

Predicting job success with the assessment center: Validity or artifact?

HEINZ HOLLING UND WOLFRAM REINERS

WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER, BUNDESREPUBLIK DEUTSCHLAND

1995

1 Description of problem and aims of this study

The selection of suitable employees is of fundamental importance to any organization. Within the broad range of methods used in personnel selection to discern the suitability of candidates Assessment Centers (ACs) have received considerable attention. It is not just that a whole number of studies have demonstrated that ACs fulfil such important test criteria as objectivity, reliability and predictive validity. They also meet many of the other justified demands made on personnel selection methods such as the the consideration of 'social validity' suggested by Schuler and Stehle (1983): In an AC information about the characteristics of the workplace and the organization is made available to candidates, management participates in the development and use of selection instruments, the implementation of the procedures and the resulting conclusions are transparent for both assessor and assessed and consequently they can relatively easily be communicated. These qualities have made the ACs highly acceptable to both candidates and those responsible for personnel selection within organizations (Holling & Leippold, 1991).

A multiplicity of individual studies, review articles and meta-analyses (e.g. Gaugler, Rosenthal, Thornton & Bentson, 1987; Hunter & Hunter, 1984; Schmitt, Gooding, Noe & Kirsch, 1984) deal with the question of the predictive validity of AC procedures. The validity coefficients which are reported in these studies lie within the range $r=0.3$ to $r=0.4$. Among the indicators of job performance criteria are appraisal by superiors, by colleagues, and self-appraisal based on anecdotal evidence or behavioral observation, assessment of potential made by managers, performance at training events, variables affecting career development such as number of

promotions, salary level or the increase in salary levels over time, job changes etc. (e.g. Thornton, Gaugler, Rosenthal & Bentson, 1992). But these indicators all share a common disadvantage. Because the management of an organization shares in both the design of the AC-procedures and in the selection of potential employees as assessors, their norm and value systems and their preconceptions of the qualities required in successful candidates heavily influence their assessment of those candidates. On the other hand, performance criteria are invoked to justify the predicative validity of the AC which are equally influenced by the norms, values and preconceptions of management. If one assumes that the socialisation process in organizations leads to a homogenisation of the norms, values and preconceptions of management, a link between predictor and criterion is always to be found. This effect, known as criterion contamination, leads to an overestimate of actual validity. A similar argument is put forward by Klimoski and Brickner (1987), Klimoski and Strickland (1977) and Maukisch (1986).

Bungard (1992) has shed some light on the problem from the perspective of social psychology. He argues that as a result of the values and norms of the observer, possibly even before the commencement of an exercise a judgement may have been reached about whether a future colleague or subordinate will fit in with the organization. Potential candidates who are able to identify the norms and culture of an organization will, in so far as they behave in conformity to these norms, be positively assessed. In a similar way this is also true for progress within the organization at a later date. The ability to conform and strategic behaviour, rather than performance, are rewarded with promotion, salary increases and positive assessments from superiors. Consequently, the predictive validity reported in long-term study of ACs can be deceptive, in the usual sense of the ability of ACs to predict future performance.

The aim of this current study is to critically examine the tenability of these objections. Its distinctiveness lies in the fact that the indicators of job performance selected are not dependent on subjective assessment. In contrast to earlier studies of ACs, the sales figures of insurance salesmen have been applied as indicators of job performance. These figures offer an objective performance criterion. If the hypothesis is correct that the empirically determined validity coefficient of ACs is to a great extent due to contamination resulting from the subjectivity of job performance criteria, then the validity coefficients of the ACs in terms of criteria for measurement

of job performance should in this study be significantly lower than the average validity coefficient in other studies ($r = 0.37$; Thornton et al., 1992). Objective job performance criteria have been used in only a few studies to assess the validity of the aptitude tests used in ACs. Squires, Torkel, Smither and Ingate (1991) have examined the test criteria of a situational test used to appraise the selling ability of sales personnel. It produced a validity of 0.31 with respect to sales quotas. Weekley and Gier (1987) found a predictive validity ($r=0.45$) of a situational interview with reference to the volume of sales achieved by sales staff.

2 Methods

The aims and objectives of the AC arose out of the need to select suitable sales staff for an insurance company involved in a broad range of insurance activities. The AC was developed under the direction of a consulting firm, which has long specialized in personnel development methods using ACs.

The AC lasts a total of four days. The first day is set aside for preparation and for refresher-training for the assessors. The AC exercises and simulations take place on the second and third days. During these exercises the assessors record the progress of the participants. The fourth day is used for an assessor's conference and a feedback session with the participants. Between 3 and 8 candidates participated in each AC ($M=5.3$, $s=1.57$). The ratio of assessors to participants was between 1:2 or better.

2.1 AC predictors

The core of the AC consists of six exercises. The performance of the participants are rated by the observers by means of a series of dimensions, with a different set of dimensions being applied in each exercise. Unfortunately, within the exercises only data for the mean of each exercise is available and not a score for each individual dimension. This is because at the outset the study had only a very general design. Due to the longitudinal nature of the study this shortcoming could not be corrected. The observer-rating is based on a scale of four levels for each exercise:

1 = seldom/hardly observable

2=occasionally/sporadically observable

3=regularly observable

4=prominent/strongly observable.

A participants score is the mean of all ratings according to all dimensions observed in an exercise.

In a conference following the exercises the assessors then agree on a score for each participant for each of the ten dimensions, the ratings being based on the observations recorded during the exercises. While the same basic four-level scale is used, intermediate scores (eg. 2.5, 3.5 etc.) are also possible. The final dimension rating for a participant is therefore the product of a group consensus of the scores awarded by the observers.

The selection of the ten dimensions emerged out of job analysis during the design phase of the AC and they reflect the ten most important elements affecting the success of an insurance representative. Table 1 gives a short description of the ten dimensions, the means and standard deviations and table 2 a short description of the contents of the six exercises together with their means and standard deviations.

Table 1: The dimensions of the AC. The mean (M) and standard deviations (s) of the dimensions are given in brackets.

Dimensions	Description of the dimension constructs
1 persistence	objective always kept in focus; does not allow oneself to be distracted from essentials; but not stubborn or obstinate (M=2.77 ; s = 0.42)
2. resistance to stress	able to withstand emotional, psychological and cognitive pressure ; no sign of nervousness/ tenseness shown ; normal emotional reactions present (M = 2.91 ; s = 0.42)
3 initiative	able to develop one's own ideas or the ideas of others and press ahead with them; appears actively involved ; achieves results ; but not unfocused activism (M = 2.70 ; s = 0.42)
4. sociability	able to establish relationships with and between others and to develop them; open and responsive to others ; eye-contact (M = 2.88 ; s = 0,42)
5. achievement orientation	able to set and pursue high standards of achievement; ambitious; goal-centered ; able to accept confrontation (M = 2.75 ; s = 0.39)
6. learning and adaption	able to be self-critical and when necessary revise opinions; to accept and apply new ideas; intellectual ability (M = 2.72 ; s = 0.45)
7.personal appearance	able to appear relaxed and open to others; to positively influence the emotional state of a conversation partner (M = 2.88 ; s = 0.54)
8. independence	able to make judgements, decisions and negotiate independently; able to express clear opinions; to make decisions even in the absence of all information (M = 2.85 ; s = 0.47)
9. self-confidence	confidence to express oneself in negotiations; able to realistically assess people and situations (M = 2.84 ; s = 0.40)
10. negotiating skills	ability to lead conversations and discussions in a direction decided by oneself, hold it there and so achieve one's goals. But not to outmanoeuvre somebody (M = 2.69 ; s = 0.49)

Table 2: The AC exercises or simulations. The mean (M) and standard deviation (s) for each exercise are given in brackets.

AC exercise	Description
1. self-presentation	Each participant introduced themselves to the group; one minute preparation time allowed (M = 2.57 ; s = 0.31)
2. leaderless group discussion	Each participant was expected to form a judgement about a complex topic and convey their point of view to the rest of the group. Group discussion (without chairman) followed with the aim of arriving at a conclusion acceptable to all participants. (M = 2.45 ; s = 0.42)
3. oral presentation	Participants were given 15 - 30 minutes to prepare an outline itinerary for a group outing. Each had then five minutes to present his/her ideas in the most convincing way to the group. (M = 2.58 ; s = 0.37)
4. Exploratory Interview	After a short preparation time each participant conducted an exploratory interview with one of the other members of the group. The aim was to discover the interviewee's opinion on a given subject. However, the interviewee was not to be directly informed about the true aim of the interview. (M = 2.62 ; s = 0.44)
5. Dealing with objections	Participants were given a case-study involving some element of conflict and asked to form an opinion on the case. The observers then discussed this opinion with the participants individually, seeking to confront him/her with objections to the judgement formed. (M = 2.75 ; s = 0.44)
6. Planning	Participants were each given a list of imaginary tasks, to be completed within a particular time scale and within given constraints. The order in which the tasks were performed was important so that participants had not only to plan the tasks but also the order in which they were done. (M = 2.38 ; s = 0,64)

On the basis of this data a decision is made at the observer conference whether to accept or reject candidates. The rate of selection at these ACs is about 50%. Due to data protection regulations no data is available about unsuccessful candidates.

2.2 *Sales figure as an objective indicator of job performance*

Job performance is measured by the sales figures of sales representatives. An internal coding system, whereby points are awarded or deducted according to the type and range of policies sold or cancelled, allows a calculation to be made of sales and income for each individual representative. The number of points gained by each sales representative during the first year of employment (V12) with the company

serves as the indicator of job performance in this study and has the advantage that it is an indicator that will not be distorted by monthly or seasonal fluctuations in sales.

Although this indicator provides an objective criterion of job performance it is, of course, only a partial measure. Other elements such as the securing of a long-term customer base, subjective job satisfaction, the long-term retention of sales representatives within the company, team support and the development of a representative's management potential might also form part of the construct 'job performance'. These other elements are not, or are only partly, covered by our own indicator. However, actual sales figures are seen within the profession, by both employers and employees, as the most important criterion of job performance and job satisfaction. The system in general use whereby representatives' salaries are dependent on sales figures supports this point of view.

2.3 *Sample*

Not all applicants reached the AC. Particular selection criteria were applied to select participants for the AC including age, job-related training, previous career patterns, number of job changes within and between companies, possession of a driving license, family situation. One can assume that these criteria were not applied too rigorously, so that the sample, as a consequence was not too heavily pre-selected.

The sample (n=105) represented a wide variety of previous professional experience, had varied educational backgrounds and an average age of 36.2 (s=7.9). These and other biographical variables were seen to have no observable influence on the results of the AC and later sales figures, so that no special sub-samples have been analysed.

3. **Results**

3.1 *Validity of the dimensions*

Table 3 contains the validity coefficients of the ten dimensions. It is clear that they help predict later sales figures, but each to varying degrees. The validity coefficients are in general low. Only 'achievement motivation' and 'self confidence' correlate in a statistically significant way with job performance during the first year. In a multiple regression the dimensions 'achievement motivation' and 'perseverance' contribute to

the prediction of job performance (cf table 4.) with a multiple correlation coefficient of $R=0.31$ ($F\text{.....}$). Both the weight and the level of the multiple correlation coefficient remain stable for a number of random samples.

Some univariate validity coefficients of the dimensions are low but positive, and not statistically significant. However, if they are ignored there is a danger of a type II error because of the expected low effect sizes and the small sample size. Because one cannot expect the dimensions to be independent of one another a factor analysis is necessary to provide information about the factor structure on which the dimensions are based. In an analysis of the main components two factors were extracted which together explain 60.1% of variance (Scree-Test, Cattell, 1966, and Kaiser-Guttman Criteria, Guttman, 1954; Kaiser & Dickmann, 1959). Table 3 illustrates the two factors in terms of the dimensions by means of a varimax rotation and gives the validity of the factor values. Diagram 1 illustrates the factor loading of the dimensions.

These results show that a combination of the dimensions 'perseverance', 'initiative' and 'achievement motivation' has a validity of ($r=0.222;p<0.05$). The dimensions 'ability under stress', 'sociability', 'ability to learn and adapt' and 'personal appearance', on the other hand are not valid. (cf table 5).

3.2 Validity of the predictors

The validity coefficients of the predictors can be inferred from table 3. Only exercise 6 (planning) has a significant and middle validity. In a multiple regression only the planning exercise contributed to a prediction of sales figures ($p=<0.01$).

The results of the AC exercises were also submitted to factor analysis. Using the same criteria as above two factors were extracted by means of an analysis of the main components. These explain 65.3% of total variance. Table 5 contains both a description of the factors in the AC exercises and gives the validities of the factor values. Diagram 2 shows the factor loading of the exercises.

From these results it is clear that of all the exercises the 'planning' exercise furnished the most independent results. In contrast to the results of the other exercises, the

results of the planning exercise are considerably more valid in respect to sales figures during the first year of employment.

4. Discussion: validity or artifact?

At the outset the question was asked whether the results of ACs are really as reliable in predicting future job performance as recent research indicates. Critical evaluation of earlier studies of the predictive validity of ACs seems to suggest the possibility that a considerable proportion of the observed connection between AC predictors and subjective indicators of job performance could be the result of criteria contamination. This alternative interpretation of the validity of ACs would mean that ACs do not, as originally intended, measure the characteristics which are required for success in a particular job. Rather such characteristics as the ability to 'fit in' with the particular elements of an organization's culture, and the skill to present ones own abilities are brought to the fore. Because on both sides the same characteristics, equally irrelevant to later performance, are emphasised, one would expect a corresponding level of validity to be observable. However, the use of an objective criterion of job performance as in this study should mean the presence of a subjective and irrelevant element on the criteria side will have no tendency to exaggerate predictive validity.

As can be shown, even when objective measurements of performance are used the AC in this study has a predictive validity, which in terms of quality is comparable with other studies, which use less objective job performance criteria (cf tables 3,4 & 5).

It might be suspected that the reported validity coefficients are based on a selected sample. However, in the AC in question only 50% of the participants were considered suitable and for the rejected participants naturally no sales figures are available. Thus the observed validity of the AC represents, due to restricted variance, a conservative estimate of the predictive validity of the selection procedure for the applicant sample. This could, in fact, be greater than the results mentioned above.

This partly reduces the force of the criticism that criteria contamination is the main cause of the observed relationship between AC predictors and the non-objective (subjective??) job performance criteria. If this objection were correct one

would expect the validity coefficients in our study to be considerably lower than those in studies operating with subjective criteria. Thus a comparison of the validities of this sample with the validities of other samples is necessary as no subjective job performance data is available.

Whether these results can be reproduced for other professions remains questionable as very often it is impossible to identify objective performance criteria. Nevertheless, it is difficult to establish a theoretical case for saying that the job performance predictions in the case of insurance salesmen should fundamentally differ in principle from such predictions in other occupations.

Table 3: Validity coefficients of the scores for character dimensions awarded at observer meetings following the completion of the AC exercises, and of the mean exercise scores, with reference to the sales results achieved during the first year with the company (hereafter S12)

Character Dimension	V12	n	Exercise Scores	V12	n
1.Persistence/Determination	-.001	105	1 Introductions	.030	95
2.Ability under stress	-.053	105	2 Group Discussion	.047	95
3.Initiative	.177	105	3 Presentation	.028	95
4.Ability to communicate with others	.127	105	4 Exploratory Interview	-.034	94
5.Achievement motivation	.218 [*]	105	5 Dealing with objections	.157	94
6.Ability to learn and adjust	-.009	105	6 Planning	.401 ^{**}	56 ¹
7.Personal manner	.116	105			
8.Independence/Self-Reliance	.165	105			
9.Self-Confidence	.213 [*]	105			
10. Negotiating ability	.184	105			

¹. This sample is smaller than that for the other exercises because the planning exercise was introduced into the AC at a later stage. There is no reason to suppose that the results are due to the sampling effect.

^{*} p < 0.05 (double sided (???)aspect) test)

^{**} p < 0.01 (double-sided (???)aspect) test)

Table 4: Multiple regression of the job performance indicator S12 (cf table 1) on the character dimensions, showing the character dimension, the partial regression coefficient (B), the standard partial regression coefficient (β) and the error probability (p).

Character Dimension	B	β	p
5 Achievement Motivation	516,31	0,215	0,0658
9 Self Confidence	564,77	0,229	0,0517
1 Persistence/Determination	-519,94	-0,227	0,0547
hier fehlt „Konstante“	-91,90		0,9060

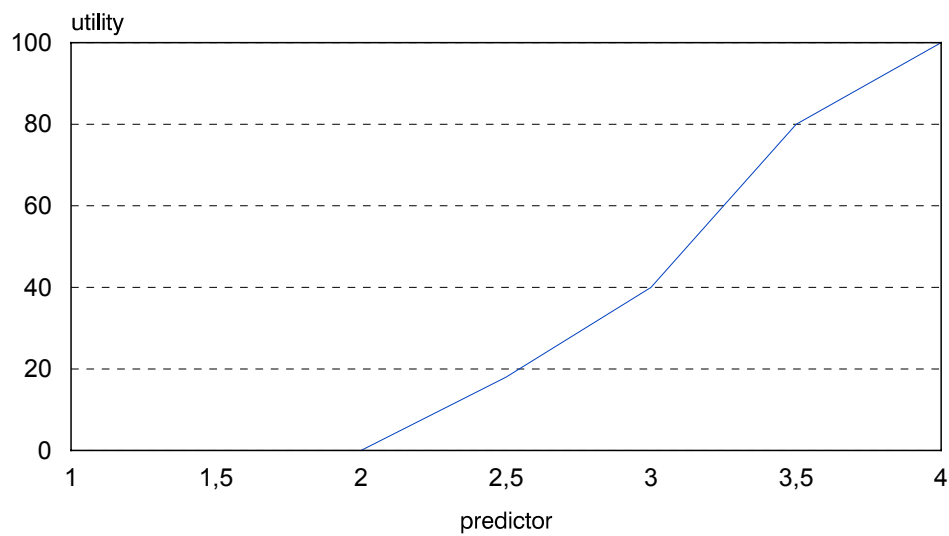


Figure 1: Utility function of a predictor

Table 5: The character dimensions (n = 105) and AC exercises (n=22) described by the factors and the validity of the factor values with reference to job performance indicator S12 (cf table 1).

Character Dimension	V12	n	Exercise Scores	V12	n
1.Persistence/Determination	,040	105	1 Introductions	,157	95
2.Ability under stress	,048	105	2 Group Discussion	,065	95
3.Initiative	,203*	105	3 Presentation	,096	95
4.Ability to communicate with others	,232*	105	4 Exploratory Interview	,021	94
5.Achievement motivation	,218*	105	5 Dealing with objections	,114	94
6.Ability to learn and adjust	,136	105	6 Planning	,403**	56 ¹
7.Personal manner	,140	105			
8.Independence/Self-Reliance	,167	105			
9.Self-Confidence	,224*	105			
10. Negotiating ability	,244*	105			

¹. This sample is smaller than that for the other exercises because the planning exercise was introduced into the AC at a later stage. There is no reason to suppose that the results are due to the sampling effect.
* p< 0.05 (double sided test)
** p< 0.01 (double-sided test)