and to ask two good friends of theirs to fill in the NEO about them. For target applicants and their friends, the correlations between the friend's and the self-ratings across the five dimensions ranged from −.85 to .93 (median = .73). Correlations among friends' ratings ranged from .51 to .77 (median = .58). We averaged the targets' personality self-report measure and the two friends' personality measure about the target for each dimension of the NEO. These aggregated ratings served as the criterion to which each participant's perception of the target applicant was compared with in order to assess accuracy of the personality ratings.

2.3.4. Perceived personality of target applicants

Participants assessed each of the six target applicants after having watched him or her in the video on the same abbreviated form of the NEO described above. Internal consistency (Cronbach's α) for the assessment of each target applicant varied between .64 and .82 (median = .74) for extraversion, between .70 and .79 (median = .76) for agreeableness, between .61 and .84 (median = .81) for conscientiousness, between .67 and .80 (median = .72) for neuroticism, and between .53 and .72 (median = .64) for openness.

2.3.5. Personality profile accuracy

To measure how well an observer (recruiter or student) assessed the personality profile (the relation of each of the NEO dimensions to each other) of each target applicant, we used a correlational approach (Davis & Kraus, 1997). For each participant, we calculated the correlation between his or her ratings of the target on each of the five NEO dimensions and the target's ratings of themselves (previously aggregated with their friends') across the five NEO dimensions. This was performed for each participant and for each target separately. For each participant, we then averaged these correlation coefficients across all six targets. The resulting correlation coefficient is treated as an indicator of accuracy. For all analyses, we first Fisher-transformed the correlation coefficients for normalization. The reported means, however, are correlation coefficients that were back-transformed into Pearson's *r*.

2.3.6. Specific personality trait accuracy

To measure how well participants (recruiters or students) assessed each of the five NEO dimensions, we calculated for each participant the correlation between his or her ratings of the target on a specific dimension (e.g., extraversion) and the target applicants' ratings of themselves for that specific dimension across all six targets. Note that these correlations are across targets, whereas the personality profile accuracy correlations are across NEO dimensions and then averaged across targets. Again, the resulting correlation coefficients are treated as indicators of accuracy and for all the analyses conducted, we first Fisher-transformed the correlation coefficients for normalization. The reported means, however, are correlation coefficients that were back-transformed into Pearson's *r*.

2.3.7. Lie detection

Participants watched four additional target applicants (two women and two men) each presenting themselves during 1 min, once by providing actual biographical data and once by presenting fictitious biographical data. After having watched each applicant, participants were asked to indicate which of the two presented versions was truthful and which one was the lie. Correct lie detection for each of the four targets was scored with 1, incorrect with 0, and we averaged the scores over the four target applicants (i.e., theoretical and actual range from 0 to 1, $M = .54$, $SD = .26$).

2.3.8. Self-evaluation of assessment skills

We selected items from the detection part of the Skill in Nonverbal Communication Scale (Zuckerman & Larrance, 1979) and developed additional new items to measure the degree to which people think that they are capable of correctly assessing their social interaction partners. Sample items are: 'The first impression I form of a person is often correct,' 'I detect immediately if people lie to me,' or 'I am usually unaware of other people's feelings' (reverse scored). The questionnaire comprised 16 items (four reverse-scored) measured on a scale of 1 = *don't agree at all* to 5 = *completely agree* ($\alpha = .71$). Scores on the items were summed and higher values indicate better self-evaluated assessment skills ($M = 58.50$, $SD = 6.12$).

3. Results

3.1. Personality assessment

To test whether the level of accuracy in assessing the target applicants' personality was above chance level, we

Table 1. Test of better than chance accuracy for each type of accuracy for recruiters and students

Type of accuracy	Overall ($df = 129$)	Recruiters ($df = 49$)	Students ($df = 79$)
Profile accuracy	.29****	.35*** a	.25*** a
Specific trait accuracy			
Openness	.38****	.44*** b	.34*** b
Extraversion	.16****	.01 a	.25*** a
Conscientiousness	.21****	.04 a	.31*** a
Agreeableness	−.16****	−.12+	−.19***
Neuroticism	−.38****	−.34***	−.40***

Note. Values are Pearson's *r*s (back-transformed from Fisher's *r*s), significance testing stemming from a *t* test against 0. Values with the letter 'a' indicate a significant difference between recruiters and students stemming from the 2 (expertise: recruiter vs. student) × 2 (participant gender) ANOVAs and are thus controlled for gender. Values with the letter 'b' indicate a marginally significant difference between recruiters and students (controlled for gender). +$p < .10$; ***$p < .001$; ****$p < .0001$.

calculated simple *t* tests against 0 for the personality profile accuracy and for each of the specific personality trait accuracies separately. We did this for recruiters and students together and separately. We used 0 as the test value because a correlation coefficient of 0 would be expected when participants assessed the traits of the target applicants randomly. Table 1 shows that personality profile accuracy could be assessed significantly better than chance level by both recruiters and students. Also, openness was assessed at better than chance level by both recruiters and students. Extraversion and conscientiousness were assessed at better than chance level only by the students but not by the recruiters. Agreeableness and neuroticism were not assessed correctly, neither by recruiters nor students (negative correlations). Note that the significant results remained significant even after a Bonferroni adjustment of the significance level for multiple tests.

To address whether the differences between the recruiters and the students were significant, we calculated a 2 (expertise: recruiter vs. student) × 2 (participant gender) ANOVA with personality profile accuracy as the dependent variable. Results show a significant expertise main effect, $F(1, 125) = 5.69$, $p = .019$, with recruiters outperforming students. There was a marginally significant participant gender main effect, $F(1, 125) = 2.82$, $p = .096$, with women ($M = .34$) being better personality profile assessors than men ($M = .25$). The interaction effect was not significant, $F(1, 125) = 0.19$, $p = .67$.

To test whether recruiters differed from students with respect to specific personality trait accuracy, we calculated separate 2 (expertise: recruiter vs. student) × 2 (participant gender) ANOVAs with each of the five personality trait dimensions as the dependent variable.

With respect to *openness*, results showed a marginally significant expertise main effect, $F(1, 124) = 2.79$, $p = .098$, with recruiters being better assessors of openness than students. There was neither a significant gender main effect, $F(1, 124) = 1.37$, $p = .25$ (M women = .42, M men = .32), nor a significant interaction effect, $F(1, 124) = 1.35$, $p = .25$.

For *extraversion*, results yielded a significant expertise main effect, $F(1, 125) = 10.72$, $p = .001$, with recruiters being worse assessors of extraversion than students. The participant gender main effect, $F(1, 125) = 0.22$, $p = .64$ (M women = .13, M men = .10), and the interaction effect, $F(1, 125) = 0.08$, $p = .78$, were both not significant.

With respect to *conscientiousness*, a significant expertise main effect emerged, $F(1, 124) = 8.29$, $p = .005$, indicating that students were better assessors of conscientiousness than recruiters. Neither the gender main effect, $F(1, 124) = 1.53$, $p = .22$ (M women = .42, M men = .32), nor the interaction effect, $F(1, 124) = 0.45$, $p = .50$, was significant.

Note that for *neuroticism* the means were negative indicating that participants' assessments of neuroticism increased when targets' self-assessments decreased. The analysis showed no significant expertise main effect, $F(1, 124) = 0.30$, $p = .58$, and no significant interaction effect, $F(1, 124) = 0.38$, $p = .55$. However, a significant participant gender main effect emerged, $F(1, 124) = 4.45$, $p = .037$, showing that women were less inaccurate ($M = -.30$) in assessing neuroticism than men were ($M = -.49$).

As can be observed in Table 1, participants were also not able to assess *agreeableness* correctly (negative correlations). The analysis showed no significant main effects and no significant interaction effect, all Fs < 2.64, all ps > .107.

3.2. Deception detection

To assess whether participants were able to detect the lies of the target applicants, we calculated a simple *t* test against .5 as the value expected by chance (scoring of the lie detection was the average of the four items, either coded as 1 if the answer was correct or as 0 if the answer was incorrect). Recruiters were able to detect lies significantly better than chance ($M = .64$), $t(50) = 4.15$, $p = .0001$, whereas students were not ($M = .47$), $t(79) = 1.00$, $p = .32$.

To test whether this difference was significant, we calculated a 2 (expertise: recruiter vs. student) × 2 (participant gender) ANOVA with lie detection as the dependent variable. Results indicated that recruiters were indeed significantly better at detecting lies in target applicants than were students, $F(1, 125) = 16.00$, $p = .0001$. No other effects were significant (all Fs < 0.83 and all ps > .36).

3.3. Self-evaluation and actual performance

Recruiters' self-evaluations of their assessment skills ($M = 3.76$, $SD = .34$) were higher than students' ($M = 3.58$, $SD = .40$), $t(129) = 2.61$, $p = .01$. We tested whether participants'

self-evaluation of their assessment skills was related to actual assessment performance by correlating scores on the interpersonal sensitivity self-report questionnaire with deception detection performance, personality profile accuracy, and specific personality trait assessment accuracy. Only one significant correlation emerged, self-reported assessment skills were positively related to deception detection, $r(131) = .25$, $p = .004$ (the effect size of this relation was comparable for recruiters, $r = .20$ and for students, $r = .19$).

3.4. Intercorrelations among interpersonal assessment measures

Table 2 shows the intercorrelations among the different assessment skills for both recruiters and students together and separately. The more accurately participants judged the personality profile of the applicants, the better they were at accurately assessing openness and neuroticism in the applicants. However, these relations only held for students (difference between students and recruiters marginally significant for openness and significant for neuroticism). Moreover, accuracy in personality profile assessment was positively related to correctly assessing the applicant's extraversion for students and negatively for recruiters (a significant difference between students and recruiters). Also, correctly assessing the personality profile was marginally significantly negatively related to the correct assessment of conscientiousness but only in recruiters (a significant difference between students and recruiters). Accuracy of assessing openness was positively related to assessing extraversion in both recruiters and students and so was accuracy in assessing openness and accuracy in assessing conscientiousness (however it was not significant for recruiters). The correct assessment of the applicants' extraversion was positively linked to correctly assessing conscientiousness but only for recruiters and not for students (a marginally significant difference between students and recruiters). Also, accuracy in assessing extraversion was negatively related to deception detection (but only when recruiters and students were merged).

Correctly assessing conscientiousness was positively related to deception detection in students but not in recruiters (a significant difference between students and recruiters). And, accuracy in assessing neuroticism was positively related to deception detection in recruiters (no significant difference between recruiters and students).

3.5. Recruiters' experience and training in relation to their assessment performance

For recruiters only, we looked at whether years of job experience and/or their training in recruitment (these variables were not related to each other: $r(50) = -.04$, $p = .78$) were related to accuracy in applicant personality assessment and/or lie detection. Recruiter *job experience* was significantly positively related to accuracy in lie detection, $r(50) = .26$, $p = .027$, and significantly negatively related to accuracy in assessing extraversion, $r(50) = -.36$, $p = .01$ (recall that extraversion could not be assessed accurately by the recruiters). *Recruitment training* was marginally significantly negatively related to accuracy in assessing the personality profile, $r(50) = -.24$, $p = .09$. Note that both recruiter experience and training were unrelated to the recruiters' self-evaluated assessment skills, $r(50) = .09$, $p = .54$; $r(50) = .02$, $p = .88$, training and experience, respectively.

4. Discussion

This study investigated the accuracy of recruiters' and laypersons' first impressions about job applicants based

Table 2. Intercorrelations among the different interpersonal assessment measures

	Openness	Extraversion	Conscientiousness	Agreeableness	Neuroticism	Deception detection
Personality profile	.28*** (.02/.35** a)	−.06 (−.40**/.25* d)	−.05 (−.25+/.18 b)	−.06 (−.20/.01)	.29*** (−.12/.42*** c)	.10 (.01/.07)
Openness		.36**** (.42**/.42***)	.19* (.14/.31**)	.08 (−.06/.15)	.01 (−.06/.03)	.01 (−.24+/.05)
Extraversion			.21* (.29*/−.01 a)	.10 (.23/.05 a)	.11 (.15/.11)	−.22** (−.20/−.16)
Conscientiousness				.10 (.12/.12)	.10 (.15/.08)	−.07 (−.23/.22* b)
Agreeableness					−.14 (−.13/−.16)	−.06 (−.16/−.04)
Neuroticism						.22** (.29*/.17)

Note. Entry on first line of the cell is for all participants ($N = 130$). First entry in parentheses is for recruiters and second entry is for students. N for recruiters is 51 and N for students is 80. a = a marginally significant difference in the correlation coefficients between the recruiter and the student ($p < .10$); b = a significant difference in the correlation coefficients between the recruiter and the student at $p < .05$; c = a significant difference in the correlation coefficients between the recruiter and the student at $p < .01$; d = a significant difference in the correlation coefficients between the recruiter and the student at $p < .001$. +$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$; ****$p < .0001$.

on a thin slice of applicant behavior during a simulated employment interview. We tested accuracy in assessing applicants' personality and deception detection. We also investigated whether recruiters are better assessors than students and how different skills (e.g., personality judgment and deception detection) are related to each other. Moreover, we were interested in whether the recruiters and students were aware of their actual assessment performance. For recruiters, we tested whether their experience or training affected their assessment performance. We do not discuss gender effects below because the ones we found are in line with existing research showing that women outperform men in interpersonal accuracy (Hall, 1984, 1998) and because none of our variables interacted with gender.

Both recruiters and students were able to accurately detect personality profiles; recruiters were better at this than students. The personality profile measures the relative weight of each of the five personality dimensions of the Big Five in a person's assessment. As an example, accuracy is high when the target's actual personality profile (thus the relative importance of each trait dimension in relation to the other trait dimensions) corresponds to the personality profile perceived by the assessor.

Looking at the different personality dimensions separately (Table 1), results showed that openness, extraversion, and conscientiousness could be assessed accurately by students but that agreeableness and neuroticism could not. For extraversion and neuroticism, these results are in line with existing research showing that laypersons are in general accurate at assessing extraversion but not neuroticism (Borkenau & Liebler, 1992; Funder & Colvin, 1988; Gifford, 1994; Watson, 1989). The RAM (Funder, 1995; Funder *et al*., 1995) explains this difference by the fact that extraversion is more readily shown in expressive behavior than neuroticism and thus easier to read for observers. Openness could be assessed accurately in some studies (Borkenau & Liebler, 1992) – consistent with our results – but not in others (Funder & Colvin, 1988; Watson, 1989). That students were able to correctly assess conscientiousness confirms the results in the Borkenau and Liebler and in the Watson studies but disconfirms findings from Funder and Colvin. Agreeableness could not be assessed accurately by our students which is in line with results from Watson and from Funder and Colvin but contradicts findings from Borkenau and Liebler.

Openness was the only trait that recruiters were able to judge accurately. Barrick *et al*. (2000) report accurate assessment of applicants' extraversion, openness, and agreeableness but not of their conscientiousness or neuroticism. Thus, our results confirm these findings with the exception of extraversion and agreeableness.

Conscientiousness is the trait most predictive of future job performance (Barrick & Mount, 1991); however, it could not be assessed accurately by the recruiters. This is in line with previous research on the assessment of job applicants in employment interviews (Barrick *et al*., 2000) but it is surprising that the students in our sample were able to assess conscientiousness accurately. Maybe conscientiousness is a trait that is salient for students given that one of their main preoccupations is to pass exams and obtain good grades. Assuming that students know the grades of at least some of their peers, they are in a good position to learn how different levels of conscientiousness are expressed in behavior. Whether indeed our laypersons were experts on conscientiousness by virtue of them being students remains an open question. This interpretation would suggest that had we targeted another lay population, conscientiousness may not have been assessed correctly.

Although students assessed more trait dimensions correctly than did the recruiters, recruiters were still significantly more accurate in judging the applicants' personality profiles. Note that these are two theoretically different approaches to assess personality and to the best of our knowledge, ours is the first study that compares these two different approaches. We think that recruiters are better at assessing applicants as a whole (i.e., their personality profile) instead of assessing how applicants differ on a given personality dimension. Support for the interpretation that recruiters are experts in Gestalt-like personality assessment – the personality profile – comes from the fact that their assessment performance for personality profiles was unrelated (or negatively related) to the correct assessment of the separate traits (first row of Table 2). In other words, being an expert in assessing the personality profile does not necessarily imply that the single personality aspects can be assessed correctly. The recruiters' job is to recommend the best applicant as a whole and not with respect to one specific trait. Qualitative studies of recruiter decision-making emphasize how recruiters construct a holistic portrait of an applicant based on multiple information sources like the resumé and interview data (Gall, 2000). Thus, recruiters might ponder the different aspects of an applicant's personality dimensions when comparing it with a competence profile for a given job. Fitting the required competence profile for a job necessitates a fit with a personality profile of *one* person. This may make recruiters experts with respect to holistic personality profiles.

We note, however, that the students in our study may not be representative of laypersons as nonexperts drawn randomly from the general population. Students have high intellectual skills which could contribute to their proficiency in this task. We measured cognitive ability with the Wonderlic intelligence questionnaire and recruiters and students did not differ in cognitive ability. However, the Wonderlic could not be administered properly for recruiters (we instructed them to solve

the task in a given time frame but we had no means to check whether they did) which is why we did not include this variable in our analyses.

With respect to deception detection, recruiters were able to detect lies in job applicants, whereas students were not. Although success on a lie detection test depends on how hard the test is, the laypersons in our study confirm existing findings that laypersons are usually unable to correctly identify lies (Ekman, 2001). Our study shows that lie detection seems to be an area of expertise of the recruiters and Table 2 also shows that their skill is unrelated to the correct assessment of personality (as a profile or as separate traits). Moreover, job experience was related to the recruiters' performance in deception detection. Recruiters' (and students') self-evaluation of assessment skills was related to lie detection. Thus recruiters seem to have some insight into how well they detect deception. This finding contradicts what is typically found in the literature (DePaulo *et al*., 1997). Recruiters might know that they need to focus on the nonverbal as opposed to the verbal channel (especially on paralinguistic cues such as laughing or vocal pitch) when trying to detect a lie because the nonverbal channel is more diagnostic in lie detection than the verbal one (Anderson, DePaulo, Ansfield, Tickle, & Green, 1999).

In a nutshell, recruiters are skilled at assessing the personality profile of applicants but laypersons could replace them in this task because they also do a good job. Recruiters are better than laypersons in deception detection. Moreover, recruiters' experience is valuable for this skill.

The present research shows that recruiters cannot simply be replaced by assessment nonexperts (i.e., students). Given the importance of selecting people (possessing a personality profile) instead of traits for a given position and given that many applicants heavily use impression management strategies, it is advantageous to use recruiters for personnel selection because they are the experts in assessing personality profile and in detecting deception.

Acknowledgement

This research was funded by a grant to the first author from the Swiss National Science Foundation (PP001-106528/1).

References

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*, 538–547.

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

Ambady, N., LaPlante, D., & Johnson, E. (2001). Thin-slice judgments as a measure of interpersonal sensitivity. In J. A. Hall, & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 89–101). Mahwah, NJ: Erlbaum.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274.

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untrapped wisdom? *Journal of Nonverbal Behavior*, *23*, 67–88.

Bardack, N. R., & McAndrew, F. T. (1985). The influence of physical attractiveness and manner of dress on success in a simulated personnel decision. *The Journal of Social Psychology*, *125*, 777–778.

Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. *Intelligence*, *13*, 263–287.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimension 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.

Bernieri, F. J., & Gillis, J. S. (1995). The judgment of rapport: A cross-cultural comparison between Americans and Greeks. *Journal of Nonverbal Behavior*, *19*, 115–130.

Bernieri, F. J., & Gillis, J. S. (2001). Judging rapport: Employing Brunswik's lens model to study interpersonal sensitivity. In J. A. Hall, & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement. The LEA series in personality and clinical psychology* (pp. 67–88). Mahwah, NJ: Lawrence Erlbaum Associates.

Blackman, M. C. (2002). Personality judgment and the utility of the unstructured employment interview. *Basic and Applied Social Psychology*, *24*, 241–250.

Blackman, M. C., & Funder, D. C. (2002). Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment*, *10*, 109–116.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 645–657.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology*, *82*, 546–561.

Cole, M. S., Feild, H. S., & Giles, W. F. (2003). Using recruiter assessments of applicants' resume content to predict applicant mental ability and big five personality dimensions. *International Journal of Selection and Assessment*, *11*, 78–88.

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In J. A. Hall, & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 47–65). Mahwah, NJ: Erlbaum.

Colvin, C. R., Vogt, D., & Ickes, W. J. (1997). Why do friends understand each other better than strangers do? In W. J. Ickes (Ed.), *Empathic accuracy* (pp. 169–193). New York: Guilford Press.

Costa, P. T., & McCrae, R. R. (1992). *Professional manual. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor*

Inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources Inc.

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

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy–confidence correlation in the detection of deception. *Personality and Social Psychology Review*, *1*, 346–357.

Dipboye, R. L. (1982). Self-fulfilling prophecies in the selection-recruitment interview. *Academy of Management Review*, *7*, 579–586.

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

Dougherty, T. W., Turban, D. B., & Callender, J. C. (1994). Confirming first impressions in the employment interview: A field study of interviewer behavior. *Journal of Applied Psychology*, *79*, 659–665.

Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton.

Ellis, A. P. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, *87*, 1200–1208.

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

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., & Colvin, C. R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, & J. A. Johnson (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego, CA: Academic Press.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, *69*, 656–672.

Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality*, *61*, 457–476.

Gall, J. M. (2000). Equilibre de jugement et jugement d'équilibre. Les faits, les mots, le corps, dans l'entretien de recrutement. [Equilibrium of judgment and judging equilibrium. Facts, words, and the body in the selection interview]. *Revue de Gestion des Ressources Humaines*, *36*, 69–83.

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology*, *66*, 398–412.

Gravesa, L. M., & Powell, G. N. (1988). An investigation of sex discrimination in recruiters' evaluations of actual applicants. *Journal of Applied Psychology*, *73*, 20–29.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.

Hall, J. A. (1998). How big are nonverbal sex differences: The case of smiling and sensitivity to nonverbal cues. In D. J. Canary, & K. Dindia (Eds.), *Sex differences and similarities in communication: Critical essays and empirical investigations of sex and gender in interaction* (pp. 155–177). Mahwah: Erlbaum.

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, *86*, 897–913.

Ickes, W. J. (Ed.). (1997). *Empathic accuracy*. New York: Guilford Press.

Ickes, W. J. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus Books.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, *87*, 530–541.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, *62*, 88–97.

Kristof-Brown, A. L., Zimmerman, E., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, *58*, 281–342.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology*, *92*, 1638–1656.

Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, *86*, 255–264.

Lievens, F., & Peeters, H. (2008). Interviewers' sensitivity to impression management tactics in structured interviews. *European Journal of Psychological Assessment*, *24*, 174–180.

Macan, T. H., & Dipboye, R. L. (1990a). The relationship of interviewers' preinterview impressions to interview outcomes. *Personnel Psychology*, *43*, 745–768.

Macan, T. H., & Dipboye, R. L. (1990b). The relationships of interviers' preinterview impressions to selection and recruitment outcomes. *Personnel Psychology*, *43*, 745–758.

Malone, B. E., & DePaulo, B. M. (2001). Measuring sensitivity to deception. In J. A. Hall, & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement. The LEA series in personality and clinical psychology* (pp. 103–124). Mahwah, NJ: Lawrence Erlbaum Associates.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683–729.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, *79*, 272–280.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality*, *71*, 465–493.

Nowicki, S., Jr, & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, *18*, 9–35.

OReilly, C. A., Chatman, J., & Caldwell, D. E. (1991). People and organizational culture: A profile comparison approach to

assessing person-organization fit. *Academy of Management Journal*, *34*, 487–516.

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, *55*, 1–81.

Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, *74*, 441–472.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: The Johns Hopkins University Press.

Schallberger, U., & Venetz, M. (1999). *Kurzversionen des MRS-Inventars von Ostendorf (1990) zur Erfassung der fünf grossen Persönlichkeitsfaktoren*. Zürich, Switzerland: Psychologisches Institut der Universität Zürich.

Schmid Mast, M., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior*, *28*, 145–165.

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.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing in social interaction. *Journal of Personality and Social Psychology*, *36*, 1202–1212.

Townsend, R. J., Bacigalupi, S. C., & Blackman, M. C. (2007). The accuracy of lay integrity assessments in simulated employment interviews. *Journal of Research in Personality*, *41*, 540–557.

Van Iddekinge, D. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, *90*, 536–552.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, *57*, 120–128.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189–205.

Zuckerman, M., & Larrance, D. T. (1979). Individual differences in perceiving encoding and decoding abilities. In R. Rosenthal (Ed.), *Skill in nonverbal communication: Individual differences* (pp. 171–203). Cambridge, MA: Delgeschlager, Gunn, & Hain.