

# “I Know What You Want to Know”: The Impact of Interviewees’ Ability to Identify Criteria on Interview Performance and Construct-Related Validity

Klaus G. Melchers

*Universität Zürich*

Ute-Christine Klehe

*Universiteit van Amsterdam*

Gerald M. Richter

*Novartis Behring, Marburg*

Martin Kleinmann and Cornelius J. König

*Universität Zürich*

Filip Lievens

*Ghent University*

The current study tested whether candidates’ ability to identify the targeted interview dimensions fosters their interview success as well as the interviews’ convergent and discriminant validity. Ninety-two interviewees participated in a simulated structured interview developed to measure three different dimensions. In line with the hypotheses, interviewees who were more proficient at identifying the targeted dimensions received better evaluations. Furthermore, interviewees’ ability to identify these evaluation criteria accounted for substantial variance in predicting their performance even after controlling for cognitive ability. Finally, the interviewer ratings showed poor discriminant and convergent validity. However, we found some support for the hypothesis that the quality of the interviewer ratings improves when one only considers ratings from questions for which interviewees had correctly identified the intended dimensions.

The idea that people differ in terms of how they assess social situations and flexibly adapt their behavior on the basis of these cues has a long history in psychology (e.g., Argyle, 1969; Thorndike, 1920). In recent years, such individual differences have known a renaissance under the general term of social effectiveness constructs. According to Ferris, Perrewé, and Dougl

effectiveness is a “broad, higher-order, umbrella term which encapsulates a number of moderately-related, yet conceptually-distinctive, manifestations of social understanding and competence” (p. 50). These social effectiveness constructs are known under various aliases such as self-monitoring (Snyder & Gangestad, 1986), social competence (Schneider, Ackerman, & Kanfer, 1996), social skill (Riggio, 1986; Witt & Ferris, 2003), social intelligence (Marlowe, 1986), emotional intelligence (Salovey & Mayer, 1990), and even practical intelligence (Neisser, 1976; Wagner & Sternberg, 1985). People high on these constructs are typically able to better “read” interpersonal situations than others and adapt their interpersonal behavior in line with the cues gathered.

Various studies have demonstrated that such individual differences in social effectiveness are related to performance on the job. For instance, Ferris et al. (2005) and Semadar, Robins, and Ferris (2006) showed how social skills are related to job performance. In a similar vein, Bowman, Markham, and Roberts (2001) reviewed research on practical intelligence in organizational, educational, and military settings and concluded that it had certain promise for predicting performance in these real-world environments. However, social skills might also be of pivotal importance to get the job in the first place. In fact, being able to quickly see through interpersonal situations better than others and to adjust one’s behavior accordingly might be beneficial in the context of going through a selection procedure. This is because many personnel selection procedures (e.g., employment interviews, assessment center exercises) represent complex interpersonal interactions that are characterized by a high degree of ambiguity and a short time span (Fox & Spector, 2000; Gilmore & Ferris, 1989). Hence, selection procedures might provide an ideal platform to observe individual differences in social effectiveness.

Recently, König, Melchers, Richter, Kleinmann, and Klehe (2007) posited that the ability to identify the criteria used in selection procedures fits in this broad category of social effectiveness constructs. They defined the ability to identify criteria (ATIC) as a candidate’s accuracy of assessing what is being measured in the specific context of a selection procedure. So, ATIC represents a context-specific operationalization of social effectiveness. In contrast to several other social effectiveness constructs (cf. Ferris et al., 2002) ATIC is not assessed by a verbal self-report inventory but instead represents an actual ability measure on which participants can give correct or incorrect answers. And in comparison to most other social effectiveness constructs, ATIC is specifically tailored to the situation in which the socially effective behavior has to be shown, thus offering an insight into the actual process by which social effectiveness influences subsequent behavior.

König et al. (2007) found that ATIC emerged as a cross-situationally consistent ability across different selection procedures (interviews and assessment centers). Furthermore, the degree to which candidates had correctly identified the targeted dimensions in one selection procedure was positively related to performance in the other procedure.

The aim of this article is to build on this work and to answer three key questions with regard to ATIC in employment interviews (both past-behavior and future-oriented interviews): First, does the ability to identify criteria lead to better interview performance? Second, is the ability to identify criteria in interviews distinct from cognitive ability, and if so, does it predict interview performance beyond cognitive ability? Third, what are the effects of the ability to identify criteria on the quality (i.e., convergent and discriminant validity) of interviewer evaluations? The remainder of the introduction draws on theory about social effectiveness to formulate hypotheses related to each of these three research questions.

This study builds on prior research on ATIC yet also offers new insights in several respects. For example, prior research on ATIC primarily examined the effects of ATIC on candidates' performance in the assessment center domain (e.g., Kleinmann, 1993; Preckel & Schüpbach, 2005). Furthermore, the effect of ATIC on the convergent and discriminant validity of interviewer ratings has not been examined in prior research. Thus, this study also examines the effect of a particular interviewee skill on the quality of interviewer ratings, which is a key extension of prior research and bears implications for employment interview theory and practice.

## BACKGROUND AND HYPOTHESES

### ATIC and Interview Performance

The first hypothesis deals with the effects of ATIC on performance (in this case, performance in the employment interview). According to Hogan and Shelton's socioanalytic theory (1998) of job performance, social effectiveness constructs (i.e., social skills) enable people to translate their intentions into actions, which in turn might provide them with better evaluations in performance-related situations. Thus, social effectiveness constructs should generally predict performance. This conceptual logic has been confirmed by recent empirical research. For instance, Ferris et al. (2005) and Semadar et al. (2006) showed that social effectiveness constructs are related to job performance. So, we posit that when some interviewees are generally better than others in identifying the relevant interview dimensions (i.e., if they have a higher ATIC), then these interviewees will perform better in the interview. For example, if a question is intended to measure cooperation, then interviewees who identify cooperation as the targeted dimension will have an advantage over candidates who incorrectly assume that the question is to assess leadership or assertiveness. This is because the former group will be more likely to present themselves in a cooperative manner, and as a consequence will presumably receive a more positive evaluation on that particular question.

So far, there exists empirical support for the role of ATIC in improving performance in assessment centers. Both Kleinmann (1993) and Preckel and Schüpbach (2005) found that there were differences between candidates in their ATIC and that those candidates who were more successful in identifying the intended dimensions assessed in various assessment center exercises also performed better in the assessment center. The present study extends this research to the interview domain and tests the validity of the arguments described in the previous paragraphs. Thereby, it also goes beyond previous research on social effectiveness in interviews that has mainly considered the effects of impression management tactics on interviewers' evaluations (e.g., Ellis, West, Ryan, & DeShon, 2002; Stevens & Kristof, 1995).

Taken together, the conceptual and empirical arguments just presented lead to the following hypothesis:

- H1: There will be a positive correlation between the degree to which interviewees identify the targeted dimensions in a selection interview (ATIC) and their performance in the interview.

### ATIC and Cognitive Ability

A second hypothesis concerns the relationship between ATIC and cognitive ability. Some researchers have proposed similar arguments for the role of cognitive ability in interviews as we

here present for ATIC. That is, candidates high on cognitive ability might be better at thinking through questions to give more appropriate answers and may also be more able to present themselves in a favorable way (Harris, 1999; Huffcutt, Roth, & McDaniel, 1996). We agree that interviewees' cognitive ability might contribute to their ATIC in employment interviews. Yet cognitive ability is a more general construct than the correct identification of the criteria to be measured in interpersonal situations such as interviews.

Furthermore, and on a more general level, the discussion about the distinctiveness of social effectiveness constructs and cognitive ability is not new. A common thread running through the research on social effectiveness is the attempt to distinguish this construct from cognitive ability. In line with this thread, social effectiveness constructs and cognitive ability often covary with each other (e.g., Jones & Day, 1997; Lee, Wong, Day, Maxwell, & Thorpe, 2000; Weis & Süß, 2007). Often, however, the correlations only reach a moderate level of about .20, and the empirical evidence shows that the two constructs can be distinguished from each other (Jones & Day, 1997; Lee et al., 2000; Weis & Süß, 2007). This indicates that social effectiveness and cognitive ability are distinct but nevertheless related constructs. However, as the correct identification of the criteria to be measured in interpersonal situations such as interviews is a more situation-specific construct than is cognitive ability, ATIC should explain variance in candidates' interview performance even after cognitive ability is already taken into account. This leads to the following hypothesis:

- H2: ATIC will explain variance in candidates' interview performance over and above the variance explained by cognitive ability.

### ATIC and Interview Construct-Related Validity

The last set of hypotheses deals with the effects of ATIC on the interviewer evaluations. In structured interviews, different interview questions often aim at different job-related dimensions (e.g., cooperation, planning, leadership, analysis). The manner in which correct identification of the targeted interview dimensions might influence interviewer evaluations on these competencies becomes evident if one considers the following scenarios: Suppose an interviewee correctly infers that cooperation is to be assessed in each of two interview questions. As a consequence, he or she will try to describe cooperative behavior in both answers. Conversely, if the interviewee only identifies cooperation as the intended dimension in the first question and incorrectly assumes that the second question is designed to measure leadership behavior, then he or she will describe different kinds of behavior in the two situations. As a consequence, the content of the interviewee's two answers will be more similar in the former case than in the latter. In turn, this should lead to larger correlations (i.e., higher convergent validity) between ratings of cooperation when this dimension is correctly identified in both questions than when it is not.

Indirect support for the suggested impact of applicants' correct identification of the targeted interview dimensions comes from an assessment center study by Kleinmann (1993). Kleinmann found that the convergent validity of assessment center ratings was moderated by whether participants had identified the targeted dimensions. Ratings of the same dimension measured in different assessment center tasks correlated nearly twice as highly with each other when participants had correctly identified the respective dimension in both tasks than when they had identified it in one task but not in the other. However, we are not aware of any research that assesses a similar beneficial effect on the convergent validity in selection interviews. Furthermore, we also expect that

ATIC will influence the discriminant validity of interviewer ratings. Such an effect has not at all been investigated by prior research but seems plausible in the case of interviews for the following reason: Suppose interviewees incorrectly assume that a particular question targets a specific dimension. In this case, they might present behavioral examples in their answers that are irrelevant with regard to the dimension actually targeted. In turn, this might cloud the distinction between the targeted dimension and the other dimensions. Conversely, interviewees who realize that two questions assess two different dimensions might give answers more in line with these different dimensions and describe more distinct patterns of behavior, leading to better differentiation among the dimensions in the evaluations of the interviewers. As a consequence, ratings of the different dimensions might correlate less. That is, their discriminant validity might be better. All of this leads to the following hypotheses:

- H3a: Interviewer ratings will show better convergent validity when interviewees have correctly identified the targeted dimensions in comparison to when they have not identified the targeted dimensions.
- H3b: Interviewer ratings will show better discriminant validity when interviewees have correctly identified the targeted dimensions in comparison to when they have not identified the targeted dimensions.

## METHODS

### Participants

Ninety-two participants (46 males, 46 females) participated in a 2-day selection simulation. This simulation was conducted as part of an application training program for recent or prospective university graduates who were currently applying for jobs or would be doing so in the near future. On average, participants had been attending university for 4 years 10 months ( $SD = 1.92$ ). More than one fourths of them (27.2%) had already obtained the German equivalent of a master's degree and 47.8% had prior work experience. Their mean age was 26.88 ( $SD = 2.97$ ). The participants came from various study majors, with the largest group having a background in business or economics. To cover part of the costs and to increase their motivation, participants were required to pay a small fee for participating in the training program as part of which the selection simulation was administered. All participants consented that their data could be used for research purposes.

### Interview Development

We used an interview that was also used in a recent study by Klehe, König, Richter, Kleinmann, and Melchers (2008, Study 2). It consisted of two different components (a set of past-behavior questions and a set of future-oriented questions) designed to be suitable for selecting candidates for a hypothetical management trainee position. This position was chosen because it represented a realistic and attractive position suitable for applicants from diverse areas of study.

The first part of the interview contained the past-behavior questions and the second part contained the future-oriented questions. The procedure to develop the original interview is described in detail by Klehe et al. (2008). In brief, it consisted of the following steps: Identification of potentially job-relevant dimensions on the basis of critical incidents, determining which of the dimen-

sions are suitable to be assessed in a structured interview, selection of a subset of dimensions that should be most independent from each other (resulting in the three following dimensions: Systematic Planning, Cooperation, and Leadership Behavior), collecting and selection of potential questions for the different dimensions (most of the questions were taken from actual interviews developed and validated for selection purposes in field settings or from earlier interview studies, e.g., Deller & Kleinmann, 1993; Melchers, Kleinmann, Richter, König, & Klehe, 2004; Schuler & Moser, 1995) and of pretesting whether the questions indeed reflected the targeted dimensions and whether the descriptive rating anchors provided for each question were understandable and were suitable to reflect poor (1), average (3), and good answers (5) on a 5-point rating scale. Only questions rated as clearly measuring the intended dimension and none of the nonintended dimensions provided during pretesting were chosen for the final set of 12 past-behavior and 12 future-oriented questions (4 questions per dimension). The questions chosen had an average rating of  $M = 8.45$  for the targeted dimension (on a scale from 0 to 10, with 10 indicating that a question optimally corresponded to the respective dimension) and average ratings between 0.19 and 3.08 for nontargeted dimensions. Examples from the final set of future-oriented and past-behavior questions are given in the Appendix.

## Procedure

**Interviewer Training.** In total, 32 master's-level students (the majority of them being work and organizational psychology students) served as interviewers. All interviewers participated in a 1-day interviewer training. In this training, they were introduced to the two interview formats as well as to definitions and behavioral examples of the three dimensions to be assessed. The interviewers received information about typical rating errors and discussed the interview questions and the behavioral scoring guides to achieve a consistent frame-of-reference for rating interviewees' answers (Woehr & Huffcutt, 1994). It was pointed out to them that it was crucial to read the interview questions to the interviewees as printed on the interview forms and not to rephrase them or to give additional cues in any way. Furthermore, interviewers were introduced to other selection procedures also included in the selection simulation for training participants to gain additional experience with them. The interviewers did not receive any information concerning the objectives of the study.

**Selection Simulation.** Prior to the selection simulation, participants received a fictitious job advertisement for a management trainee position within a large technology corporation and were asked to prepare a written application for the job described. They were also informed that the interview would focus on this position. The advertisement included information concerning the requirements for the job. This information was formulated specifically for the dimensions we intended to assess in 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 selection simulation consisted of the structured interview and several other selection procedures, among them two modules of a cognitive ability test (see next) as well as of additional information sessions. Each participant was interviewed individually by a panel of two interviewers. The interviewers first asked the 12 past-behavior questions and then the 12 future-oriented questions. Both interviewers independently recorded and scored the responses on the basis of the scoring guides before asking the next question. The interviewers knew the dimension that each ques-



tion was intended to measure but did not inform interviewees about the respective dimensions. The interviewers were not given access to the participants' written applications until the interviews were finished.

Each of the two interview components (i.e., the set of past-behavior questions and the set of future-oriented questions) lasted approximately 20 min. Upon completion of each component, participants received a questionnaire. This questionnaire repeated all the interview questions and participants were asked to write down possible hypotheses that they had entertained during the actual interview as to what a certain question was trying to assess. Space was provided to note a maximum of two such hypotheses.

Following completion of the interview process, the interviewers discussed their ratings. They were instructed to focus on answers to questions for which their ratings were 2 or more points apart from each other (on the 5-point scale). Although interviewers did not have to agree with each other, most differences could be resolved after a short debate, meaning that the ratings did not usually differ by more than 1 point.

## Variables

**Interview Performance.** For all analyses, the overall ratings of the two interviewers were averaged. These overall ratings were the statistical means of the ratings across all 24 interview questions. To determine the reliability of the interview, we calculated intraclass correlations (ICC 1,1) between the overall ratings of the two interviewers after their discussion of the ratings. The mean interrater reliability (i.e., the reliability of a single interviewer) for these overall ratings was .88.

**ATIC.** To ascertain the degree to which the interviewees' hypotheses as to what the different questions were trying to assess corresponded to the intended dimensions, we employed two different procedures. The first procedure was similar to that used in the assessment center study by Kleinmann (1993): At the end of the training, participants were introduced to six dimensions that structured interviews often intend to assess. Three of these dimensions were the dimensions used in the interview, and three were distractor dimensions often used in other interviews (Job Knowledge and Experience, Self-Confidence, and Acquisition and Handling of Information; cf. Huffcutt, Conway, Roth, & Stone, 2001). Participants also received a list of behavioral examples for each dimension. After having read the list, the interviewees were handed back the questionnaires on which they had written down their own hypotheses. They were then asked to indicate which of their hypotheses corresponded to which of the six dimensions and to rate the strength of this correspondence on a scale from 1 (*fits somewhat*) to 4 (*fits completely*). Alternatively, participants could also indicate that a hypothesis did not correspond to any of the dimensions.

The second procedure was similar to the first one with the important difference that participants did not have to categorize their hypotheses themselves but that two trained master's-level work and organizational psychology students coded these hypotheses. In a first step, these coders determined whether a hypothesis corresponded to one of the three target dimensions (and if so to which), and in a second step, they rated the strength of this correspondence from 1 to 4 on the same scale that had been used by the interviewees. To determine their interrater agreement, both external coders coded the interviewees' hypotheses for a sample of 14 interviewees. Concerning the categorization of the hypotheses with regard to the different target dimensions in the first step, Cohen's  $\kappa$  was .61, which can be considered to be satisfactory (Fleiss & Cohen, 1973), and concern-

ing the strength of the correspondence ratings from the second step, we found an intraclass correlation (ICC 2,1) of  $r = .84$  between the two coders which can be considered as good given that this intraclass correlation takes into account both consistency and absolute agreement (Shrout & Fleiss, 1979).

Both of the procedures to code interviewees' hypotheses reflect the situationally specific nature of ATIC and the fact that ATIC is a measure of ability. Yet both have potential strengths and weaknesses. The first procedure relying on interviewees' self-coded hypotheses is similar to situational judgment tests where candidates themselves indicate what they would do (e.g., Motowidlo, Dunnette, & Carter, 1990). Thus, this procedure follows an established measurement approach. Furthermore, it has the advantage that only the interviewees themselves know how to interpret the idiosyncrasies captured in their specific hypotheses and how much these idiosyncratic responses relate to the different dimensions that they were introduced to after the selection simulation. The disadvantage of this procedure is that an additional bias might occur when interviewees indicate the extent to which they have identified the targeted dimensions. Specifically, the general tendency to try to make a good impression in one's own eyes as well as in the eyes of others (Leary & Kowalski, 1990) might lead to a self-serving bias at this stage, which in turn should result in inflated ATIC scores. The second procedure using external coders has the advantage that it prevents the occurrence of such a self-serving bias and the associated inflation of ATIC scores. Furthermore, this approach extends previous ATIC research that largely relied on candidates' self-coding of hypotheses (e.g., Kleinmann, 1993). The potential drawback of using external coders, however, is that those coders might have difficulties to correctly understand the idiosyncratic meaning of interviewees' hypotheses and to fully comprehend how far these hypotheses actually correspond to the targeted dimensions.

Independent from the procedure to code interviewees' hypotheses, the final step in the determination of the ATIC scores was the same. Specifically, we determined interviewees' ATIC scores for each question and then averaged these scores across all questions to calculate an overall ATIC score for the interview. For each question, we focused on ratings indicating to what degree the specific hypotheses corresponded to the intended dimension, and took the rating from the hypothesis that had the highest fit for that dimension. If both hypotheses corresponded to incorrect dimensions or to none of the dimensions at all, a score of 0 was assigned to this rating. This led to ATIC scores for individual questions ranging from 0 (indicating no fit with the correct dimension) to 4 (indicating perfect fit with the correct dimension). An example for how ATIC scores were determined on the basis of the interviewees' self-coded hypotheses is provided in the Appendix. The internal consistency of the ATIC scores (coefficient alpha) was .79 for the self-coded hypotheses and .60 for the externally-coded hypotheses. Furthermore, the correlation between ATIC scores on the basis of the self-coded and of the externally-coded hypotheses was  $r = .61$  ( $p < .01$ ).

**Cognitive Ability.** The interviewees completed two modules of the IST 2000 (Amthauer, Brocke, Liepmann, & Beauducel, 1999), a widely used German cognitive ability test with good criterion-related validity in selection settings (Hülsheger, Maier, Stumpp, & Muck, 2006). The first module contains three subtests (Sentence Completion, Analogies, and Similarities) and comprises a measure of reasoning for verbal tasks. The second module also contains three subtests (Figures, Cubes, and Matrices) and comprises a measure of reasoning for figural tasks. In previous investigations, coefficient alpha was .88 for the verbal and .87 for the figural module (Amthauer et



al., 1999). Scores for the IST 2000 were determined on the basis of norm values for people who were eligible for university studies in Germany.

## RESULTS

### Preliminary Analyses

There was considerable variability between interviewees concerning the degree to which they had identified the targeted dimensions. First, we considered the number of questions for which the targeted dimension was completely identified (i.e., the questions for which interviewees had obtained the maximum ATIC value of 4). When correct identification was determined on the basis of interviewees' self-coded hypotheses, the average number of correctly identified targeted dimension was 7.36 ( $SD = 5.60$ ) out of the 24 questions (range = 0–19). In contrast to this, the average number of correctly identified dimensions was 11.17 ( $SD = 3.52$ ) of the target dimensions when identification was determined on the basis of the externally-coded hypotheses (range = 2–19). A paired-samples  $t$ -test confirmed that the average for the self-coded hypotheses was smaller than for the externally-coded hypotheses,  $t(91) = 6.80, p < .01$ . Second, we considered interviewees' mean ATIC scores. Here, the mean on the basis of the self-coded hypotheses was 2.09 ( $SD = 0.63$ ) and the mean on the basis of the externally-coded hypotheses was 2.37 ( $SD = 0.49$ ). Again, the mean on the basis of the self-coded hypotheses was lower than for the externally-coded hypotheses,  $t(91) = 5.22, p < .01$ . These results make the presence of a self-serving bias in interviewees' self-coded hypothesis unlikely as such a bias would have inflated their ATIC scores in comparison to the external coders. Furthermore, the lower internal consistency of the ATIC measure on the basis of the externally-coded hypotheses (.60 in comparison to .79 for the ATIC score on the basis of the self-coded hypotheses) might indeed reflect the problem that the external coding does not fully capture the idiosyncrasies of the interviewees' hypotheses, suggesting that the self-coded hypotheses represent a more accurate and reliable estimate.

Next, we considered the transparency of the different interview questions. This transparency was calculated as the mean of the interviewees' ATIC scores for the respective question and could range between 0 and 4. The actual range of these transparency scores for the self-coded and the externally-coded hypotheses varied from 0.64 to 3.52 ( $M = 2.09, SD = 0.58$ ) and between 0.99 and 2.97 ( $M = 2.37, SD = 0.79$ ), respectively, meaning that no question was completely transparent or completely nontransparent to the interviewees. Overall, most interview questions were moderately transparent.

Additional descriptive and correlational information for participants' interview performance and for the other dependent variables can be found in Table 1.

### Interviewees' Ability to Identify Criteria and Interview Performance

H1 stated that the degree to which interviewees identify the targeted dimensions in an interview will be positively related to their performance in the interview. In line with this prediction, we found significant correlations between mean ATIC scores and interview performance. For the ATIC scores on the basis of the self-coded hypotheses, the correlation was  $r = .36 (p < .01)$ , and for the ATIC scores on the basis of the externally-coded hypotheses it was  $r = .26 (p < .05)$ .

TABLE 1  
Descriptive Information and Correlations for Interview Performance,  
ATIC Scores and Cognitive Ability Measures

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5
1. Interview performance	2.73	0.40					
2. ATIC self-coded	2.10	0.62	.36**				
3. ATIC externally-coded	2.37	0.49	.26*	.61**			
4. IST verbal	101.89	15.17	.23*	.30**	.19†		
5. IST figural	99.10	15.13	.09	.12	.04	.33**	

Note. *N* = 92. ATIC = score for interviewees' ability to identify criteria; IST verbal = cognitive ability score from the verbal module of the IST 2000; IST figural = cognitive ability score from the figural module of the IST 2000.

†*p* < .10. \**p* < .05. \*\**p* < .01.

To illustrate the impact of interviewees' ATIC on their performance we also compared extreme groups of interviewees. A comparison of interviewees with very high or with very low ATIC scores (top vs. lowest 10%), for example, shows that the former were approximately one standard deviation better in the interview than the latter (*M* = 3.34, *SD* = 0.45 vs. *M* = 2.92, *SD* = 0.39 for ATIC scores on the basis of the self-coded hypotheses and *M* = 3.45, *SD* = 0.39 vs. *M* = 3.06, *SD* = 0.27 for the ATIC score on the basis of the externally-coded hypotheses).

### Interviewees' Ability to Identify Criteria and Cognitive Ability

H2 predicted that ATIC will explain variance in candidates' interview performance over and above the variance explained by cognitive ability. Before we tested this hypothesis, we first evaluated the relationships between cognitive ability and ATIC and between cognitive ability and interview performance.

We found a moderate correlation between ATIC and the score from the verbal module of the cognitive ability test. Verbal ability correlated  $r = .30$  ( $p < .01$ ) and  $r = .19$  ( $p < .08$ ) with ATIC scores determined on the basis of the interviewees' self-coded hypotheses and the externally-coded hypotheses, respectively. Of interest, and in contrast to this, no significant correlations between ATIC scores and cognitive ability were obtained for the score from the figural module (both  $r$ s < .13). Thus, it appears that verbal ability somewhat increases the degree to which interviewees manage to identify the intended dimensions but that other aspects of cognitive ability are less important in this regard. Furthermore, the size of the correlations between cognitive ability and ATIC shows that a considerable amount of variance in interviewees' ATIC cannot be explained by cognitive ability alone. Concerning the relationship between cognitive ability and interview performance, we found that the correlation between verbal ability and interview performance was  $r = .23$  ( $p < .05$ ), which is similar to previous meta-analytic findings (Berry, Sackett, & Landers, 2007; Huffcutt et al., 1996) but that the correlation for the figural module was not significant ( $r = .09$ ).

To test whether ATIC scores can explain variance in interview performance beyond cognitive ability (as predicted by H2) we conducted hierarchical multiple regressions. Insofar as our assessment of ATIC captures other aspects that are important for success in the interview in addition to cognitive ability, ATIC scores should account for variance in interview performance even when cognitive ability is controlled for. Therefore, in the first step, we included the two measures of

TABLE 2  
Standardized Regression Weights, Total  $R^2$ , and  $\Delta R^2$  for the Hierarchical Regressions  
of Interview Performance on Cognitive Ability and ATIC

<i>Predictor</i>	<i>ATIC Self-Coded</i>		<i>ATIC Externally-Coded</i>	
	<i>Step 1</i>	<i>Step 2</i>	<i>Step 1</i>	<i>Step 2</i>
IST verbal	.22*	.13	.22*	.18
IST figural	.02	.01	.02	.02
ATIC		.32**		.23*
Total $R^2$	.05†	.14**	.05†	.10*
$\Delta R^2$		.09**		.05*

*Note.*  $N = 92$ . ATIC = score for interviewees' ability to identify criteria; IST verbal = cognitive ability score from the verbal module of the IST 2000; IST figural = cognitive ability score from the figural module of the IST 2000.

† $p < .10$ . \* $p < .05$ . \*\* $p < .01$ .

cognitive ability, and in the second step, we included the ATIC score to predict candidates' performance in the interview.

As can be seen in Table 2, cognitive ability accounted for only a limited amount of variance ( $R^2 = .05$ ,  $p < .10$ ). However, in each analysis, ATIC accounted for significant amounts of additional variance in predicting interviewees' performance in the interview ( $\Delta R^2 = .09$ ,  $p < .01$ , and  $\Delta R^2 = .05$ ,  $p < .05$ , for the ATIC scores determined on the basis of the interviewees' self-coded hypotheses and the externally-coded hypotheses, respectively).

### Construct-Related Validity of the Interviewer Ratings

To evaluate the construct-related validity of the interviewer ratings and the role of ATIC for interview construct-related validity, we followed prior interview studies (e.g., Conway & Peneno, 1999; Huffcutt, Weekley, Wiesner, DeGroot, & Jones, 2001; Schuler & Funke, 1989) and calculated a multitrait-multimethod (MTMM) matrix. This enables to get estimates of the interview's convergent and discriminant validity. The method factors in this matrix were the two interview components (the past-behavior and the future-oriented questions) and the trait factors were the three interview dimensions (Systematic Planning, Cooperation, Leadership Behavior). Table 3 shows the MTMM matrix that resulted from correlating the mean scores for the different dimensions from the two interview formats with each other. With regard to convergent validity, all correlations between ratings for identical dimensions across the two interview components were significant. The average of the three monotrait-heteromethod coefficients (using an  $r$ -to- $Z$  transformation) was  $r = .25$ . Concerning discriminant validity, correlations between ratings for different dimensions within the same component were also significant, with an average correlation of  $r = .33$ . Thus, within-dimension correlations were smaller than within-component correlations, thereby indicating poor construct-related validity. These results are similar to results from previous studies that used a comparable approach to investigate the construct-related validity of structured interviews (e.g., Conway & Peneno, 1999; Huffcutt, Weekley, et al., 2001; Melchers et al., 2004; Schuler & Funke, 1989).

Regarding the analyses assessing the influence of correct identification of the evaluation criteria on convergent and discriminant validity (H3a and H3b), we focused on questions for which the

TABLE 3  
Multitrait–Multimethod Matrix for the Interview

<i>Dimension</i>	<i>M</i>	<i>SD</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
Past-behavior questions								
1. Systematic Planning	3.34	0.65						
2. Cooperation	3.25	0.58	.38**					
3. Leadership Behavior	3.13	0.86	.25*	.33**				
Future-oriented questions								
4. Systematic Planning	3.27	0.68	.21*	.27**	.36**			
5. Cooperation	3.39	0.62	.15	.27**	.00	.21*		
6. Leadership Behavior	3.27	0.55	-.01	.33**	.27**	.38**	.40**	

*Note.*  $N = 92$ .

\* $p < .05$ . \*\* $p < .01$ .

relevant dimensions had either been completely identified (ATIC value = 4) or not identified at all (ATIC value = 0). On the basis of those questions for each interviewee for which the relevant dimensions had either been completely identified or not identified at all, we calculated new MTMM matrices and new sets of corresponding within-dimension and within-component correlations. These correlations were based on a varying number of participants, depending on how many interviewees had ATIC values of 0 or 4, respectively, for the specific cell of the MTMM matrix. The mean sample size upon which the correlations were based was 45, with a minimum of 31 and a maximum of 66 interviewees for the analyses that made use of interviewees' self-coded hypotheses. For the analyses based on the externally-coded hypotheses the mean  $n$  was 53, with a minimum of 24 and a maximum of 85. For the subsequent analyses, the correlations were weighted by their respective sample sizes. A summary of the results for these analyses is shown in Table 4. For the sake of completeness we also calculated new within-dimension and within-component correlations for questions for which the targeted dimensions were partially identified. Even though we do not go into these results here they are also given in Table 4.

For questions for which a common dimension had been correctly identified according to interviewees' self-coded hypotheses, the mean convergent validity (i.e., within-dimension correlation) was  $r = .35$  and the mean discriminant validity (i.e., within-component correlation) was  $r = .12$ . Thus, within-dimension correlations were larger than within-component correlations. Furthermore, in comparison to the MTMM analysis for the complete data set this indicates better construct-related validity with an improvement of the mean convergent validity of  $\Delta r = .10$  and an improvement for the mean discriminant validity of  $\Delta r = -.22$ . In contrast to this, ratings based on questions for which the dimension had not been identified yielded a mean convergent validity coefficient of  $r = .06$  and a mean discriminant validity coefficient of  $r = .11$ . Similarly, when participants had correctly identified the dimension in one component but not in the other, the mean convergent validity was  $r = .05$  and the respective mean discriminant validity was  $r = .12$ . Thus, although within-component correlations were close to 0 in the last two analyses (suggesting an improvement of  $\Delta r = -.21$  compared to the overall discriminant validity), the low correlations of the respective within-dimension correlations (reflecting a decline of convergent validity of approximately  $\Delta r = -.20$ ) nevertheless indicate a lack of construct-related validity. Taken together, these results are in agreement with H3a and H3b predicting that the interview shows better con-

TABLE 4  
Influence of ATIC on the Internal Construct-Related Validity of the Interview

<i>Degree to Which Targeted Dimensions Were Correctly Identified for Two Cells of the MTMM Matrix</i>	<i>Convergent Validity</i>	<i>Discriminant Validity</i>
ATIC self-coded		
Not at all – not at all	.06	.11
Not at all – partially	.18	.04
Not at all – completely	.05	.12
Partially – partially	.19	.17
Completely – partially	.12	.19
Completely – completely	.35	.12
ATIC externally-coded		
Not at all – not at all	.28	.22
Not at all – partially	.12	.14
Not at all – completely	.07	.05
Partially – partially	.12	.02
Completely – partially	.06	.20
Completely – completely	.16	.16
Construct-related validity complete data set	.25	.33

*Note.* Row entries in this table refer to the degree to which the different interview dimensions were identified for combinations of different cells of the new MTMM matrices. “Completely – not at all”, for example, means that the validities depict the mean correlations for cells for which interviewer evaluations stem from questions that were completely identified (ability to identify criteria [ATIC] value = 4) for one of the correlated variables (e.g., Leadership in future-oriented questions) but not at all identified (ATIC value = 0) for the other variable (e.g., Leadership in past-behavior questions).

struct-related validity when interviewees have correctly identified the targeted dimensions in comparison to when they have not identified the targeted dimensions.

ATIC was not related to clear improvements of interview construct-related validity in the analyses based on externally-coded hypotheses. The mean convergent validity was low and comparable to the mean discriminant validity for questions for which a common dimension had been correctly identified (both  $r$ s = .16). Ratings based on questions for which the dimension had not been identified even yielded a mean convergent validity of  $r = .28$  that was higher than the mean discriminant validity ( $r = .22$ ). Thus, for the externally-coded hypotheses the results are at variance with H3a and H3b.

## DISCUSSION

### Main Conclusions

This study adds various new findings to the literature on employment interviews. First, our results show that interviewees' ATIC was related to their performance in the interview. Specifically, we found significant correlations between the degree to which interviewees correctly identified the targeted interview dimensions and their interview performance. These results were obtained independently from whether ATIC scores were determined on the basis of candidates' self-coded hy-

potheses or on the basis of externally-coded hypotheses. It should be noted that we found that ATIC scores were even a better predictor of interview performance than measures of cognitive ability. Moreover, even after controlling for cognitive ability, ATIC accounted for considerable variance in interviewees' performance. These results confirm that ATIC might play a similar role as social effectiveness constructs in leading to better performance (Ferris et al., 2005; Semadar et al., 2006).

Second, we found that ATIC was correlated with cognitive ability but that the size of the correlations was only low to moderate. Thus, differences in interviewees' cognitive abilities accounted for part of the individual differences in their ATIC scores. However, the size of the correlations suggest that interviewees' ATIC was not tantamount to cognitive ability. This result is in line with the literature on social effectiveness constructs (Jones & Day, 1997; Lee et al., 2000; Weis & Süß, 2007). It is also notable that the correlation between ATIC and verbal ability was significant (albeit only marginally for the externally-coded ATIC scores), whereas the correlations with figural ability were not significant. Clearly, verbal ability helps to "better understand" interview questions. Similar findings have been obtained in the literature about social effectiveness where stronger relationships between different social effectiveness constructs and verbal ability have been reported in comparison to non-verbal cognitive abilities when the social effectiveness construct was also measured for the verbal domain (e.g., Jones & Day, 1997; Weis & Süß, 2007). Taken together these results confirm König et al.'s argument (2007) that ATIC works similar to a context-specific form of social effectiveness.

Third, the present study is the first to find evidence for the hypothesized influence of ATIC on the convergent and discriminant validity of interviewers' evaluations which represents a key extension of prior research. When we analyzed the entire dataset, we found that the interview had poor internal construct-related validity, which parallels previous findings (Conway & Peneno, 1999; Huffcutt, Weekley, et al., 2001; Melchers et al., 2004; Schuler & Funke, 1989). In comparison to this, when only questions were used for the MTMM analysis for which the targeted dimensions were identified correctly (according to the interviewees' self-coded hypotheses), the mean convergent validity coefficient of the interview exceeded the mean discriminant validity coefficient (indicating better construct-related validity). The reason for this finding lies in the fact that there was both an increase in the within-dimension correlations (i.e., an increase in convergent validity) and a decrease in the within-components correlations (i.e., an increase in discriminant validity). In contrast to the improved construct-related validity of ratings of correctly identified dimensions, the convergent validity deteriorated when the intended dimension was only identified in one interview component but not in the other, or when it was not identified in either component.

The influence of ATIC on the construct-related validity of the interview did not show up when the analyses were based on the hypotheses as coded by external experts. In that case, the mean convergent correlations did not improve when only questions were used for the MTMM analysis for which the targeted dimensions were identified correctly. In contrast to this, the mean convergent correlations were (at least somewhat) larger than the discriminant correlations when the targeted dimensions were not identified at all in either component.

There are at least two possible reasons for the divergent results concerning the moderating influence of ATIC on interview construct-related validity. First, these results are less stable than our other analyses because of the reduced sample sizes on which the mean convergent and discriminant correlations were based and because scores for the two interview components were



often based on less than four questions (e.g., when interviewees had identified the targeted dimension on only two of the four questions, performance on only these two questions was analyzed). Second, it might be that the divergent results are due to the use of external coders in contrast to interviewees self-categorizing their hypotheses. The rationale for using external coders was to prevent biases in the measurement of ATIC, a proposition not supported in our preliminary analyses. Yet, external coders might also introduce additional errors as those coders might not discern the specific idiosyncrasies captured by interviewees' hypotheses. The lower reliability found for external coders supports this interpretation. All of this suggests that the self-coded hypotheses seem a more appropriate measure for assessing ATIC. However, we encourage future research to design and compare different measures of ATIC.

### Limitations

Some limitations should be addressed. First, our data were not obtained from a real sample of applicants. However, many of our participants were currently seeking employment (often for management trainee positions similar to the hypothetical position in the present study) and were using the selection simulation as an opportunity to prepare for their real job interviews and to see how well they can handle such situations. Furthermore, we also questioned candidates after the selection simulation concerning the perceived realism of the simulation on four items (e.g., "Did you perceive the application training to be realistic?" or "During the training, did you act as you would during a real job application?"), on 4-point scales ranging from 1 (yes) to 4 (no). The participants' average rating of 1.86 on these items was significantly lower than 2 (rather yes),  $t(91) = 2.42, p < .05$ , indicating that most participants perceived the selection simulation as being a realistic reflection of an application situation, and as such, acted accordingly. Thus, our sample seemed to be motivated to show high performance and also seemed to be comparable to a sample of actual applicants. Nevertheless, future research is needed to test whether ATIC is related to candidates' experience with selection procedures. Given that ATIC is conceptualized as an ability, we expect to find such a relationship.

Second, our conclusions concerning the relationship between ATIC and interview performance are restricted to structured interviews such as the one used in the present study. Thus, future research is necessary to evaluate whether the present findings generalize to unstructured interviews. However, as was suggested by a reviewer of this article, it might well be the case that ATIC (and social effectiveness constructs in general) even play a greater role in unstructured interviews where the evaluation criteria and the potentially targeted performance dimensions are less clear than in structured interviews. This is because initial success at meeting the interviewers' expectations may additionally impact the behavioral responses with which the interviewer reacts towards the interviewee, thus facilitating a successful social interaction from the perspective of the candidate (Dipboye, 2005).

Finally, with regard to our findings concerning the construct-related validity of the interview, one might ask whether the use of students as interviewers might be responsible for the poor construct-related validity. Although we cannot rule out this possibility, it seems unlikely because our results are comparable to previous findings concerning interviewer reliability (Conway, Jako, & Goodman, 1995; Taylor & Small, 2002) and interview construct-related validity obtained in field

samples (Conway & Peneno, 1999; Huffcutt, Weekley, et al., 2001; Schuler & Funke, 1989; Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004).

## Implications

This study opens various interesting avenues for future research. First, it is possible that people's ATIC, as a specific social effectiveness construct, is not limited to performance in the interview but also influences their performance on the job especially in interpersonal situations. In turn, ATIC might partially account for the good criterion-related validity of structured interviews (e.g., Huffcutt, Conway, Roth, & Klehe, 2004; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Taylor & Small, 2002). People 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 on that job. Future research should therefore examine whether similar to their success in selection procedures (interviews or assessment centers), employees' success in a job will be influenced by their ability to correctly interpret situational cues and to discern the criteria for success in a given situation on the basis of those cues.

Second, we need studies that investigate the relationship between ATIC and other social effectiveness constructs such as social skills or self-monitoring. We expect ATIC to covary with those other constructs as they all share a common ground (Schollaert & Lievens, 2008). That is, they all capture aspects of social effectiveness. But in contrast to most of the constructs reviewed by Ferris et al. (2002), ATIC is more situation specific and might therefore allow better prediction of performance in specific selection situations. However, with regard to job performance it seems necessary to assess the relative contribution of ATIC in comparison to other social effectiveness constructs (Semadar et al., 2006).

Third, the finding of better construct-related validity of our interview when only correctly identified dimensions were considered suggests that one might improve the construct-related validity of interviews by telling interviewees in advance which dimensions will be assessed. Such information would allow them to concentrate on dimension-relevant behavior with the result that the intended dimensions could be measured more accurately. Recent research has indeed confirmed this assumption (Klehe et al., 2008). However, given the finding that interviewees' ATIC is an important determinant of their performance in the interview and could also be a factor influencing job performance as well, important information about applicants could be lost if the dimensions were to be revealed to them beforehand. Thus, even though revealing the dimensions might improve the construct-related validity of an interview, it might have a negative effect on its criterion-related validity.

## CONCLUSION

Taken together, this study's results are in line with the hypotheses that the identification of evaluation criteria influences interview performance—even after controlling for the effects of cognitive ability. Thus, it confirms the assumption that social effectiveness constructs are relevant not only for performance on the job but also during selection situations to get the job in the first place. Our results also suggest that ATIC—or knowledge about the targeted dimensions in gen-

eral—might affect the interview's convergent and discriminant validity. Therefore, the results provide important directions for further research on factors related to interview performance and validity.

## ACKNOWLEDGMENTS

The research reported in this article was supported by grant KI 823/6–1 from the German Science Foundation (Deutsche Forschungsgemeinschaft) to Martin Kleinmann. We thank Thomas Hartstein, Dorit Auge, Katja Nicht, Peter Guzzardi, and Torsten Biemann for their help with the data collection, Alice Inauen and Janine Thut for their help with data coding, and Deniz Ones, Stephan Dilchert, and Jose Cortina for helpful comments on previous versions of this article.

## REFERENCES

- Anthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000 [Intelligence Structure Test 2000]*. Göttingen, Germany: Hogrefe.
- Argyle, M. (1969). *Social interaction*. London: Methuen.
- Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology, 60*, 837–874.
- Bowman, D. B., Markham, P. M., & Roberts, R. D. (2001). Expanding the frontier of human cognitive abilities: So much more than (plain) g! *Learning and Individual Differences, 13*, 127–158.
- 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.
- Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology, 13*, 485–506.
- Deller, J., & Kleinmann, M. (1993). Das situative Interview [The situational interview]. In A. Gebert & W. Hacker (Eds.), *Arbeits- und Organisationspsychologie 1991 in Dresden* (pp. 336–343). Bonn, Germany: Deutscher Psychologen Verlag.
- Dipboye, R. L. (2005). The selection/recruitment interview: Core processes and contexts. In A. Evers, N. Anderson, & O. Smit-Voskuil (Eds.), *The Blackwell handbook of personnel selection* (pp. 121–142). Malden, MA: Blackwell.
- 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.
- Ferris, G. R., Perrewé, P. L., & Douglas, C. (2002). Social effectiveness in organizations: Construct validity and research directions. *Journal of Leadership & Organizational Studies, 9*, 49–63.
- Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C., et al. (2005). Development and validation of the political skill inventory. *Journal of Management, 31*, 126–152.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613–619.
- Fox, S., & Spector, P. E. (2000). Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It's not all just 'G.' *Journal of Organizational Behavior, 21*, 203–220.
- Gilmore, D. C., & Ferris, G. R. (1989). The effects of applicant impression management tactics on interviewer judgments. *Journal of Management, 15*, 557–564.
- Harris, M. M. (1999). What is being measured? In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 143–157). Thousand Oaks, CA: Sage.
- Hogan, R., & Shelton, D. (1998). A socioanalytical perspective on job performance. *Human Performance, 11*, 129–144.
- 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.
- Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., DeGroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, 54, 619–644.
- Hülshager, U. R., Maier, G. W., Stumpp, 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, 145–162.
- Jones, K., & Day, J. D. (1997). Discrimination of two aspects of cognitive-social intelligence from academic intelligence. *Journal of Educational Psychology*, 89, 486–497.
- Klehe, U.-C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion related validity. *Human Performance*, 21, 107–137.
- 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.
- König, C. J., Melchers, K. G., Richter, G. M., Kleinmann, M., & Klehe, U.-C. (2007). The ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment*, 15, 283–292.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34–47.
- Lee, J.-E., Wong, C.-M. T., Day, J. D., Maxwell, S. E., & Thorpe, P. (2000). Social and academic intelligences: A multi-trait-multimethod study of their crystallized and fluid characteristics. *Personality and Individual Differences*, 29, 539–553.
- Marlowe, H. A., Jr. (1986). Social intelligence: Evidence for multidimensionality and construct independence. *Journal of Educational Psychology*, 78, 52–58.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616.
- 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., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640–647.
- Neisser, U. (1976). General, academic, and artificial intelligence. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 135–144). Hillsdale, NJ: Earlbaum.
- 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.
- Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, 51, 649–660.
- Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, 9, 185–211.
- Schneider, R. J., Ackerman, P. L., & Kanfer, R. (1996). To “act wisely in human relations”: Exploring the dimensions of social competence. *Personality and Individual Differences*, 21, 469–481.
- Schollaert, E., & Lievens, F. (2008, July). *The effects of exercise instructions on the observability of assessment center behavior*. Paper presented at the 29th International Congress of Psychology, Berlin, Germany.
- Schuler, H., & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183–192). Newbury Park, CA: Sage.
- Schuler, H., & Moser, K. (1995). Die Validität des Multimodalen Interviews [Validity of the multimodal interview]. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 2–12.
- Semadar, A., Robins, G., & Ferris, G. R. (2006). Comparing the validity of multiple social effectiveness constructs in the prediction of managerial job performance. *Journal of Organizational Behavior*, 27, 443–461.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, 51, 125–139.

- Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology, 80*, 587–606.
- 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.
- Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Monthly Magazine, 140*, 227–235.
- Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., & Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance, 17*, 71–93.
- Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436–458.
- Weis, S., & Stüb, H.-M. (2007). Reviving the search for social intelligence: A multitrait-multimethod study of its structure and construct validity. *Personality and Individual Differences, 42*, 3–14.
- Witt, L. A., & Ferris, G. R. (2003). Social skill as moderator of the conscientiousness-performance relationship: Convergent results across four studies. *Journal of Applied Psychology, 88*, 809–821.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.

## APPENDIX

Examples from the final set of future-oriented and past-behavior questions (the targeted dimensions are given in brackets) and for the calculation of ATIC scores on the basis of hypothetical hypotheses for these questions:

*Future-oriented question:* You have newly been placed in charge of a branch. Your employees don't have enough trust in you yet and act somewhat diffident towards you. What would you do? [Leadership Behavior]

Assume an interviewee had written down the following two hypotheses what this future-oriented question was trying to assess and coded them as described:

1. taking initiative for building confidence in my employees and for getting accepted as the new branch manager;
2. convincing my employees about my technical knowledge so that they will accept me better in the long run.

In the first step, the interviewee categorized both hypotheses according to the target dimensions they corresponded to. In both cases, this was Leadership Behavior. In the second step, he rated the strength of this correspondence as 4 (*fits completely*) for H1 and as 2 (*fits partially*) for H2. As a consequence, the ATIC score for this question was 4 as this was the highest rating for the correct target dimension.

*Past-behavior question:* You probably remember such days: Your day was packed with appointments, for example at school, sports, with doctors, and you had invited friends over for a dinner at your place but still needed to do the grocery shopping. Finally, there were several important phone-calls to make that couldn't be delayed. What did you do? [Systematic Planning]

Assume an interviewee had written down the following two hypotheses what this past-behavior question was trying to assess and coded them as described:

1. systematic approach to analyze the problem and take action;
2. stress tolerance

In the first step, the interviewee categorized the first hypothesis as corresponding to Systematic Planning and the second as corresponding to none of the three target dimensions. Then he rated the strength of the correspondence as 4 (*fits completely*) for H1. This rating was used as his ATIC score for the question.

*Past-behavior question:* You definitely are familiar with the following situation: Two good friends of yours have been quarreling about something that you weren't personally involved in. Both friends told you about this quarrel and asked for your advice. You thought that one of them was right and that the other was rather wrong. What did you do? [Cooperation]