Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from

an assessment center and a structured interview

Cornelius J. König, Klaus G. Melchers, and Martin Kleinmann

Universität Zürich

Gerald M. Richter

Novartis Behring

Ute-Christine Klehe

Universität Zürich and Universiteit van Amsterdam

Author Note

Cornelius J. König, Klaus G. Melchers, and Martin Kleinmann, Psychologisches Institut, Universität Zürich, Switzerland; Gerald M. Richter, Novartis Behring, Marburg, Germany; Ute-Christine Klehe, Psychologisches Institut, Universität Zürich, Switzerland, and Universiteit van Amsterdam, the Netherlands.

Correspondence concerning this article should be addressed to Cornelius J. König, Institut für Psychologie, Universität Zürich, Rämistrasse 62, CH-8001 Zürich, Switzerland (phone +41-44-63 44132). Electronic mail may be sent to c.koenig@psychologie.unizh.ch.

Abstract

In selection procedures like assessment centers (ACs) and structured interviews, candidates are often not informed about the targeted criteria. Previous studies have shown that candidates' ability to identify these criteria (ATIC) is related to their performance in the respective selection procedure. However, past research has studied ATIC in only one selection procedure at a time, even though it has been assumed that ATIC is consistent across situations, which is a prerequisite for ATIC to contribute to selection procedures' criterion-related validity. In this study, 95 candidates participated in an AC and a structured interview. ATIC scores showed cross-situational consistency across the two procedures and accounted for part of the relationship between performance in the selection procedures. Furthermore, ATIC scores in one procedure predicted performance in the other procedure even after controlling for cognitive ability. Implications and directions for future research are discussed.

Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from

an assessment center and a structured interview

People usually try to control how others perceive them in social interactions (Hogan,

1991; Hogan, Hogan, & Roberts, 1996). This should particularly be the case in a selection

context, in which individuals attempt to present themselves positively in order to increase

their chances of receiving a job offer (Motowidlo, 1999). However, positive self-presentation

is often easier said than done because it is not always clear what "positive" entails.

Accordingly, the current study addresses the extent to which candidates can identify the

criteria used to evaluate their performance in two selection procedures, an assessment center

(AC) and a structured interview, and investigates how correct identification of those criteria is

related to their performance.

Selection procedures differ in the degree to which they reveal to applicants the

required behavior: In a cognitive ability test, applicants know that their performance will be

evaluated with regard to the number of correct solutions to the different items, but in an

interview or an AC, finding out the targeted performance dimensions is often much less

straightforward: For example, when interviewees are asked about a hypothetical conflict

situation, or AC participants face a conflict situation, it is usually not obvious whether they

should argue their point of view and negotiate fiercely or whether it would be more

advantageous to indicate that they are endeavoring to understand the other party's point of

view and wish to search for compromises. As applicants usually do not learn beforehand

whether one kind of behavior or the other will be evaluated more positively in the respective

interview question or AC task, applicants may receive better or worse evaluations depending

on the degree to which they are able to discern the targeted performance dimensions. Thus, if

applicants have the ability to identify the criteria (ATIC) that are used for evaluating their

performance, this should considerably influence their success in these and other personnel

selection situations in which these criteria are relatively nontransparent. In line with this suggestion, researchers have found that participants' ATIC in an AC predicts their performance in this AC (Kleinmann, 1993; Preckel & Schüpbach, 2005), that interviewees' ATIC in a structured interview predicts their performance in the interview (Melchers, Kleinmann, Richter, König, & Klehe, 2004), and that candidates' ATIC in an integrity test predicts their integrity scores (König, Kleinmann, Melchers, Richter, & Klehe, 2006).

Furthermore, it seems quite likely that individuals' ATIC is not limited to performance in the interview but also influences their later performance on the job. Accordingly, this ability might contribute to the good criterion validity of structured interviews and ACs (shown, for example, in meta-analyses by Gaugler, Rosenthal, Thornton, & Bentson, 1987; Huffcutt, Conway, Roth, & Klehe, 2004; Taylor & Small, 2002). Individuals who correctly interpret cues in their environment and use those cues for deciding on their plans of action are likely to do so not only when applying for a job but also once they are working in that job. In other words, ATIC might also be important for later job performance as in many work situations it is not exactly clear what other people (e.g., customers, managers, colleagues) actually want (cf., Beehr & Juntunen, 1990). Being able to identify what others consider to be important (i.e., their evaluation criteria) should help to create interactions that are satisfying for both parties involved. Thus, ATIC might be considered as a specific aspect of social intelligence (Peterson & Seligman, 2004; Zaccaro, 2002).

The suggestion that ATIC might (at least in part) account for the good criterion-related validity of ACs and interviews might also help to solve the validity paradox from which both of these selection procedures suffer. Although ACs and interviews show good criterion-related validity, past research has repeatedly found that ACs (e.g., Lance, Lambert, Gewin, Lievens, & Conway, 2004; Woehr & Arthur, 2003) as well as interviews (e.g., Huffcutt, Conway, Roth, & Stone, 2001; Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004) suffer from serious internal construct validity problems. Both ACs and interviews do not seem to

measure the intended constructs. Often convergent validities (e.g., correlations between ratings of the same constructs in different AC exercises or different interview components) are found to be lower than discriminant validities (e.g., correlations between ratings of different constructs in the same AC exercise or interview components). Thus, although both procedures measure something that is important for later job performance, neither seems to measure the intended constructs (i.e., the performance dimensions that are supposedly evaluated).

Until now, research on applicants' ATIC has been restricted in that researchers have studied ATIC in only one personnel selection procedure at a time. However, to support the claim that ATIC contributes to the criterion-related validity of ACs or interviews, several open questions have to be answered. First, it has to be shown that ATIC is not bound to a specific situation or selection procedure, but that it is a more general ability, because ATIC can only account for the predictive validity of selection procedures if it is not just a very specific skill for one specific situation (e.g., one interview). Second, if ATIC generalizes across situations, then measures of candidates' ATIC from one situation should allow for a prediction of their performance in other situations. And third, if ATIC as a general factor influences candidates' performance in different contexts, the correlation between their performance in those different contexts should decrease if ATIC is partialled out from this relationship.

The present study investigated these questions in the context of a simulated selection context in which participants took part in two different selection procedures, an AC and a structured interview. This setting allowed us to use candidates' performance in one procedure as a proxy criterion, for which we assessed how well it can be predicted on the basis of their ability to identify the criteria that were used during the other selection procedure. Our first objective was to confirm that assessments of ATIC show cross-situational consistency, meaning that measures of ATIC from different selection procedures correlate substantially

with one another (Hypothesis 1). Furthermore, we expected that candidates' ATIC assessed in

one selection procedure would predict their performance in another procedure (Hypothesis 2).

Given that we regard ATIC as an ability, it also seemed necessary to relate it to

cognitive ability, a main predictor of job performance (e.g., Salgado, Anderson, Moscoso,

Bertua, & de Fruyt, 2003; Schmidt & Hunter, 1998). This is even more important bearing in

mind that performance in ACs (Collins et al., 2003; Scholz & Schuler, 1993) and interviews

(Huffcutt, Roth, & McDaniel, 1996; Salgado & Moscoso, 2002) correlates with cognitive

ability. If ATIC was only a situated facet of cognitive ability, it would not be surprising to

find that ATIC assessed in one procedure is related to performance in another procedure. It is

therefore important to show that ATIC in one procedure predicts performance in another

procedure even after controlling for cognitive ability (Hypothesis 3). And finally, we expected

ATIC to account for at least part of the correlation between candidates' performance in the

two selection procedures (Hypothesis 4).

Methods

Participants

Ninety-five participants (47 males, 48 females) took part in a two-day application

training program that was organized by a German university and a regional branch of the

German Bureau of Labor Exchange. Participants were recent or prospective university

graduates who were currently applying for jobs or would be doing so in the near future.

Participants had been attending university on average for four years and ten months ($SD =$

1.82) and 27.4% had already obtained the German equivalent of a Master's degree. Nearly

half of them (47.4%) had prior work experience. Participants' median age was 26 years (range

between 21 and 36 years). The majority of participants had a background in business or

economics (41.3%) or in natural science (21.7%). Participants had to pay a small fee for

participating in the training to cover part of the costs and to ensure their commitment. They

were not given any information concerning the objectives of the study.

AC and Interview Development

The application training consisted of three AC exercises, a cognitive ability test and a structured interview that included past-oriented questions (cf. Janz, 1989) and future-oriented questions (cf. Latham, Saari, Pursell, & Campion, 1980). The AC and the interview were designed to be suitable for selecting candidates for a hypothetical management trainee position. Based on a job analysis conducted for an earlier study, three dimensions were chosen to be assessed in the AC and the interview: Systematic Planning (prioritizing tasks, making plans for tasks and projects, making appointments in due time and allocating tasks), Leadership Behavior (striving for and assuming responsibility for tasks and groups, coordination of teams, and arguing for one's point of view in groups), and Cooperation (assisting others with problems they may have, considering the needs of others, being prepared to compromise with others, and mediating between diverging points of view). These three dimensions correspond to the thinking-power-feeling taxonomy suggested for ACs (Kolk, Born, & van der Flier, 2004). As in many ACs (cf. Collins et al., 2003), a group discussion with assigned roles, a group discussion without assigned roles, and an in-basket task were used as AC exercises. For the interview, two subject matter experts collected a pool of 34 past-oriented and 34 future-oriented questions (most of which had already been used in previous studies). Scoring guides provided behavioral anchors for outstanding (5), acceptable (3), and unacceptable (1) answers for each interview question. Behavioral checklists were provided for each dimension in the AC. The AC exercises and interview questions were pre-tested with ten Master's level work and organizational psychology students who had experience in serving as AC observers and/or interviewers. The exercises, interview questions, and behavioral anchors were modified where needed (e.g., to increase the understandability) and interview questions that were deemed unsuitable to assess the intended dimensions were deleted from the item pool, leaving a final set of 12 past-oriented and 12 future-oriented questions for the actual interview (4 past-oriented and 4 future-oriented

questions per dimension). In the AC, all three dimensions were assessed in each exercise, whereas in the interview, each question targeted only one dimension.

Rater Training

Thirty-two Master's level work and organizational psychology students served as observers for the AC and as interviewers for the structured interview. They received a one-day rater training session during which they learned about the AC and the interview as well as about definitions and behavioral examples for all three dimensions. The training also included information about typical rating errors, frame-of-reference elements (cf. Woehr & Huffcutt, 1994), and advice on how to give behavioral feedback to participants. The raters were not told about the objectives of the study.

Assessment of Participants' ATIC

ATIC was measured in a very similar manner to Kleinmann (1993). Participants had to fill out a questionnaire following completion of each of the three AC tasks and each of the two interview components. In these questionnaires, they were asked to write down the hypotheses that they had entertained during the actual interview questions or AC exercises regarding what a certain question or exercise was intended to assess. The questionnaires provided space for a maximum of two hypotheses for each interview question and up to six hypotheses for each AC exercise. All interview questions were repeated in the questionnaires to prevent memory problems. Participants were told that their responses would not be rated for the training but served to improve the application training.

To assess the degree to which participants' hypotheses corresponded to the intended dimensions, we again employed a procedure similar to the one used in the AC study by Kleinmann (1993): At the end of the application training, participants received a list of six dimensions that are commonly used for ACs and interviews. Three of these dimensions were the dimensions used in this study and three were distractor dimensions (Job Knowledge and Experience, Self-Confidence, and Acquisition and Handling of Information). A list of

behavioral examples for each dimension was also given to participants. After having been introduced to the six dimensions, participants' questionnaires, containing their answers on assumed dimensions, were returned to them. They were asked to indicate which of their ideas corresponded to which dimension. In addition (and unlike Kleinmann, 1993), they were also asked to rate the strength of this correspondence on a scale from 1 (= *fits somewhat*) to 4 (= *fits completely*). Participants could also indicate that an idea did not correspond to any of the dimensions.

For the measurement of candidates' ATIC, we focused on ratings of hypotheses corresponding to the correct dimensions, and took the rating from the hypothesis for which candidates had indicated the highest fit for that dimension. If a dimension was not correctly identified at all, a score of 0 was assigned to the respective rating. This led to ATIC values for interview questions or AC dimensions ranging from 0 (= no fit with the correct dimension) to 4 (= perfect fit with the correct dimension). We then calculated the mean of the correspondence ratings across the different questions or different dimensions in the AC exercises and used these as candidates' overall ATIC score in the interview or in the AC, respectively.

Cognitive Ability

Cognitive ability was measured with two modules of the Intelligenz-Struktur-Test 2000 (IST 2000, Amthauer, Brocke, Liepmann, & Beauducel, 1999; Beauducel, Brocke, & Liepmann, 2001; see also Bühner, Krumm, Ziegler, & Schmidt-Atzert, 2006), a widely-used (cf. Schorr, 1995) and valid German intelligence test (Hülsheger, Maier, Strumpp, & Muck, 2006). The first module contains three subtests (Sentence completion, Analogies, and Similarities) and measures verbal reasoning. The second module also contains three subtests (Figures, Cubes, and Matrices) and measures figural reasoning. The verbal reasoning module and the figural reasoning modules are reported by Amthauer et al. to be highly correlated with *g* measured by a Raven test (.54 and .50, respectively), whereas the numerical module

correlates only at .09 with *g*. Thus, we did not use this third module. Amthauer et al. report a coefficient alpha of .88 for the verbal module and .87 for the figural module. For the present study, participants' scores for the IST 2000 were determined on the basis of norm values for people who were eligible to study at a German university.

Procedure

Participants received a fictitious job advertisement for a management trainee position within a large technology corporation prior to the training, and were asked to prepare a written application for the job described. They also learned that the selection procedures would focus on this position. The job advertisement was similar to real advertisements and included hints concerning the relevant performance dimensions for the job. This information was formulated specifically for the dimensions we intended to assess in the AC and the interview. For Leadership Behavior, for example, the advertisement stated that applicants should be prepared to take responsibility for themselves as well as for others.

The actual training consisted of general information sessions about personnel selection, the AC and the structured interview as well as of the cognitive ability test. Four participants took part in each group discussion and were rated by four raters. For the in-basket task, participants had to work out a written solution that was then evaluated by two raters. Finally, each participant was interviewed individually by a panel of two interviewers. One interviewer read the questions to the participant, and both interviewers independently recorded and scored the responses on the basis of the scoring guide. Raters observed a candidate either in the AC or in the interview, with no candidate being observed by the same raters in both selection procedures. This was done to ensure that a correlation between performance in the AC and in the interview was not attributable to common rater variance. The raters were not given access to the participants' written applications before the AC and the interview were finished, so that prior knowledge of them could not influence their ratings.

At the end of the application training, raters discussed their ratings for the AC dimensions or for answers to interview questions if these ratings were two or more points apart from each other (on the 5-point scale). Most differences could be resolved after a short debate even though there was no requirement that raters had to agree. For the later analyses, average ratings of these final ratings were used. In addition, applicants received feedback on their performance in the AC and the interview and on their written applications.

Results

Descriptive information and correlations between the variables from this study are shown in Table 1.

Reliability of the Different Measures

The average interrater agreement for the overall ratings for each AC task (averaged across all 3 dimensions) was .69 and .75 for the two group discussions and .90 for the in-basket. These values are comparable to the values from the meta-analysis by Collins et al. (2003). The average interrater agreement (i.e., the average correlation between two raters) for the overall interview ratings (averaged across all 24 questions) was .88 and coefficient alpha for the interview was .72. These values are comparable to meta-analytically derived values for structured interviews (Conway, Jako, & Goodman, 1995). Given that these interrater agreement values reflect the reliability of only one rater, the Spearman-Brown prophecy formula was used to calculate the reliability of the average ratings for two or four raters, respectively. As each of the two group discussions was rated by teams of four observers, the reliability of the overall ratings was .90 and .92, respectively. The interview and the in-basket were rated by two raters, and the reliability of the overall interview rating was .94 and of the overall in-basket rating .95.

To assess the reliability of the two ATIC measures, we determined their internal consistency by calculating coefficient alphas, which were .78 for the interview and .59 for the

AC. These values are comparable to results from earlier studies (Kleinmann, 1993; Melchers et al., 2004). The larger coefficient alpha for the interview stems only from the fact that the ATIC score in the interview was based on a larger number of items than the ATIC score in the AC (24 interview questions versus 9 AC dimension ratings).

----

insert Table 1 about here

Correlation Between ATIC Scores Across Different Selection Procedures

To assess the cross-situational consistency of participants' ATIC, we calculated the correlation between their overall ATIC scores from the AC and the interview. As predicted by Hypothesis 1, ATIC scores between the two selection procedures correlated significantly with each other, $r = .40$, $p < .01$ ($N = 95$ for this and all subsequent analyses unless indicated otherwise). Furthermore, we also assessed the extent to which this correlation may be underestimated due to the somewhat limited reliability of the ATIC measures. Accordingly, we also calculated the true correlation between participants' ATIC in the two procedures by correcting for unreliability in both measures. Insertion of the respective values for coefficient alpha in the disattenuation formula raised the correlation to $r = .59$, thereby indicating common variance among ATIC scores across selection procedures.

Predictive Power of Participants' Ability to Identify Criteria for their Performance in Another Selection Procedure

As can be seen in Table 1, ATIC scores from the AC correlate significantly with participants' performance in the AC, and ATIC scores from the interview correlate significantly with participants' performance in the interview. This parallels previous findings concerning the impact of ATIC for performance (Kleinmann, 1993; König et al., 2006; Melchers et al., 2004; Preckel & Schüpbach, 2005). However, concerning the question that ATIC not only predicts performance in the specific situation for which an ATIC score is

determined but also performance in other situations, we found that ATIC scores from the interview were significantly correlated with performance in the AC, $r = .34$, $p < .01$, and ATIC scores from the AC were significantly correlated with performance in the interview, $r = .29$, $p < .01$. Thus, these results confirmed Hypothesis 2.

To test Hypothesis 3, we conducted two hierarchical multiple regressions in order to test whether ATIC scores from one selection procedure predict performance in the other selection procedure even after controlling for cognitive ability. In the first step, the two measures of cognitive ability were included as predictors. In the second step, we included the ATIC score from the interview to predict performance in the AC and the ATIC score from the AC to predict performance in the interview. Table 2 shows the results of these analyses. It can be seen that cognitive ability accounted for only a limited amount of variance in the first step of either regression. In line with the zero-order correlations from Table 1, the analyses showed that only the verbal aspect of cognitive ability was a significant predictor in these regression analyses, but the figural aspect was not. More importantly, however, ATIC scores from one selection procedure accounted for significant amounts of variance in predicting participants' performance in the other selection procedure even after controlling for cognitive ability, thereby confirming Hypothesis 3.

----

insert Table 2 about here

ATIC as a Source for the Correlation Between Performance in the AC and the Interview

Hypothesis 4 predicted that ATIC accounts for at least part of the relationship between two selection procedures. To test this hypothesis, we calculated the partial correlation between participants' performance in the interview and their performance in the AC by partialling out the two ATIC scores. This partial correlation was $r = .25$, $p < .05$, thus lower than the zero-order correlation of $r = .38$, $p < .01$. To test whether this difference between the

zero-order and the partial correlation was statistically significant, we used a procedure suggested by Olkin and Finn (1995) and extended by Graf and Alf (1999). This procedure showed that the 95% confidence interval for this difference did not include zero but ranged from .03 to .22. Thus, statistically controlling for the impact of ATIC in fact significantly lowered the correlation between AC and interview performance to a significant amount, thereby lending support to Hypothesis 4.

For purposes of comparison, we also assessed the degree to which partialling out cognitive ability would influence the correlation between AC and interview performance. It emerged that the partial correlation in this case was $r = .34$, $p < .01$ ($N = 94$), which was only slightly lower than the zero-order correlation of $r = .39$, $p < .01$ ($N = 94$). Accordingly, the difference between the two coefficients, which was .05, was far from significant.

Discussion

Taken together, this study revealed that people's ability to identify evaluation criteria in one situation is related to a comparable measure of this ability in another situation and, more importantly, also to the performance in that other situation. Furthermore, ATIC had considerable predictive power across situations even after controlling for cognitive ability. Finally, ATIC was found to account for at least parts of the relationship between AC and interview performance.

The study supports the claim that ATIC may partially account for the predictive validity of selection procedures. If ATIC was only important in the context of a specific selection procedure, it could not be assumed to have effects on the job and, consequently, on job performance. Or to put it another way, ATIC can only be responsible for the predictive validity of personnel selection procedures if it shows some stability across different situations. Previous studies, however, were not able to investigate this issue because ATIC and performance were assessed for one procedure only (Kleinmann, 1993; König et al., 2006;

Melchers et al., 2004; Preckel & Schüpbach, 2005). By contrast, the present study assessed ATIC in both an AC and an interview and showed that ATIC does indeed have some predictive validity across procedures. The present study therefore also provides a first step with regard to assessing the criterion-related validity of participants' ATIC by using the performance in one selection procedure as a proxy criterion for the ATIC measure from the other procedure.

This study also showed that ATIC is not mere cognitive ability. Even though ATIC is conceptualized and measured as an ability construct (and hence the A in ATIC), controlling for cognitive ability did not eliminate the effects of ATIC. Cognitive ability was found to be moderately related to AC and to interview performance, thus replicating earlier findings (Collins et al., 2003; Huffcutt et al., 1996), yet ATIC explained variance beyond cognitive ability. On the one hand, cognitive ability might help a person to identify evaluation criteria. For example, people with high cognitive ability might have more mental capacities to think about the evaluation criterion of a particular interview question and simultaneously consider which options might fit this criterion. On the other hand, ATIC seems to be more than what which is measured by conventional cognitive ability tests. One potential reason for the finding that ATIC is a better predictor of performance (even across different selection procedures) than cognitive ability could be that it is more situation specific and a better measure of the actual mental processes taking place in candidates' minds during an AC or an interview.

The findings of this study could also be used to explain why ACs and interviews predict job performance beyond cognitive ability (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Dayan, Kasten, & Fox, 2002): If ATIC influences who performs better and who performs worse in an AC, an interview, and later on the job, then it could be responsible for additional predictive variance that is not included in the cognitive ability test measure.

It might be argued that ATIC is just a facet of faking. However, such a position is erroneous in that ATIC can clearly be differentiated from faking (e.g., Donovan, Dwight, &

Hurtz, 2002). ATIC is mainly a cognitive concept. If you are able to identify the evaluation criteria in a personnel selection situation, you can orientate your behavior towards these criteria. Such an orientation can be achieved, for example, by talking about a past situation in which you mastered a particular problem in an appropriate way. However, if you are motivated to fake, you might talk about the past situation in a distorted way or even make up such a situation, so that the recruiter does not notice that you had not actually mastered this situation (i.e., you give a faked answer). Thus, ATIC is a precondition for effective faking if applicants want to fake, but it is also the prerequisite for effective self-presentation without faking. Furthermore, ATIC is regarded as a positive personal attribute (cf. Preckel & Schüpbach, 2005), whereas faking typically has a negative connotation in the literature (e.g., Ones & Viswesvaran, 1998).

Limitations and Future Research

Although we believe that the present results are informative and provide important directions for future research, some limitations should be noted. First, our data were obtained in the course of an application training program and not from a sample of applicants for a real job. However, our sample can be assumed to be similar to many applicant samples for entry-level jobs for university graduates. Many of our participants had already finished their university degree or were about to finish it and used the application training to prepare for the selection procedure for a job for which they had already applied or planned to apply.

Second, we measured only verbal reasoning and figural reasoning as cognitive abilities, using two modules of the IST 2000. It could be questioned whether the relationship between ATIC and cognitive abilities had been greater if a more *g*-loaded measure of general mental ability would have been used.

Third, our raters/interviewers were psychology students and not human resource managers, who might have given different performance ratings. However, in an assessment center study with videotaped candidates, Lievens (2001) found that ratings obtained from

student observers were quite similar to the ratings obtained from a sample of managers--if anything, students distinguished somewhat better between the various dimensions, meaning that their ratings had slightly better construct validity. In addition, many human resource managers do not hold degrees in HR management, industrial/organizational psychology or organizational behavior (see, e.g., Hoque & Noon, 2001), and having gone through such academic training seems to be beneficial for observational skills (Sagie & Magnezy, 1997) and, at least in the case of ACs, also for construct- and criterion-related validity (Gaugler et al., 1987; Woehr & Arthur, 2003).

Fourth, our arguments regarding the importance of ATIC are only valid if the evaluation dimensions are not made transparent. Sometimes, candidates are told what the dimensions in an AC or interview are. If an AC or an interview is conducted transparently, ATIC is not needed anymore and the AC or the interview should become easier. In line with this argument, participants in transparent ACs (Kleinmann, Kuptsch, & Köller, 1996) achieve higher ratings than participants in nontransparent ACs.

Even though this study contributes to the growing evidence that ATIC is an important construct for personnel selection (e.g., Kleinmann, 1993; König et al., 2006; Melchers et al., 2004; Preckel & Schüpbach, 2005), several questions for future research remain. One such question addresses the nomological network of ATIC. We know from the current research that ATIC and cognitive ability are correlated and from Kleinmann (1997) that ATIC is correlated with social judgment skills. However, we do not know which additional constructs it is also related to. Such research on the nomological network of ATIC could also shed new light on the question why AC and interview scores are related to other constructs (cf. Collins et al., 2003; Salgado & Moscoso, 2002). For example, Salgado and Moscoso (2002) report meta-analytic evidence that both conventional and structured interviews are related to job experience. It might be easier for applicants with much work experience to identify evaluation criteria because they know better what is required for doing the job. Consequently, they might

get higher ATIC scores, which leads to better performance in the interview. As another example, ATIC might be related to the Big Five construct Openness to Experience, in particular to the facet "imagination", because this facet is measured with items like "I do not have a good imagination" (negatively keyed item from the International Personality Item Pool, cf. Goldberg et al., 2006). A good imagination might be helpful for figuring out what recruiters like to see or hear. Such a correlation could explain why AC ratings are correlated with Openness to Experience (Collins et al., 2003). Thus, future research could try to test whether ATIC at least partially mediates the relationship between AC or interview performance and variables of the nomological net of ACs /interviews.

Another open question concerns the predictive validity in real field settings. The current study relied on a proxy criterion (i.e., performance in a second nontransparent personnel selection procedure). Future research should therefore look at the correlation between ATIC and real job performance. Such research is needed to assess whether ATIC is indeed responsible for the predictive validity of a personnel selection procedure as suggested by the present study.

Practical Implications

If ATIC is at least partly responsible for the criterion-related validity of personnel selection procedures, recruiters should think carefully about the degree to which they give hints about their evaluation criteria, as the criterion-related validity might decrease if evaluation criteria are made transparent to applicants. Furthermore, it is important to ensure that some applicants are not given more information (or more hints from which these criteria might be inferred) than others with regard to the evaluation criteria. It would be unfair if recruiters provided more information about their company and their values at a job fair at one college than at another college.

Additionally, ATIC may be a construct of interest for personnel selection professionals in itself. If ATIC is assessed during selection procedures, it may help practitioners to gain a

better understanding of why some participants handled a given selection situation better than others.

References

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000 [Intelligence-Structure-Test 2000]*. Göttingen, Germany: Hogrefe.

Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: Facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences, 30*, 977-994.

Beehr, T. A., & Juntunen, D. L. (1990). Promotions and employees' perceived mobility channels: The effects of employee sex, employee group, and initial placement. *Human Relations, 43*, 455-472.

Bühner, M., Krumm, S., Ziegler, M., & Schmidt-Atzert, L. (2006). Ist der I-S-T 2000 R Rasch-skalierbar? [Is the I-S-T 2000 R Rasch scaleable?]. *Diagnostica, 52*, 119-130.

Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment, 11*, 17-29.

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*, 565-579.

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325-351.

Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center: An efficient tool or a hammer to kill a fly? *Personnel Psychology, 55*, 827-849.

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2002). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16*, 81-106.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of

assessment center validity. *Journal of Applied Psychology, 72*, 493-511.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et

al. (2006). The International Personality Item Pool and the future of public-domain

personality measures. *Journal of Research in Personality, 40*, 84-96.

Graf, R. G., & Alf, E. F., Jr. (1999). Correlations redux: Asymptotic confidence limits for

partial and squared multiple correlations. *Applied Psychological Measurement, 23*,

116-119.

Hogan, R. (1991). Personality and personality measurement. In L. M. Hough (Ed.), *Handbook

of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 873-919). Palo Alto,

CA: Consulting Psychologists Press.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment

decisions: Questions and answers. *American Psychologist, 51*, 469-477.

Hoque, K., & Noon, M. (2001). Counting angels: A comparison of personnel and HR

specialists. *Human Resource Management Journal, 11*(3), 5-22.

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U.-C. (2004). Evaluation and

comparison of the situational and behavior description interview formats.

International Journal of Selection and Assessment, 12, 262-273.

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.

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of

cognitive ability in employment interview evaluations: Moderating characteristics and

implications for incremental validity. *Journal of Applied Psychology, 81*, 459-473.

Hülsheger, U. R., Maier, G. W., Strumpp, T., & Muck, P. M. (2006). Vergleich

kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von

Ausbildungserfolg in Deutschland: Ergebnisse einer Metaanalyse [Comparing criterion-related validities of different intelligence tests for the prediction of training success in Germany: A meta-analysis. *Zeitschrift für Personalpsychologie, 5*, 154-162.

Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. In G. R. Ferris (Ed.), *The employment interview: Theory, research, and practice* (pp. 158-168). Thousand Oaks, CA: Sage.

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*, 988-993.

Kleinmann, M. (1997). Transparenz der Anforderungsdimensionen: Ein Moderator der Konstrukt- und Kriteriumsvalidität des Assessment-Centers [Transparency of the required dimensions: A moderator of assessment centers' construct and criterion validity]. *Zeitschrift für Arbeits- und Organisationspsychologie, 41*, 171-181.

Kleinmann, M., Kuptsch, C., & Köller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An International Review, 45*, 67-84.

Kolk, N. J., Born, M. P., & van der Flier, H. (2004). A triadic approach to the construct validity of the assessment center: The effect of categorizing dimensions into a feeling, thinking, and power taxonomy. *European Journal of Psychological Assessment, 20*, 149-156.

König, C. J., Kleinmann, M., Melchers, K. G., Richter, G. M., & Klehe, U.-C. (2006). The relationship between the ability to identify evaluation criteria and integrity test scores. *Psychology Science, 48*, 369-377.

Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology, 89*, 377-385.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427.

Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior, 22*, 203-221.

Melchers, K. G., Kleinmann, M., Richter, G. M., König, C. J., & Klehe, U.-C. (2004). Messen Einstellungsinterviews das, was sie messen sollen? Zur Bedeutung der Bewerberkognitionen über bewertetes Verhalten [Do selection interviews measure what they intend to measure? The impact of applicants' cognitions about evaluated behavior]. *Zeitschrift für Personalpsychologie, 3*, 159-169.

Motowidlo, S. J. (1999). Asking about past behavior versus hypothetical behavior. In M. M. Harris (Ed.), *The employment interview handbook* (pp. 179-190). Thousand Oaks, CA: Sage.

Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin, 118*, 155-164.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245-269.

Peterson, C., & Seligman, M. E. P. (2004). Social intelligence (emotional intelligence, personal intelligence). In C. Peterson & M. E. P. Seligman (Eds.), *Character strengths and virtues: A handbook and classification* (pp. 337-353). Washington, DC: American Psychological Association.

Preckel, D., & Schüpbach, H. (2005). Zusammenhänge zwischen rezeptiver Selbstdarstellungskompetenz und Leistung im Assessment-Center [Correlations between receptive self-presentation competence and performance in an assessment center]. *Zeitschrift für Personalpsychologie, 4*, 151-158.

Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology, 70*, 103-108.

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.

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology, 11*, 299-324.

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.

Scholz, G., & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: Eine Metaanalyse [The nomological network of the assessment center: A meta-analysis]. *Zeitschrift für Arbeits- und Organisationspsychologie, 37*, 73-85.

Schorr, A. (1995). Stand und Perspektiven diagnostischer Verfahren in der Praxis: Ergebnisse einer repräsentativen Befragung westdeutscher Psychologen [Status and perspectives of applied diagnostic techniques: Results of a representative survey of West German psychologists]. *Diagnostica, 41*, 3-20.

Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology, 75*, 277-294.

Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., Jr., & Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance, 17*, 71-93.

Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231-258.

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

Zaccaro, S. J. (2002). Organizational leadership and social intelligence. In R. E. Riggio, S. E. Murphy, & F. J. Pirozzolo (Eds.), *Multiple intelligences and leadership* (pp. 29-54). Mahwah, NJ: Lawrence Erlbaum.

Table 1

Descriptive Information and Correlations for Interview and AC Performance, ATIC Scores and Cognitive Ability Measures.

Variable	*M*	*SD*	1	2	3	4	5
1. AC performance	2.94	0.54					
2. Interview performance	3.28	0.40	.38**				
3. ATIC AC	1.95	0.67	.39**	.29**			
4. ATIC interview	2.09	0.62	.34**	.35**	.40**		
5. IST verbal	101.85	15.01	.33**	.23*	.29**	.30**	
6. IST figural	99.03	15.01	.15	.09	.16	.12	.33**

Note. ATIC = score for interviewees' ability to identify criteria, IST verbal = cognitive ability score from the verbal module of the Intelligenz-Struktur-Test 2000, IST figural = cognitive ability score from the figural module of the Intelligenz-Struktur-Test 2000, $N = 95$ with the exception of cognitive ability measures where $N = 94$.

p < .05, **p* < .01.

Table 2

Standardized Regression Weights (Betas), ΔR^2, Total R^2, and Adjusted R^2 for the Hierarchical

Regressions of Assessment Center (AC) and Interview Performance.

Predictor	Interview performance		AC performance	
	Step 1	Step 2	Step 1	Step 2
IST verbal	.22*	.15*	.31**	.24*
IST figural	.02	.00	.05	.04
ATIC AC		.25*		
ATIC interview				.25*
ΔR^2		.06*		.06*
Total R^2	.05	.11*	.11**	.17**
Adjusted R^2	.03	.08	.09	.14

Note. IST verbal = cognitive ability score from the verbal module of the Intelligenz-Struktur-

Test 2000, IST figural = cognitive ability score from the figural module of the Intelligenz-

Struktur-Test 2000, ATIC = score for interviewees' ability to identify criteria. $N = 94$.

*$p < .05$, **$p < .01$.