APPLIED PSYCHOLOGY: AN INTERNATIONAL REVIEW, 2003, *52* (4), 648–668

The Transparent Assessment Centre: The Effects of Revealing Dimensions to Candidates

Nanja J. Kolk*
Berenschot, Utrecht, The Netherlands

Marise Ph. Born
Erasmus University, Rotterdam, The Netherlands

Henk van der Flier
Vrije Universiteit, Amsterdam, The Netherlands

Quelles retombées peut avoir le fait de dévoiler aux participants les dimensions mesurées dans un centre d'évaluation ? Cette question est abordée dans deux études indépendantes qui font appel à des exercices individuels. Les résultats de la première étude n'indiquent aucune différence dans la validité de construction entre un groupe d'étudiants universitaires "transparent" ($N = 99$) et un autre "non transparent" ($N = 50$) ; ceci est contraire à ce qu'avaient trouvé Kleinmann & al. (1996) et Kleinmann (1997) avec des exercices de groupe. Les évaluations moyennes ne changent pas à l'exception de la "sensibilité" qui augmente légèrement avec la transparence. Par contre, les résultats de la deuxième étude, qui faisait appel à un échantillon de candidats à un poste réel, débouchèrent sur une amélioration significative de la validité de construction chez le groupe "transparent" ($N = 297$) pa rapport au group "non transparent" ($N = 393$). La encore, les évaluation moyennes des deux groupes n'ont pas différé. Les apports de ces résultats pour la pratique et des suggestions pour de futures recherches sont présentés dans cet article.

What are the effects of revealing dimensions to candidates in an assessment centre? This question is addressed in two independent studies, using individual

* Address for correspondence: Nanja J. Kolk, Berenschot, Europalaan 40, 3526 KS Utrecht, The Netherlands. Email: n.kolk@berenschot.com

We gratefully acknowledge LTP in Amsterdam for making this research possible. We are also indebted to Hajo van Wimersma Greidanus for his assistance in conducting the experiment. We would also like to thank Peter Dekker for his helpful suggestions regarding the analyses and the three anonymous reviewers for their comments on an earlier version of this manuscript.

An earlier version of this paper was presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans (April 2000).

exercises. Results in Study 1 showed no difference in construct-related validity between a transparent (N = 99) and a non-transparent group of university students (N = 50), contrary to previous findings by Kleinmann, Kuptsch, and Köller (1996) and Kleinmann (1997), who used group exercises. Also, mean ratings did not alter, the exception being the dimension "Sensitivity", which increased slightly after the transparency treatment. Conversely, results in Study 2, which contained a sample of actual job applicants, showed a significant improvement in construct-related validity for the transparent group (N = 297) compared with the non-transparent group (N = 393). Again, mean ratings did not differ between these two groups. Implications of these findings for practice and suggestions for future research are discussed in this paper.

INTRODUCTION

More than a decade ago, Sackett (1987, p. 21) raised a challenging question to assessment centre (AC) administrators: "Would you object to my providing candidates with a 30-minute coaching session prior to the Assessment Centre?" This question was posed in reaction to the fact that in some ACs the behavioural dimensions are revealed (i.e. they are made "transparent"), while in others they are not. Should the answer to this question be: "No, because it leads to reduced variance", Sackett argued that any variance that a 30-minute session can remove, can by no means be regarded as true score variance.

Disclosing the dimensions to candidates has a number of benefits. First of all, since all candidates have equal access to information, they also have equal opportunity to show dimension-relevant behaviour (Kleinmann et al. 1996). Sackett (1987, p. 21) concurs that the goal of giving instructions to candidates should be to minimise the effects of extraneous factors on ratings through a systematic attempt to put all candidates on an equal footing. Second, transparency has been shown to facilitate eliciting dimension-related behaviours, which, as will be elaborated upon below, benefits construct-related validity of the dimensions (Kleinmann et al., 1996). Third, from a practical standpoint, transparency is an important issue, because almost 30 per cent of the ACs in Spychalski, Quiñones, Gaugler, and Pohley's (1997) survey convey information about dimensions to candidates. Considering that besides transparency there are numerous other ways of determining which dimensions are relevant in an AC (e.g. reading books, talking to others, former AC experience, training, etc.), it is important to be aware of the consequences of transparency for AC performance.

Despite the obvious relevance for AC practice and research, since the Kleinmann et al. 1996 study, only one study has investigated the effects of transparency on construct-related validity (Kleinmann, 1997) and two studies have examined the effects on criterion-related validity (Kleinmann, 1997; Smith-Jentsch, 1996). So far, no study has examined the effects of transparency on mean ratings. This paper reports two independent studies examining

the effects of transparency on construct-related validity of the dimensions on the one hand, and on differences in mean ratings on the other hand.

EFFECT OF TRANSPARENCY ON CONSTRUCT-RELATED VALIDITY

While the AC has been shown to predict various aspects of job performance (Gaugler, Rosenthal, Thornton, & Bentson, 1987), evidence for construct-related validity of AC ratings has not been established convincingly (Neidig & Neidig, 1984; Robertson, Gratton, & Sharpley, 1987; Sackett & Dreher, 1982; Sagie & Magnezy, 1997). Heterotrait-monomethod (htmm) correlation coefficients (discriminant validity) are consistently higher than monotrait-heteromethod (mthm) correlation coefficients (convergent validity). The unitary conceptualisation of validity states that different kinds of validity (i.e. content-, criterion-, and construct-related validity) are merely different strategies for demonstrating the validity of a certain measure's construct (Binning & Barrett, 1989). Therefore, because ACs have predictive validity, they also must have construct-related validity (see also Arthur, Woehr, & Maldegen, 2000). These results have urged many researchers to scrutinise this seemingly contradictory result of the AC's validity, in an attempt to unravel its underlying grounds (see Lievens, 1998, for an overview).

One of the factors proposed to be of influence to construct-related validity was the transparency of the AC's dimensions. Kleinmann (1993), who pioneered in this area of research, investigated to what extent candidates are aware of which dimensions are relevant for a particular exercise. First, he found that dimensions are not transparent per se. Second, he showed that when candidates identified a dimension as being relevant in one exercise but not in another (while both exercises in fact tapped the same dimension), convergent validity was lower than when the dimension was correctly identified in both exercises. Individuals who identify the dimensions more accurately, outperform people who identify the dimensions less accurately. This means that the "demand characteristics" of the AC are different depending on the extent to which people identify dimensions. Kleinmann postulated that, as candidates apparently do not necessarily know which behavioural dimensions are required, they may not show dimension-related behaviour, even though they may have the ability to do so. Thus, dimension-related behaviours are not automatically elicited in every exercise. Therefore, the constructs that the exercises are supposed to tap may in fact not be tapped when the dimensions (i.e. the intended constructs) are not identified. This could prevent finding evidence for construct-related validity. If so, ratings contain unwanted variance accounted for by the candidate's judgment regarding the appropriateness of a targeted dimension (cf. Smith-Jentsch, Salas, & Brannick, 2001).

Kleinmann et al. (1996) investigated the hypothesis that the transparency treatment benefits construct-related validity. This study showed in a between-subjects experiment using a student sample that transparency indeed led to the desired increase in construct-related validity. In their study, Kleinmann et al. used group exercises, but no individual exercises. For this reason, it is conceivable that participants were able to adjust their behaviour towards a dimension because they received behavioural cues from fellow participants who all attempted to modify their behaviour towards that dimension. For instance, it might be easier to be cooperative when one's fellow participant is also trying to be co-operative.[1] Conversely, it may be more difficult to be persuasive when one's fellow participant is also trying to be persuasive. Thus the demand characteristics of group exercises compared to individual exercises differ due to the presence or absence of others, which may affect the outcome of the transparency treatment. In practice, a large number of commonly used exercises involve participants either acting on their own (e.g. an in-basket) or acting vis-à-vis a confederate (e.g. an interview simulation) (Thornton, 1992). Therefore, to be able to generalise the findings regarding the effects of transparency on the AC as a whole, it is also important to study the effects of transparency in individual exercises. Also, it is important to study the effects of transparency on the performance of actual job applicants, as until now only the effects on the performance of students have been studied. Job applicants with more work experience than students may be more able to adjust their behaviour on the basis of behavioural cues.

The purpose of the present study is to further examine Kleinmann et al.'s (1996, 1997) finding that the construct-related validity of AC ratings improves due to transparency. The effect of transparency is observed in two independent studies, Study 1 using a student sample, and Study 2 using a sample of actual job applicants. In both studies we use individual rather than group exercises, to examine whether transparency also has an effect without the presence of others. Besides construct-related validity, mean AC ratings may be affected by the disclosure of initially unidentified dimensions. When participants are made aware of which behaviours are relevant, they may be able to adjust their behaviour accordingly, thus receiving higher ratings. Transparency may induce participants to fake target behaviours. Whereas the influence of faking has been studied within the range of non-cognitive measures (e.g. McFarland & Ryan, 2000), to our knowledge faking in the AC has not been studied. As a first step, both studies examine the effect of transparency on mean AC ratings per dimension.

---

[1] We acknowledge Martin Kleinmann for this helpful suggestion.

STUDY 1

General Method

Summary. The subjects in Study 1 were 149 university students (56 male), a few months before or after reaching their MA (in most cases an Economics, Psychology, or Law degree). Their mean age was 26 (SD = 5). The incentive for participation in this study was receiving training and multi-source feedback. In order to increase their motivation, participants were paid a small fee to take part in the AC. They participated in a one-day developmental AC, which consisted of four interview simulations and several tests and inventories. Participants were randomly assigned to either the non-transparency group (N = 50) or the transparency group (N = 99). The experimental scheme was a pretest-posttest procedure. Both groups participated in a pretest (morning) and a posttest (afternoon) condition, each condition consisting of two interview simulations. Assessors as well as assessees were blind to the true purpose of the study. Data were collected from ACs on seven different occasions during a one-year period (1998–99).

Assessors. To ensure the fidelity of our AC experiment, we used 23 professional assessors, and 11 trained role-players, following the suggestion by Lievens (1998, p. 145) and Thornton (1992, p. 71). The rater–ratee ratio was 2:1. Both the assessors and the role-players received recurring assessor training sessions, focusing on a shared frame-of-reference. Inter-rater reliabilities were moderate (i.e. the mean PPM correlation coefficient r = .62).

Exercises and Dimensions. Four interview simulations were used, between which the content but not the format differed. In all of the interviews, participants were required to change a subordinate's behaviour or attitude, or to persuade him or her to comply with a certain task (e.g. working overtime). Each exercise took 15 minutes to prepare and 15 minutes to play. The dimensions used in Study 1 were Sensitivity, Analytical Skills, and Persuasiveness. These dimensions were rated immediately after the completion of each exercise on a scale ranging from 1 to 5.

Tests and Personality Measures. To test whether the ability to identify dimensions (before they are made transparent) is related to intelligence, the Verbal Analogies sub-test of a Dutch translation of the Differential Aptitude Test was administered (Evers & Lucassen, 1991). To test whether this ability is related to personality, we administered the NEO Personality Inventory Revised (NEO PI R; Hoekstra, Ormel, & De Fruyt, 1996).

Transparency Treatment. Revealing the dimensions to participants was done following the Kleinmann et al. (1996, p. 73) procedure. Participants in

the transparency group received a 15-minute training session, in which the dimensions were explained, including the definitions and examples of effective and ineffective behaviours. As a reminder, participants were given a handout on which all definitions and behavioural examples were printed. They were asked to withhold this information from the assessors. The non-transparency group did not receive this treatment, but instead participated in a session in which information was provided on "irrelevant" topics, such as how to prepare for an employment interview, how to write a letter and résumé, how to look up information on the Internet, etc.

Analyses. We analysed our data by visually examining the MTMM matrices of the transparency and non-transparency group. A formal test was performed on their corresponding covariance matrices. The MTMM covariance matrices of posttest scores within the control and the transparency group (i.e. the effect of transparency between subjects) were analysed with LISREL 8.20 (Jöreskog & Sörbom, 1989). Criteria for evaluating the competing models are measures of overall fit: χ^2/degrees of freedom ratio (should approach 1), χ^2's *p*-value (should not be significant), the Root Mean Square Error of Approximation (RMSEA), which evaluates the closeness of fit given the number of degrees of freedom (should be lower than .05), and the Tucker-Lewis index (nonnormed fit index: NNFI; the model with the highest NNFI should be selected) which is fairly independent of sample size and can be used when comparing models across samples with varying sample sizes (Bentler & Bonett, 1980). Variances of the two methods (i.e. exercises) were set equally throughout the models, as these can be assumed to be roughly similar. A multiple group analysis was performed to test for differences between groups.

Results

Before moving on to the effect of transparency on construct-related validity, we made sure that the dimensions were not transparent beforehand. After the pretest exercises, participants in both groups were asked which three dimensions they thought had been observed, out of a list of ten possible dimensions and their definitions, following the procedure of Kleinmann et al. (1996, p. 74). Results indeed indicated that the dimensions were not transparent beforehand. None of the dimensions were correctly identified by all participants. The mean amount of accurate identifications was 1.7 (scale 0 to 3 correct identifications, SD = .60). The proportion of participants choosing each of the ten dimensions were: Persuasiveness: 28 per cent; Sensitivity: 19 per cent; Firmness: 14 per cent; Decisiveness: 10 per cent; Analytical Skills: 9 per cent; Tenacity: 7 per cent; Assertiveness: 6 per cent; Sociability: 3 per cent; Planning and Organising: 2 per cent; Self-confidence:

2 per cent. Thus, the dimension Persuasiveness was identified most often, Sensitivity less often and Analytical Skills were identified the least often. Note that having identified a dimension does not pertain to the treatment, because we not only revealed the dimension labels, but also their accompanying behavioural examples and effective and ineffective behaviours.

Kleinmann (1993) and Smith-Jentsch (1996) argued that the ability to identify the dimensions adequately does not vary at random between participants, but is related to intelligence and various social skills. Results in this study revealed that this ability was significantly related to verbal intelligence ($r = .27$, $p < .05$), and to the dimension Sensitivity in the second and fourth exercise ($r = .26$; $r = .25$, $p < .05$), but unrelated to personality traits.

Does Construct-Related Validity Improve Due to Transparency? According to the hypothesis that transparency enhances construct-related validity, the mthm correlations of the two posttest exercises should be higher in the transparency group than in the non-transparency group, and the htmm correlations of the two posttest exercises should be lower in the transparency group than in the non-transparency group. Kleinmann et al. (1996) found that the improvement in construct-related validity was greater for the subgroup of participants who stated they had indeed oriented their behaviour towards the dimensions. Thus, we hypothesise that the effect of transparency is greater for the group of people who stated that they oriented their behaviour towards the dimensions, defined as a score ≥ 4 on a Likert scale from 1 to 5. Within the transparency group, the same pattern should emerge comparing pretest correlations to posttest correlations. The results are presented in Table 1. Table 1 should be interpreted as follows: mean heterotrait-monomethod (htmm) correlations are indicative of discriminant validity (the lower the correlations the better the discriminant validity). Mean monotrait-heteromethod (mthm) correlations are indicative of convergent validity (the higher the correlations the better the convergent validity).

Table 2 shows an overview of mean intercorrelations between dimensions and exercises, obtained from the MTMM matrix in Table 1 (also including means and standard deviations). Also, Table 2 shows the dimension and exercise correlations for the subjects in the transparency subgroup who stated that they had oriented their behaviour towards the dimensions (transparency group′), and the transparency subgroup who stated they did not do so (transparency group″).

Visual examination of Table 2 does not support the hypothesis that transparency increases construct-related validity within and between the groups. The mean posttest convergent validities were $r = .31$ for the non-transparency group, $r = .28$ for the transparency group and $r = .30$ for the transparency subgroup′. As for discriminant validity, these values were $r = .40$, $r = .49$ and $r = .47$, respectively. Comparing pretest to posttest within the transparency

TABLE 1
MTMM Matrix: Observed Correlations among Ratings for Study 1

	Non-transparent group														*Transparent group*											
	M	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	M	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Ex. A																												
1 S	5.20	1.87	–												5.16	1.87	–											
2 A	5.60	1.72	.40	–											5.44	1.52	.57	–										
3 P	5.47	1.71	.22	.72	–										5.72	1.73	.22	.51	–									
Ex. B																												
4 S	5.07	1.93	.38	.10	.03	–									5.11	1.90	.22	.04	.03	–								
5 A	5.30	1.80	.28	.22	.13	.55	–								5.24	1.70	.03	.03	.19	.67	–							
6 P	5.53	2.11	.24	.35	.20	.33	.68	–							5.85	1.82	.03	.09	.37	.12	.45	–						
Ex. C																												
7 S	4.70	1.84	.30	.13	−.02	.27	.25	.24	–						5.51	2.02	.25	.01	−.06	.30	.18	.09	–					
8 A	5.58	1.61	.17	.29	.23	.07	.19	.27	.42	–					5.62	1.65	.19	.13	.21	.28	.32	.35	.53	–				
9 P	5.77	1.63	.21	.21	.21	.04	.34	.39	.34	.57	–				5.68	1.79	.11	.17	.32	.22	.23	.41	.35	.68	–			
Ex. D																												
10 S	4.97	1.94	.54	.23	.14	.21	.08	−.01	.27	.23	.29	–			5.12	1.74	.12	−.01	.09	.12	.13	.05	.40	.24	.15	–		
11 A	5.31	1.62	.34	.12	.14	.14	.10	.04	.00	.29	.20	.33	–		5.03	1.63	.25	.16	.20	.07	.11	.05	.20	.21	.14	.54	–	
12 P	5.60	1.76	.26	.39	.29	.16	.14	.23	.10	.35	.36	.27	.50	–	5.63	1.92	.12	.18	.22	.05	.16	.26	.06	.16	.24	.26	.59	–

Note: Ex.: exercise; Ex. A and Ex. B: pretest; Ex. C and Ex. D: posttest; S: Sensitivity; A: Analytical Skills; P: Persuasiveness. Means and correlations are based on the summed ratings of the two assessors.

TABLE 2
Mean Dimension and Exercise Intercorrelations among Ratings for Study 1

	Pretest				*Posttest*			
	NT	*T*	*T′*	*T″*	*NT*	*T*	*T′*	*T″*
Dimension (monotrait-heteromethod)								
Sensitivity	.38	.22	.28	.14	.27	.40	.41	.39
Analytical Skills	.22	.03	−.10	.08	.29	.21	.19	.20
Persuasiveness	.20	.37	.47	.23	.36	.24	.30	.16
Mean (convergent validity)	.27	.21	.27	.15	.31	.28	.30	.25
Mean heterotrait-heteromethod	.19	.07	.07	.18	.19	.16	.17	.16
Exercise (heterotrait-monomethod)								
Exercise A–C	.45	.43	.38	.51	.44	.52	.52	.51
Exercise B–D	.52	.41	.35	.50	.37	.46	.43	.48
Mean (discriminant validity)	.48	.42	.36	.50	.40	.49	.47	.50

Note: NT = non-transparency group, T = transparency group; T′: experimental subgroup ($n = 60$) who stated that they oriented their behaviour toward the dimensions (i.e. who scored ≥ 4; on a Likert scale ranging from 1 to 5); T″: experimental subgroup ($n = 39$) who stated that they did not orient their behaviour toward the dimensions (i.e. who scored < 4; on a Likert scale ranging from 1 to 5).

group, the mean mthm correlations (convergent validity) were $r = .21$ and $r = .28$, respectively. Discriminant validity coefficients were $r = .42$ and $r = .49$, respectively. Regarding the transparency group′, convergent validity coefficients were $r = .27$ and $r = .30$, respectively, whereas discriminant validities were $r = .36$ and $r = .47$.

As in the majority of studies examining MTMM data (e.g. Lance, Newbolt, Gatewood, Foster, French, & Smith 2000), we ran into estimation problems using the traditional CFA approach when testing some of the competing models. In an attempt to overcome empirical problems (i.e. out-of-range estimates and convergence problems) with the traditional CFA approach, some researchers have suggested the so-called correlated uniqueness (CU) approach which is not subject to the aforementioned problems (Kenny & Kashy, 1992; Marsh, 1989). This approach specifies trait factors and does not create method factors, but allows its unique factors to correlate across measures within the same method. Kenny and Kashy (1992, p. 170) noted that a potential downside of the CU model is that it assumes zero method-method correlations. When this assumption is not met (e.g. when the methods are similar), it can have a biasing effect by artificially enhancing convergent validity and worsening discriminant validity. For each of the competing CFA models, we therefore report the traditional model as well as the CU model and ran further analyses using models with admissible solutions.

TABLE 3
Fit Indices for the CFA Models of Posttest Ratings for Study 1

Model	*Admissible?*		*df*	χ^2		χ^2/df		*NNFI*		*RMSEA*	
	NT	*T*		*NT*	*T*	*NT*	*T*	*NT*	*T*	*NT*	*T*
1-3D2E	No	No	2	1.28	.37	.64	.18	1.11	1.07	.00	.00
1′-3D-CU	Yes	Yes	3	1.77	1.03	.59	.34	1.12	1.06	.00	.00
2-2D-2E	No	Yes	4	2.21	3.46	.55	.86	1.12	1.01	.00	.00
2′-2D-CU	Yes	Yes	5	2.88	6.83	.58	1.37	1.13	.96	.00	.06
3-1D-2E	Yes	Yes	5	3.75	6.13	.75	1.23	1.07	.98	.07	.05
3′-1D-CU	Yes	Yes	6	4.70	7.50	.78	1.25	1.06	.98	.00	.05
4-0D-2E	Yes	Yes	11	8.38*	25.54*	.76	2.32	1.06	.88	.00	.12
4′-0D-CU	Yes	Yes	9	14.26	23.29*	1.58	2.59	.78	.85	.11	.13
5-3D-0E	Yes	Yes	9	17.54*	68.20*	1.95	7.58	.73	.40	.14	.26
Null	Yes	Yes	15	80.23*	195.33*	5.35	13.02	.00	.00	.30	.35

Note: #D and #E = number of dimension and exercise factors; NT = non-transparency group, T = transparency group; CU = correlated uniqueness model; admissible? = whether all model parameter estimates were contained within admissible ranges; df = degrees of freedom; NNFI = nonnormed fit index; RMSEA = root mean square error of approximation.
* $p < .01$.

Table 3 reports the results of the CFA: χ^2/df ratio, the χ^2's *p*-value, the NNFI, and the RMSEA of the competing posttest models of the non-transparency group and the transparency group. The transparency group was not split in the CFA, because the sample size of the transparency group″, who stated they had not oriented their behaviour towards the dimensions, was only 39. Results from Table 3 suggest that the traits × methods CU model fits well under both conditions, regarding the fit indices (the traditional CFA model does not yield an admissible solution). Neither in the non-transparency group nor in the transparency group, did the three-factor model provide a significantly better explanation than a one-factor model.[2]

A multiple-group analysis was performed on model 1′ to test whether the loading patterns in the non-transparency and in the transparency group differed significantly. χ^2 difference tests indicated that there was no significant difference between the two matrices regarding discriminant and convergent validity.

Does Transparency Affect Mean AC Ratings? To test whether mean ratings were affected by the transparency treatment on a dimension level, a covariance analysis was conducted on the posttest scores, using the pretest

[2] The model parameter estimates of Study 1 can be obtained from the first author.

scores as the covariate. Pretest scores were calculated by summing up the first two exercises; posttest scores by summing up the last two exercises. First of all, means and standard deviations on the pretest did not differ significantly between groups. Therefore, posttest scores did not need to be standardised. Paired samples *t*-tests revealed no significant increases or decreases in mean ratings. The mean posttest scores of Sensitivity in the non-transparency group decreased compared to pretest scores, but not significantly. In other words, there was no pretest-posttest effect. Also, there appeared to be no decrease in variance in the transparency group as compared to the non-transparent group. Levene's tests showed no significant differences in variance within the groups.

Results of the covariance analyses (performed in three separate analyses) indicated significant differences between pretest and posttest only for scores on Sensitivity (F: 3.61, $p < .05$). However, this result may have been influenced by group differences. In the non-transparent group posttest scores on Sensitivity decreased, whereas in the transparent group these scores increased. Therefore, the difference between groups is more likely to reach significance, and should therefore be regarded cautiously.

Discussion

Results of Study 1 demonstrate that the predicted increase in construct-related validity, as shown by Kleinmann et al. (1996, 1997), cannot be reaffirmed using individual exercises. The MTMM matrices do not show that the correlations altered in the expected direction. CFA shows no significant difference in construct-related validity between the two groups. This is also true for the transparent group′ (those participants in the transparent group who stated they had adjusted their behaviour toward the dimensions).

The second research question dealt with the issue of whether mean ratings alter due to transparency. Covariance analysis showed no significant difference in mean AC scores between the transparent group and the non-transparent group. An exception was the dimension Sensitivity. Ratings on this dimension increased significantly after the transparency treatment. Apparently, being nice and showing understanding to other people is something participants can indeed pick up. Yet, this significant increase might well have been strengthened by the fact that mean ratings within the non-transparent group decreased, whereas mean ratings within the transparent group increased. Consequently, the difference between pretest and posttest is more likely to reach a level of significance, and should not be over-interpreted.

One possible alternative explanation for the lack of influence of transparency in Study 1 was that our student sample might have been unsuitable

for the research issue under study (though Kleinmann et al. [1996] used the same type of subjects). Students may lack the behavioural repertoire to make adjustments after they are told which behaviour is expected, when taking part in individual exercises. The demand characteristics of this type of exercise (reflected in the behaviour of the role-player) might be too subtle for students to grasp. It is conceivable that a sample of experienced job applicants would show a stronger effect of transparency. In order to test this alternative explanation, Study 2 examines the effect of transparency in a between-subjects design using a sample of actual job applicants.

STUDY 2

Method

Summary. Six hundred and ninety Dutch job applicants (470 male) participated in one-day ACs at a Dutch consultancy firm during the year 2000. They applied for a variety of jobs, mostly in management. The ACs were part of the regular selection procedure. The mean age of the participants was 35 (SD = 9). The AC consisted of several psychological tests and inventories, an interview simulation, in which participants had a one-to-one talk with a subordinate (this was the same type of exercise as Study 1), and an analysis/presentation exercise, in which participants presented a fictitious business problem and defended a solution to this problem to the "board of directors" (i.e. two confederates). The dimensions measured in Study 2 were Sensitivity, Sociability, Judgment and Tenacity. Together with the instructions for each AC simulation, 297 participants received a written handout which explained the dimensions, including their definitions and examples of effective and ineffective behaviours. The other group of 393 candidates did not receive this treatment. The two groups do not differ with respect to demographic variables and the jobs for which they applied. The assessors came from the same pool of assessors as Study 1. In this study, the assessors were told, in general terms, that applicants in the transparency condition received information on the target dimensions. Thus, strictly speaking, they were not blind to the experimental treatment. They were, however, blind to the experimental research question (does transparency affect construct-related validity?). In addition, because the assessors rotated in each exercise (i.e. they observed each candidate only once), possible alterations in convergent validity would be meaningful and could not be attributed to cross-exercise bias. Also, the experiment ran for several months, and the sample does not include applicants from the first few weeks after we began running the transparency condition. This makes an effect of the assessors' knowledge of the study's purpose unlikely.

Results

Before moving to the hypothesis testing, we looked at how participants perceive the transparency treatment in terms of stress. On the one hand, transparency may cause people to become more agitated due to increased behavioural demands. On the other hand, transparency may decrease perceived stress due to being familiar with the exercise requirements. However, Kleinmann et al. (1996) showed that their participants did not differ in subjectively perceived stress. This may be different for actual job applicants, whose interest in the AC's outcome is greater. We asked candidates in Study 2 afterwards how agitated they felt during the course of the AC (rated on a 5-point scale). As in Kleinmann et al.'s study, we found no significant difference ($p < .05$, two-sided) between the transparent and the non-transparent group.

A visual examination of the MTMM matrix displayed in Table 4 conveys an increase in convergent correlations (mthm) from $r = .22$ to $r = .31$, and quite similar discriminant correlations (htmm), $r = .52$ and $r = .50$, respectively. These results are comparable to those of Kleinmann et al.'s (1996), which found an increase in mthm correlations from $r = .30$ to $r = .35$, and similar htmm correlations, $r = .60$ and $r = .61$, respectively.

The fit indices displayed in Table 5 suggest that the 3 traits × 2 methods CU model fits under both conditions (as in Table 3, the traditional CFA model in Table 5 did not yield an admissible solution).[3] For the non-transparent group, model 1′ did not fit significantly better than the more restrictive model 2′. Therefore, model 2′ provided the best explanation for the data, also in terms of fit indices. Compared to the more parsimonious one-factor model (model 3′), model 2′ fitted significantly better. To summarise, for the data in the non-transparent group, only two dimension factors and CUs were necessary to explain the variance in the ratings. As for the transparent group, model 1′ fitted significantly better than model 2′, and yielded a non-significant χ^2. The χ^2/degrees of freedom ratio, the NNFI, and the RMSEA showed an adequate fit. Therefore, model 1′, which incorporates all three dimension factors and CUs, provided the best explanation for the data in the transparent group.

The next step was a multiple group analysis for both conditions. First of all, we compared the loading pattern in the non-transparent and transparent groups using model 2′. This resulted in a significant χ^2 test ($\chi^2 = 29.47$ [10], $p < .01$). Thus, the CU model with two latent dimension factors does not fit under both the transparent and the non-transparent conditions. Next we compared the loading patterns in both groups using model 1′. Results

[3] The model parameter estimates of Study 2 can be obtained from the first author.

TABLE 4
MTMM Matrix: Observed Correlations among Ratings for Study 2

	Non-transparent group								*Transparent group*							
	M	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	M	*SD*	*1*	*2*	*3*	*4*	*5*	*6*
1. Sensitivity	55.78	18.07	–						54.69	17.08	–					
2. Judgment	57.52	16.59	.62**	–					56.52	16.04	.58**	–				
3. Tenacity	62.60	15.50	.31**	.64**	–				60.64	15.07	.33**	.69**	–			
4. Sociability	60.83	13.60	.18**	.13*	.19**	–			61.12	14.93	.31**	.22**	.16**	–		
5. Judgment	58.44	17.04	.11*	.14**	.14**	.52**	–		58.81	17.94	.15**	.25**	.24**	.42**	–	
6. Tenacity	59.92	16.01	.12*	.22**	.33**	.41**	.62**	–	59.30	15.98	.15**	.30**	.38**	.33**	.66**	–

Note: Correlations are based on the summed ratings of the two assessors.
* $p < .05$; ** $p < .01$.

TABLE 5
Fit Indices for the CFA Models of Posttest Ratings for Study 2

Model	*Admissible?*		*df*	χ^2		χ^2/df		*NNFI*		*RMSEA*	
	NT	*T*		*NT*	*T*	*NT*	*T*	*NT*	*T*	*NT*	*T*
1-3D2E	No	No	2	206.12*	6.88*	103.06	3.44	–1.32	.94	.51	.09
1′-3D-CU	Yes	Yes	3	13.81*	7.37	4.60	2.46	.93	.96	.10	.07
2-2D-2E	Yes	No	4	13.37*	13.49*	3.34	3.37	.95	.93	.08	.09
2′-2D-CU	Yes	Yes	5	15.99*	13.59*	3.20	2.72	.96	.96	.07	.08
3-1D-2E	Yes	Yes	5	24.27*	27.75*	4.85	5.55	.92	.87	.10	.12
3′-1D-CU	Yes	Yes	6	42.16*	36.44*	7.03	6.07	.88	.87	.12	.13
4-0D-2E	Yes	Yes	11	88.18*	61.79*	8.02	5.62	.86	.88	.13	.12
4′-0D-CU	Yes	Yes	9	62.75*	68.91*	6.97	7.66	.88	.82	.12	.15
5-3D-0E	Yes	Yes	9	324.86*	192.56*	36.10	21.39	.31	.45	.30	.26
Null	Yes	Yes	15	773.63*	639.89*	51.57	42.66	.00	.00	.37	.36

Note: #D and #E = number of dimension and exercise factors; NT = non-transparency group, T = transparency group; CU = correlated uniqueness model; admissible? = whether all model parameter estimates were contained within admissible ranges; df = degrees of freedom; NNFI = nonnormed fit index; RMSEA = root mean square error of approximation.
* $p < .01$.

showed again a significant χ^2 test ($\chi^2 = 21.16$ [6], $p < .01$). This means that the CU model with three dimension factors does not fit equally well under both conditions.

Covariance analyses (performed separately for each dimension and exercise) revealed no significant increases or decreases in mean ratings.

GENERAL DISCUSSION

This study contributes to the literature by seeking answers to the question "What are the effects of revealing dimensions to candidates?" Initially, we saw in Study 1 that transparency does not seem to affect mean ratings or the construct-related validity of these ratings when a student sample is used. As mentioned previously, students may lack the behavioural repertoire to be able to modify their behaviour towards the revealed dimensions. Nonetheless, Kleinmann et al. (1996), who did find an effect on construct-related validity, also used student subjects. An important difference between the Kleinmann et al. (1996) and Study 1 was that the former used group exercises, whereas the latter administered individual exercises. The students in the Kleinmann et al. study might have had more demand characteristics at their disposal due to the presence of others. To ascertain that the absence of the effect of transparency in Study 1 is not due to the type of subjects, and to be able to generalise the results to AC practice, Study 2 used a larger sample

of actual job candidates, in a between-subjects design. Besides having a broader behavioural repertoire, job applicants may also be more motivated to use the behavioural cues from the transparency condition. In Study 2, the expected positive effect of transparency on construct-related validity does turn up. Yet, mean ratings remain unaffected. Thus, it would seem that transparency leads to more consistency in behaviour with respect to the various dimensions and not to higher scores.

A limitation of both studies reported in this paper is the limited number of different exercises included in the ACs. However, when combining our findings with the Kleinmann et al. studies (1996, 1997), we can conclude that transparency seems to have little effect on inexperienced participants taking part in individual exercises, but does have a greater effect on inexperienced participants when they are interacting with others. Also, transparency has an effect on the performance of actual job candidates with work experience taking part in individual exercises. In both studies, the participants' mean ratings were not affected by transparency. We also ran several analyses on AC ratings, verbal intelligence, and personality, yet we found no moderators on the effect of transparency. These conclusions leave the question open as to whether the performance of actual job candidates taking part in group exercises alters due to transparency. In view of Kleinmann et al.'s finding that students were able to modify their behaviour in group exercises, it seems likely that job applicants would do the same.

How Does Transparency Relate to Practice, Tutoring, and Coaching?

Generally, techniques aimed at improving test performance are classified in three categories: practice, tutoring, and coaching (Maurer, Solamon, & Troxtel, 1998, p. 128). Practice is aimed merely at becoming acquainted with a particular test; learning from experience. Tutoring involves instruction within the content domain measured by the test. Coaching concerns a variety of instructional techniques to improve test performance. Maurer et al. noted that coaching may involve identification and explanation of the dimensions being measured. As such, transparency could be regarded as coaching. In the absence of feedback, however, transparency may best be described as tutoring.

Earlier studies on the effect of instructional interventions on AC performance yielded mixed results. For instance, research on performance in Leaderless Group Discussions has shown that practice did not affect the ratings, nor did a brief description of the definitions of the dimensions (tutoring) (Denning & Grant, 1979). However, a short training session, involving both a dimension description (including examples of effective performance) and practice with feedback, did enhance performance (Kurecka,

Austin, Johnson, & Mendoza, 1982; Petty, 1974). Performance on in-basket exercises appeared to be positively affected by specific training in several studies (Brannick, Michaels, & Baker, 1989; Brostoff & Meyer, 1984; Gill, 1982), although one study did not find any training effect (Jaffee & Michaels, 1978). With regard to the interview simulation, one study showed that taking part in a management effectiveness training programme improved performance significantly (Moses & Ritchie, 1976).

The results of both studies that transparency does not alter mean ratings add to the literature on coaching instructions that merely disclosing the definitions of the dimensions and their behavioural examples does not affect AC ratings substantially. A concern expressed by Petty (1974), that participants may be able to fake certain characteristics after having been made aware of the required AC dimensions, is not upheld by the present study. In conclusion, it seems that if AC performance is affected, it probably only will be so by specific coaching, and not by mere tutoring or practice.

Implications for Research and Practice

As stated in the introduction, it is not uncommon in operational ACs to convey dimensions to candidates. Spychalski et al. (1997) noted in their survey of US-based AC practices, that almost 30 per cent of the respondents indicated that their assessees receive information concerning the dimensions prior to an AC. It is also important to study the effects of transparency in view of the fact that transparency is not a manipulation one can either implement or leave out. As said earlier, applicants have numerous ways of finding out which dimensions are important in an AC. The following section deals with the question which implications the previous and present studies have on transparency, and which questions are left open for future research.

The ability to recognise rating dimensions appeared to differ between individuals (Kleinmann, 1993). This finding was a starting point for a study on transparency, focusing on its potential *dis*advantages. Smith-Jentsch (1996) hypothesised that transparency would decrease criterion-related validity because people who are able to make these adequate inferences concerning the required behaviours in an AC will also make these inferences on the job. Transparency eliminates the possibility of measuring these inferences, hence decreasing the predictor–criterion correlation. This study resulted in a significant decrease in the predictive validity of the transparent group as compared to the non-transparent group, using a later self-evaluation as a criterion (Smith-Jentsch, 1996). Yet, the limited sample size of this study (i.e. $N < 20$) prevents drawing any sound conclusions. Since this study only included an assessment of the dimension Assertiveness, no construct-related validity outcomes were available.

Following Smith-Jentsch's (1996) rationale, Kleinmann (1997) used a design that included a first (non-transparent) AC as a criterion and a subsequent (transparent) AC (taking place on the same day) as the predictor. Kleinmann postulated that if the correlation between the first AC and the second AC decreased in the transparent group, transparency affects criterion-related validity negatively. This hypothesis was supported by his data. Still, the fact that this correlation decreased does not necessarily say much about the predictive power of the transparent AC, compared to the traditional AC. The alteration in construct-related validity between pretest and posttest could by itself have led to a lower correlation between pretest and posttest.

In conclusion, both Smith-Jentsch (1996) and Kleinmann (1997) expressed their doubts concerning the appropriateness of revealing dimensions to candidates with regard to predictive validity, despite its presumed benefits for construct-related validity. However, both have been unable to show convincing evidence for this idea. Results in the present two studies, however, do not give reason to expect that transparency has a major impact on mean AC ratings.

Nevertheless, we feel that it might be worthwhile studying differences in criterion-related validity between a transparent and a non-transparent AC. As a first step towards external validation, in Study 2 we examined for both groups the correlations between AC ratings and verbal intelligence, which has been reported as being related to some, but not all, aspects of AC performance (Lance et al., 2000). It appeared that the correlation between verbal intelligence (measured by the DAT sub-test Verbal Analogies) and AC performance on the dimension Tenacity in a cognitive exercise (i.e. the analysis/presentation exercise) was significantly higher (Fisher's $z = 1.99$, $p = .02$) for the transparent group than for the non-transparent group (respectively $r = .38$ and $r = .19$). Though there is some discussion over the question whether the AC represents typical or maximum performance (Sackett, Zedeck, & Fogli, 1988), these results suggest that at least some of the behaviours in a transparent AC may lean towards maximum rather than typical performance. How this result is to be interpreted in terms of criterion-related validity of the transparent AC, depends on one's standing on whether, in general terms, the AC should or should not correlate with cognitive ability. Schmidt and Hunter (1998) maintain that the AC has little incremental validity over General Mental Ability (GMA). This suggestion is countered by recent findings showing that GMA does not uniformly show positive relationships with AC ratings (Lance et al., 2000). Future empirical research should conduct an actual criterion study in order to examine the effect of transparency on criterion-related validity.

To summarise, Study 1 has shown no significant improvement in construct-related validity using a student sample and individual exercises. Yet Study 2, using actual job candidates participating in individual exercises, did

reveal improved construct-related validity. Both studies showed that mean AC ratings do not alter substantially due to transparency. This is good news for practitioners because faking does not seem to be an issue when the dimensions are made transparent. Thus, there is also little reason for practitioners to fear self-help books on how to survive an AC, which obviously make the AC largely transparent for those who read these books.

Other research questions remain unanswered. For instance, how do candidates *perceive* being made aware of the dimensions? We know that they do not get more agitated, as was shown by Kleinmann et al. (1996) and confirmed in Study 2. But do they also perceive being treated more fairly? Are candidates more willing to accept feedback when they are aware of the behaviour requirements prior to the exercises? Future research should be directed at obtaining empirical insight into the effect of transparency on candidates' perceptions and reactions regarding procedural fairness (candidates' perception of fairness of the selection procedure), perceived control (candidates' perception of control over task performance) (Macan, Avedon, Paese, & Smith, 1994), and perceived predictive validity (candidates' perception of how well the procedure predicts future job performance) (Ployhart, Ryan, Conley, & West, 1999).

REFERENCES

Arthur, W., Woehr, D.J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center contruction-related validity paradox. *Journal of Management*, *26*, 813–835.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*, 478–494.

Brannick, M.T., Michaels, C.E., & Baker, D.P. (1989). Construct construct-related validity of in-basket scores. *Journal of Applied Psychology*, *74*, 957–963.

Brostoff, M., & Meyer, H.H. (1984). The effects of coaching on in-basket performance. *Journal of Assessment Center Technology*, *7*, 17–21.

Denning, D., & Grant, D. (1979). Knowledge of assessment process: Does it affect candidate ratings? *Journal of Assessment Center Technology*, *2*, 7–12.

Evers, A., & Lucassen, W. (1991). *Handleiding DAT '83. Differentiele Aanleg Testserie*. Lisse: Swets & Zeitlinger.

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.

Gill, R.W.T. (1982). A trainability concept for management potential and an empirical study of its relationship with intelligence for two managerial skills. *Journal of Occupational Psychology*, *52*, 185–197.

Hoekstra, H.A., Ormel, J., & De Fruyt, F. (1996). *NEO PI-R/NEO FFI; Big Five persoonlijkheidsvragenlijsten*. Lisse: Swets & Zeitlinger BV.

Jaffee, C., & Michaels, C. (1978). Is in-basket performance subject to coaching effects? *Journal of Assessment Center Technology*, *1*, 13–17.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd edn.). Chicago, IL: SPSS Inc.

Kenny, D.A., & Kashy, D.A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.

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

Kleinmann, M. (1997). Tranaparenz der Anforderungsdimensionen: Ein Moderatur der Konstrukt- und Kriteriumvaliditat des Assessment-Centers. *Zeitschrift fur Arbeits- und Organisationspsychologie*, *41*, 171–181.

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

Kurecka, P.M., Austin, J.M., Johnson, W., & Mendoza, J.L. (1982). Full and errant coaching effects on assigned role leaderless group discussion performance. *Personnel Psychology*, *35*, 805–812.

Lance, C.E., Newbolt, W.H., Gatewood, R.D., Foster, M.R., French, N.R., & Smith, D.B. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance*, *13*, 323–353.

Lievens, F. (1998). Factors which improve the construct construct-related validity of assessment centers: A review. *International Journal of Selection and Assessment*, *6*, 141–152.

Macan, T.H., Avedon, M.J., Paese, M., & Smith, D.E. (1994). The effects of candidates' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, *47*, 715–738.

McFarland, L.A., & Ryan, A.M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*, 812–821.

Marsh, H.W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*, 335–361.

Maurer, T., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology*, *83*, 128–136.

Moses, J.L., & Ritchie, R.J. (1976). Supervisory relationships training; A behavioral evaluation of a behavior modeling program. *Personnel Psychology*, *29*, 337–343.

Neidig, R.D., & Neidig, P.J. (1984). Multiple assessment center exercises and job relatedness. *Journal of Applied Psychology*, *69*, 182–186.

Petty, M.M. (1974). A multivariate analysis of the effects of experience and training upon performance in a leaderless group discussion. *Personnel Psychology*, *27*, 271–282.

Ployhart, R.E., Ryan, A.M., Conley, P.R., & West, B.J. (1999). Effects of test preparation programs on perceived predictive validity and self-efficacy. Paper

presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, April.

Robertson, I., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, *60*, 187–195.

Sackett, P.R. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology*, *40*, 13–25.

Sackett, P.R., & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, *67*, 401–410.

Sackett, P.R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*(3), 482–486.

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

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.

Smith-Jentsch, K.A. (1996). Should rating dimensions in situational exercises be made transparent for participants? Paper presented at the 11th Annual Conference of the Society for Industrial and Organizational Psychology Conference, San Diego, CA, April.

Smith-Jentsch, K.A., Salas, E., & Brannick, M.T. (2001). To transfer or not to transfer? Investigating the combined effects of trainee characteristics, team leader support, and team climate. *Journal of Applied Psychology*, *86*, 279–292.

Spychalski, A.C., Quiñones, M.A., Gaugler, B.B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, *50*, 71–90.

Thornton, G.C. (1992). *Assessment centers in human resource management*. Reading, MA: Addison-Wesley.