

Human Relations, Vol. 48, No. 4, 1995

The Managerial Performance Implications of a Developmental Assessment Center Process

A. S. Engelbrecht^{1,3} and A. H. Fischer²

The purpose of this study was to determine if a developmental assessment center process improves managerial performance in the workplace. Focus was placed on the behavioral level of evaluation. The research design made use of a two-group design with random selection and a control group. A sample of 76 managers, at supervisory level, was used. Behaviorally anchored rating scales were developed to measure the job performance of participating managers. The results indicated significant differences between the experimental and the control groups for six performance dimensions. Significant differences were also found in all the second-order factors and the total managerial performance score. Thus, the developmental assessment center process had a positive impact on managerial performance, and this effect was still measurable 3 months after center attendance. In order to generalize the results of this study, it is essential to do further research on the utility of the developmental assessment center.

KEY WORDS: assessment center; management development; managerial performance; evaluation of management development.

INTRODUCTION

Over the last four decades, assessment center technology has evolved from being used only for selection to having at present diverse applications (Woodruffe, 1990). Given the evolution of assessment center technology from the early selection center to the newest development centers, one would expect research to reflect the same progression. However, by far the most research has been done on assessment centers used for selection purposes (Thornton &

¹Senior Lecturer in Industrial Psychology, Department of Industrial Psychology, University of Stellenbosch, P.O. Box X5018, 7599 Stellenbosch, South Africa.

²Old Mutual Life Assurance Society, South Africa.

³Requests for reprints should be addressed to Amos S. Engelbrecht, Senior Lecturer in Industrial Psychology, Department of Industrial Psychology, University of Stellenbosch, P.O. Box X5018, 7599 Stellenbosch, South Africa.

Bentson, 1987). Assessment centers as predictors of performance is one of the best research methodologies in industrial psychology (Woodruffe, 1990; Kriek, 1991; Cascio, 1991). In contrast, assessment centers for development seem to be underresearched (Robertson & Rout, 1989). Research on the management development application is mostly aimed at the feedback process (Slivinsky, McDonald, & Bourgeois, 1979; Thornton & Byham, 1982; Fleenor, 1988) and assessor exposure and benefits (Lorenzo, 1984; Beardsley, 1985).

The developmental approach has become necessary because of the need of participants to understand and agree with center results, cost-effective procedures, and an increased focus on management development (Griffiths & Allen, 1987; Baldwin & Padgett, 1993). The assessment of managerial skills that is gained through an expensive, comprehensive, and time-consuming process should be put to development uses (McCloskey & Slivinsky, 1983).

The ultimate objective of management development is to enhance effectiveness, specifically the job performance of the manager and the performance of the organization as a whole (Berry, 1990). In order for effective performance to occur a match between the manager's individual competencies, the job's demands and the organizational environment needs to be present (Boyatzis, 1982). Management development thus focuses on the managerial competencies, especially the skills level, in order to enhance performance (Spangenberg, 1990).

A large variety of management development vehicles are presently being used (Marsh, MacCormick, & Robinson, 1989). These can primarily be categorized as formal training, on-the-job training (coaching, appraisals, projects, action learning, task forces, and feedback from the manager's senior), and developmental assessment centers. In focusing on developmental assessment centers, Boehm (1985) identifies five applications for this technique in management development:

1. The feedback session in itself, by providing insight and formulating development strategies. The objective is to convert self-diagnosis into performance (Thornton & Byham, 1982). Three major components influence the effect of feedback on performance, namely the source of the feedback, the characteristics of the feedback, and the characteristics of the recipient. Further, feedback must influence three interrelated cognitive processes before it will be translated into practice: the perception of the feedback (accuracy), the acceptance of the feedback, and the motivation or willingness to use the information in future tasks (Camp, Blanchard, & Huszczo, 1986, p. 121).

Orpen and King (1989) found that subordinates' reactions to feedback from superiors are relatively unaffected by perceived credibility or expertise. Responses were, however, significantly affected by the kind of

feedback. Ilgen and Moore (1987) further found that when evaluating performance on more than one dimension (such as an assessment center), it may be useful to provide feedback separately on the dimensions, and to allow recipients to choose feedback on each dimension in order to reduce redundant information and the time needed to receive and evaluate the feedback. In a recent study by Fleenor (1988), participant responses indicated a positive perception of the assessment center feedback and a fairly high level of acceptance. Furthermore, Jones and Whitmore (Baldwin & Padgett, 1993) found that participants' self-efficacy and perceived support of developmental activities had a significant impact on feedback acceptance from the developmental assessment center. In contrast to this, Frank, Seifeik, and Jaffe (Boehm, 1985, p. 42) cited evidence of participants attending assessment centers as many as five times without demonstrating any performance improvement, despite feedback. In summary, however, more of the relevant research evidence indicates that the feedback in itself can be of developmental value when it is detailed and behaviorally specific (Boehm, 1985). Goodge (Baldwin & Padgett, 1993) discusses guidelines and pitfalls of development centers, which he defines as an off-site process resulting in effective development actions that go beyond the feedback and assessment of a traditional assessment center.

2. The early identification of managerial talent in order to prepare the identified individuals for rapid advancement in the organization. An important objective of early identification programs is thus the acquisition of valid information to guide the use of development resources (Thornton & Byham, 1982).

3. The identification of strengths and development areas (at the skills level) of managers in order to formulate individual training and development programs. The latter can be either formally based or on-the-job. Participation in an assessment center can create insight due to a process of self-evaluation. Schmitt, Ford, and Stults (1986, p. 332) investigated changes in self-perceived ability as a product of assessment center participation. Participants ($N = 1693$) provided self-ratings on eight ability dimensions, before and immediately after the center without receiving feedback. The results indicate that significant change in self-perceptions do occur as a result of center participation, even in the absence of specific feedback from either an observer or an administrator (Schmitt et al., 1986).

The information generated during a developmental assessment center is used primarily for identifying managerial strengths and weaknesses (ratings per dimensions). Based on this quantitative data, observers also make developmental recommendations by combining this information with more qualitative data (observations, trends during the assessment center). This information is then conveyed to the participant during the feedback session.

Often additional recommendations and action plans are formulated during the feedback, by incorporating the input and views of the participant and his/her senior.

Participation in an assessment center is seen as a development experience providing certain assumptions underlie developmental assessment centers. Boehm (1985) identifies three: First, it is assumed that particular persons can actually improve their skills in those dimensions measured by an assessment center. Research evidence appears mixed, as related to an individual's ability to change basic managerial skills. Skill improvement can only be assumed, according to Thornton and Byham (1982, p. 335), when participants "are subjected to a major developmental effort targeted to the individual's particular and specific needs." Second, it is assumed that participants will be sufficiently motivated to undertake and persist in developmental activities. Third, it is assumed that the participant is provided with sufficient developmental experiences to positively impact on managerial skill deficiencies as identified by the assessment center.

4. Observer training and practice serve as a developmental experience for line managers throughout the organization. Observer training and subsequent service improves observing, recording, categorizing, and evaluating of behavior (Boehm, 1985).

5. The assessment center as a tool for organizational planning and development.

The historical failures to evaluate both costs and benefits of development programs have rendered many management development functions vulnerable for cost-cutting (Carnevale & Schultz, 1990). Evaluation can justify the cost of developmental interventions, establish their effects on participants, measure the effect on job performance and measure the effect on the profitability, performance, flexibility, or survival of the organization as a whole (Harrison, 1989).

A wide variety of evaluation procedures and methods exist to evaluate development programs (Phillips, 1991, pp. 44-51; Bushnell, 1990; Tannenbaum & Yukl, 1992). The Kirkpatrick (1979) typology, however, remains the prevalent framework for categorizing training criteria (Tannenbaum & Yukl, 1992). Kirkpatrick proposes four levels for evaluating programs namely, reaction, learning, behavior and results. Reaction may best be defined as how well the trainees or participants like the program. However, there is no assurance that any learning or behavior change has taken place at this stage. Neither is there any indication that the program resulted in any benefits for the organization as a whole. The learning level deals with the principles, facts, and techniques understood and cognitively absorbed by participants. The third evaluation level deals with performance or behavior in the workplace. When learning does not transfer to the job, the

two most likely reasons are that the work environment does not support the learning, or that the participant thinks the program was irrelevant (Carnevale & Schultz, 1990). Results evaluation deals with the impact of the development program on the effectiveness of the organization. Factors generally considered are changes in output, costs, turnover, accident frequency, and profitability in departments or the organization as a whole (Harrison, 1989).

In a meta-analysis of previous training studies, Alliger and Janak (1989) examined the correlations among the four levels of training effectiveness. They found virtually no relationship among trainee reactions and other levels, but slightly higher correlations among the other levels. However, Alliger and Janak's findings were based on a small number of studies (Tannenbaum & Yukl, 1992). Alliger and Janak (1989, p. 3340) concluded that the assumptions of several other models of training evaluation criteria, most of them very similar to Kirkpatrick's, can also be logically questioned.

A study by Noe and Schmitt (1986) revealed important data to be considered when evaluating development programs. They found that participant satisfaction with training (reaction level) was not related to learning and the latter was not related to behavior change. The data supported only a significant relationship between behavior change and performance improvement (results level). These results support the collection of multiple criterion data, when evaluating development programs.

The fundamental issue of the developmental application is whether development actually takes place and improves job performance of managers. Although inferential evidence exists, probably not one empirical study has yet been done to measure directly the impact on performance at a behavioral level. Only one study (Fleenor, 1988) could be found which investigated this matter. Fleenor (1988) used a written situational test for analysis purposes. The problem with this research is that one cannot be certain that learning (situational test responses) will transfer to the workplace (Fleenor, 1988). Further, the evaluation measure was pitched only at the second level (learning) of the Kirkpatrick model (Kirkpatrick, 1979). Transfer of learning to the workplace (third level) and actual performance improvement could be only inferred (Carnevale & Schultz, 1990). These deficiencies highlight the need to do evaluation research at the behavioral level.

It thus becomes evident that the empirical research in the field remains focused on the selection center, while industrial developments have progressed to the developmental assessment center and, recently, to the purely development or collaborative center (Woodruffe, 1990). The present trend in industry seems clearly to favor the developmental use of the assessment center method (Dulewitz, 1991). More and more organi-

zations can therefore be expected to start using development centers, and unless research follows this trend, the present empirical gap will only increase.

Internationally a movement away from traditional assessment centers toward developmental assessment centers can be seen. Despite the enormous financial and human resources spent, very little research evidence backs developmental assessment centers. The objective of this research is therefore, to determine whether participation in a developmental assessment center, feedback session and subsequent on-the-job development, improves managerial performance. This research intends to evaluate the effect of a developmental assessment center process on participating managers, focusing on the behavioral level of the Kirkpatrick model.

METHOD

Hypotheses

The object of this research yields three hypotheses:

Hypothesis 1. Participation in a developmental assessment center and feedback session, followed by on-the-job developmental, will significantly improve managerial performance as measured according to the following skills (dimensions): (a) action orientation, (b) task structuring, (c) development, (d) empathy, (e) managing information, (f) probing, (g) synthesis, and (h) judgment.

Hypothesis 2. Participation in a developmental assessment center and feedback session, followed by on-the-job development, will significantly improve managerial performance as measured according to the following skill-clusters: (a) human resources management cluster, and (b) problem resolution cluster.

Hypothesis 3. Participation in a developmental assessment center and feedback session followed by on-the-job development, will significantly improve overall managerial performance as measured by behaviorally anchored rating scales (BARS).

Sample

The multiracial, multigender sample used for this study consists of 76 first-level supervisors, employed by a large South African assurance society. The average age of the respondents was 28 years, 26% was non-white, and 50% female.

Research Design

The approach which was followed in the evaluation of the developmental assessment center can be classified as action research (Baird, Schneider, & Laird, 1983, p. 205). It occurs in real situations and all the normal constraints are imposed as the organization goes about doing its development work. This use of field settings unavoidably undermines some conditions necessary for proper experimental designs (Baird et al., 1983, p. 196). In choosing a suitable design, the principles of both rigor (quality and quantity of information and control of variance) and practicality (time, expense, and feasibility) are applied in organizational settings (Carnevale & Schultz, 1990).

The sample of 76 managers was randomly selected from managers who had been screened and accepted for assessment centers by a screening committee. The design consisted of a two-group design, with an experimental group ($n = 41$) and a control group ($n = 35$). A pre- and post-test design was not used due to the sensitizing effect BARS measurements could have on respondents. This sensitizing effect was a possibility because other techniques than BARS were used as performance appraisal techniques in the particular organization.

The experimental group consisted of those participants who were scheduled to attend developmental assessment centers over a 10-month period. The control group consisted of different participants who had also been accepted, but were on a waiting list for the same period (the waiting list was a function of participant availability on specific dates and numbers which could not be accommodated on centers). The purpose was to select two groups who had both been accepted to attend centers in order to control independent variables. Neither subjects nor groups could be assigned randomly, due to practical considerations such as scheduling and limited numbers attending assessment centers.

Due to the relatively limited control of extraneous independent variables, other than through random selection, some relevant variables were statistically tested to determine whether a significant difference occurred between the experimental and control groups. The results are reported in Table I. These independent variables were the same as those used by the screening committee to select managers for center attendance and are essentially, criteria for managerial effectiveness in the current job.

From Table I, it is evident that no significant ($p < 0.05$) differences exist between the experimental and control groups, with regard to variables which could influence managerial performance. A reasonable deduction could be made that important extraneous independent variables were seem-

Table I. Differences Between Control ($N = 35$) and Experimental ($N = 41$) Groups Before Treatment

Variable	Experimental group		Control group		T-value	P-value
	<i>X</i>	<i>SD</i>	<i>X</i>	<i>SD</i>		
Performance appraisal scores	3,56	0,50	3,65	0,54	-0,80	0,42
Number of subordinates	6,29	4,11	5,06	5,60	1,11	0,27
Managerial experience in months	21,41	17,74	16,20	8,63	1,67	0,10
Managerial courses attended	2,63	1,26	3,06	1,11	-1,54	0,13
Age	29,05	5,51	27,57	5,87	1,13	0,26
Managerial level in grade (Peromnes)	12,10	0,66	12,10	0,56	0,08	0,93

ingly equal and thus controlled. In addition, a control group and random selection further safeguard the control of the research.

Evaluation was conducted 3 months after a developmental assessment center, in order to establish the transfer of learning and skills over the long term (Kirkpatrick, 1979, p. 86).

Measuring Instruments

This research utilized two measuring instruments. Behaviorally anchored rating scales (BARS) were used in order to determine whether managerial job performance improved as a result of participation in the developmental assessment center process.

Developmental Assessment Center

The developmental assessment center was used to measure the strengths and development areas of managers in order to formulate individual training and development programs. The developmental assessment center is pitched at the supervisory level of management. The process and simulations (with the exception of the feedback) is classically that of the conventional assessment center. Thorough job analysis was used to ensure that simulations and dimensions be related to important job activities.

The simulation exercises used are an in-basket, performance interview and a team exercise. The following managerial dimensions and clusters are

measured in the assessment center as depicted in Fig. 1. The four clusters of dimensions were postulated on an *a priori* basis.

The assessors received 5 days of formal training and acted as co-assessors on two centers prior to their first assessors experience. They are all line managers of the organization, and function at levels senior to that of participants. Each participant are observed by two assessors as the participants go through the various exercises. The assessors much reach consensus on ratings for each dimension, for each participant they observe.

The process followed in this developmental assessment center can be categorized in three phases. First, the assessment center itself, where the participant goes through the exercises as he/she would have done with the conventional assessment center. Second, there is a feedback session subsequent to the center attendance. Feedback is given by the assessors to the senior of the participant, detailing strengths and weaknesses and also focusing on developmental action plans for the participant. A detailed and lengthy feedback is also given to the participant in an interactive (facilitative) manner. A copy of the participant's assessment center report is also given to the senior and the participant. In addition, the participant completes a workbook during the feedback session which details his/her developmental actions and behavior observed during the center. Third, on-the-job development is undertaken by the participant assisted by his/her senior, based on the recommendations and action plans of the feedback session.

Behaviorally Anchored Rating Scales (BARS)

After participation in a developmental assessment center, job-performance of participating managers were evaluated by their direct seniors.

DIMENSIONS	CLUSTERS
Action orientation	Action management cluster
Task structuring Empathy Development	Human resources management cluster
Managing information	Information management cluster
Probing Synthesis Judgment	Problem resolution cluster

Fig. 1. Managerial dimensions and clusters measured in the developmental assessment centre.

Behaviorally anchored rating scales (BARS) were developed for this purpose (Fischer, 1992). BARS are judgmental scales developed to define the rating points in terms of observable and well-researched job behaviors (Gatewood & Feild, 1990; Kingstrom & Bass, 1981; Spangenberg, Esterhuysen, Visser, Briedenhann, & Calitz, 1989).

The procedure used to construct BARS for this study was based on the pioneering work by Smith and Kendall (1963). The method involves five steps, which are iterative. First, the generation of dimensions/competencies which describe areas of required job effectiveness. Second, the generation of behavioral statements representing effective, ineffective, and average performance for the job in question. Third, these behavior statements are allocated to the appropriate dimensions. Fourth, reallocation of the behavioral statements to dimensions, by a separate but comparable group of judges. Last, these behavior examples for each dimension are rated on a scale ranging from poor performance to outstanding performance. Only those behavior examples for which there is high agreement among the judges are retained. The smallest measures of standard deviation is normally used for this purpose (Gatewood & Feild, 1990).

During the comprehensive job analysis phase of the developmental assessment center, dimensions and behavior examples were generated by utilizing the critical incident technique. Behavior examples were assigned to each dimension, and managers (at higher levels) were asked to rate these examples in terms of effective, ineffective, and poor performance. The net result was the formulation of dimensions, definitions of these dimensions, and behavior examples per dimension arranged over five levels of performance (poor to outstanding).

In constructing the BARS for this study, three steps were followed. First, the existing behavior examples were analyzed by an expert panel (14 senior line managers) to retain those examples which were clear and readily understandable for use by non-assessment center staff. Other examples were also more clearly defined and it was ensured that these examples adequately covered the performance levels of excellent, adequate, and poor. These behavior statements were then rearranged in a random fashion in order to ensure more objective rating in the next step. Second managers (at higher levels than those to be rated) were asked to assign values to the behavior examples according to a 5-point rating scale (5 would represent outstanding performance and 1 would represent very poor performance).

Last, the final BARS were constructed by calculating the mean and standard deviation for each behavior example as rated by the managers. Each behavior example was analyzed in order to only retain those with the lowest standard deviations. A high-level of dispersion could otherwise lead to unreliability in the interpretation of the examples. Then, behavior ex-

amples with mean scores closest to the high, medium-high, average, medium-low, and low positions were selected in order to ensure an adequate spread of behavior examples over the vertical scales. Finally, the process resulted in eight behavior scales that were designed as BARS.

The behavior anchored rating scales reflect the same dimensions employed in the assessment center and serve as a aid around which the feedback and development recommendations are structured.

RESULTS

Table II contains a summarized version of the results. From Table II, it is evident that there is a significant difference between the experimental and control groups in action orientation, task structuring, development, empathy, managing information, and probing. No significant difference was found in either synthesis or judgment. The results thus indicate an improvement in managerial performance regarding the six first-mentioned dimensions.

Table II also demonstrates the effect of the developmental assessment center on each cluster (accumulated dimensions) of managerial skills. The experimental group scores are significantly higher for all the managerial clusters. The results also indicate that there is a significant difference be-

Table II. T-Values, Means, Standard Deviations and Omega-Values for All the Dependent Variables

Variable	Experimental group		Control group		T-value	W2-value
	X	SD	X	SD		
Action orientation	3,64	0,60	3,23	0,68	2,85**	0,0857
Task structuring	3,35	0,51	3,07	0,47	2,42*	0,0601
Development	3,45	0,62	2,92	0,73	3,36***	0,1192
Empathy	3,37	0,69	3,00	0,73	2,27*	0,0518
Managing information	3,62	0,57	3,32	0,64	2,15*	0,0455
Probing	3,35	0,57	2,95	0,49	3,29**	0,1145
Synthesis	3,35	0,59	3,12	0,70	1,57	—
Judgment	3,33	0,60	3,10	0,60	1,71	—
Human resources management	10,17	1,41	9,00	1,36	3,65***	0,1395
Problem resolution	10,03	1,38	9,16	1,36	2,76**	0,0801
Overall managerial performance	27,47	3,32	24,71	3,12	3,70***	0,1431

* $p < 0,05$.

** $p < 0,01$.

*** $p < 0,001$.

tween the experimental and control groups in the problem resolution cluster, despite the fact that the difference in scores for synthesis and judgment (these two dimensions combined with probing form this cluster) is not statistically significant.

Table II also indicates that there is a statistically significant difference between the experimental and control groups for the total score of managerial performance ($p < 0.001$). The scores of the experimental group are significantly higher than those of the control group.

Statistically significant differences indicate whether a relationship exists between the independent (developmental assessment center) and dependent (managerial job performance) variable. It does not indicate the strength of the relationship. The omega square (indication of practical significance) indicates the strength of the treatment effect. It thus indicates the proportion of the total variability in a set of scores that can be accounted for by the independent variable (Shavelson, 1981). The omega square values in Table II, indicate that between 4,55% and 11,92% of the total variance in the dimensions of managerial performance can be attributed to the developmental assessment center. Similarly, between 4,55% and 13,95% of the total variance in the clusters and 14,31% of the total variance in the overall measure of managerial performance can be attributed to the developmental assessment center.

The developmental assessment center process can thus be considered to be accountable for 14,3% of the total variance in overall managerial performance as measured by BARS, 3 months after center attendance.

These results support all three hypotheses, except for the individual dimensions of synthesis (hypothesis 1g) and judgment (hypothesis 1h). In addition, omega values of practical significance were obtained for all the variables which demonstrated statistically significant differences. The overall managerial performance, human resources cluster, development, and probing dimensions exceeded the 10% level of practical significance.

DISCUSSION

From the results, it is evident that significant statistical differences were found between the experimental and control groups in all the dimensions except synthesis and judgment. The greatest difference was found in development. These results indicate that the developmental assessment center process had the greatest effect on the skill of development (especially development of subordinates). The next greatest differences were found in action orientation and probing.

No significant difference was found in either synthesis or judgment. This seems to indicate that the developmental assessment center process

had no significant effect on improving those dimensions which are of a cognitive nature. Thornton and Byham (1982, p. 402) support this finding in stating that: "There is general agreement that change is quite difficult in personality characteristics such as cognitive style or flexibility." Boehm (1985) also supports this notion that decision-making ability does not change easily or appreciably. Both synthesis and judgment can be categorized as cognitive skills, using Schroder's (Spangenberg, 1990) model of managerial effectiveness.

Further, significant statistical differences were found between the experimental and control groups in all the managerial clusters. These results indicated that the developmental assessment center had a significant effect on all the clusters of managerial performance. The highest improvement seems to be in the area of human resources (task structuring, development, and empathy). This is supported by Thornton and Byham (1982, p. 402) in that dimensions such as sensitivity (similar to empathy), leadership, management control (similar to task structuring) and oral communication seem to show the biggest improvement over time.

Although the differences between the experimental and control groups in synthesis and judgment were not significant, the overall cluster of problem resolution was highly significant ($p < 0,01$). This improvement could be attributed to the influence that the remaining dimension of this cluster (probing) had on the overall difference. In considering the definitions of the above three dimensions, it becomes evident that probing is more behavioral than synthesis and judgment, which are cognitive.

The effect of the developmental assessment center process seemed to be highly significant ($p < 0,001$) on overall managerial performance, which is a summation of all the dimension variables.

Only two studies could be found which investigated the effects of the developmental assessment center on subsequent performance. An unpublished study by Barber (Thornton & Byham, 1982, p. 329) found that participants who received feedback from the assessment scored significantly higher on performance criteria than the group who received no feedback. A study by Fleenor (1988) also indicates that participants were able to improve their managerial performance after receiving feedback. A situational test was used for measurement purposes (the test provides hypothetical situations and requires written responses by giving participants choices). The posttest scores for the experimental group were significantly higher than the posttest scores for the control group ($t = 2,28; p < 0,05$) (Fleenor, 1988, p. 121). A limitation of this study was the fact that measurement was limited to the second level (learning) of the Kirkpatrick model. The degree of actual transfer of this learning to the job is thus unknown.

Practical significance (that is, the strength of the relationship between the developmental assessment center and performance) was also indicated by the results. An omega value of 14,31% was found for overall managerial performance. This indicates that 14,31% of variance in managerial performance can be attributed to the developmental assessment center process. Another measure of practical significance is that of effect size. It can be described as the normalized difference between the treatment group and a comparison group. According to this formula (Burke & Day, 1986, p. 237) effect size is equal to the difference between the scores of the experimental group and the control group divided by the within-group standard deviation. Cohen (1977) provides the following guidelines for the interpretation of effect size: 0,2 (small effect); 0,5 (medium effect), and 0,8 is a large effect. A measure of 0,5 is considered to be an indication of practical significance (Fleener, 1988, p. 151). The effect size for this study is 0,85 which constitutes not only a large effect, but a very strong indication of practical significance for the developmental assessment center process.

The results are supported by those of Fleener (1988, p. 151) who found an effect size of 0,64 for the developmental assessment center improving job performance. However, the level of measurement was limited to the learning level of Kirkpatrick's model as opposed to the behavior level of this study. Further support comes from Landy and Farr (Fleener, 1988, p. 151) who conducted a literature review on the effect of feedback on managerial performance. They concluded that the effect size of evaluation and feedback was 0,60.

Another meta-analysis by Guzzo, Jette, and Katzell (Fleener, 1988, p. 151) on the effect of various intervention programs on productivity found an effect size of 0,41 for appraisal and feedback.

Given the above standards, the effect of the developmental assessment center on the experimental group (effect size = 0,85) seems strong and of considerable practical significance.

CONCLUSION

The results of this study indicate that the developmental assessment center process and subsequent coaching have a positive impact on managerial performance. This is supported by measures of both statistical and practical significance. The effect also seems to persist over an extended period (3 months) and successful transfer of learning is thus demonstrated.

The improvement in managerial performance could be attributed to the fact that the post-center feedback sessions were specific and behaviorally based. In addition to the participants' receiving feedback, they also gained insight into the effect of both positive and negative behavior. Prag-

matic action plans for formal and on-the-job development also were formulated with the participant and his/her senior. This process resulted in a personal development plan which formalizes the development action of an individual manager.

This study therefore provides significant and new evidence on the effect on performance of assessment center methodology used for development. It also provides the first direct evidence on the behavioral effect (third level of the Kirkpatrick model) of centers on managerial performance, without having to make tenuous inferences.

The conclusions drawn are obviously limited by a few possible shortcomings of the study. Like the majority of validity studies on assessment centers (Sackett, 1987), this study also has the possible limitation that the results are not totally free from criterion contamination. The methodological problem is that the seniors of the participants were given detailed feedback on the results of the developmental assessment center, and they provided the performance ratings. This could have created demand effects which could discount the results. The Management Progress Study (Thornton & Byham, 1982) is one of only two empirical studies in which results were not made available to the participants' seniors or the organization at large. The dilemma with which the empirical researcher is faced is that from a developmental, financial, and ethical viewpoint, the business organization is not prepared to withhold feedback from the assessment center to the seniors of the participants, who are tasked with the responsibility of their performance management. The direct senior of the participant is therefore in the best position to do the coaching and performance evaluation. In this study, the possibility of criterion contamination was counteracted by the systematical development of the BARS and the thorough training of seniors in completing it, in such a way that the possibility of the halo effect could be minimized. The literature also implies that rater errors are minimized through the use of BARS (Gibson, Ivancevich, & Donnelly, 1994). Presumably, these errors are reduced as a result of the independence between the dimensions rated and the relatively high reliability of BARS (Gibson et al., 1994; Spangenberg et al., 1989).

The challenge for future research in this field therefore lies in a trade-off between a better research design to control criterion contamination, and the cost-effective utilization of the assessment center feedback and on-the-job coaching for developmental purposes. It is advisable that in future studies the performance of participants are rated by not only the direct seniors but also by their peers and subordinates, to determine the exact nature of the possible contamination in the senior ratings.

Owing to the importance of assessing current skills correctly and making development recommendations which will be accepted and acted upon

in the work environment, it is recommended that the construct (discriminant and convergent) validity of this developmental assessment center be determined (Reilly, Henry, & Smither, 1990). There is also a need for a confirmatory factor analysis to confirm the existence of the four clusters of dimensions postulated on *a priori* grounds in the developmental assessment center. Furthermore, it is essential to determine the reliabilities and intercorrelations of the dimensions of managerial performance. In this study, the following steps have been taken to overcome some of the problems associated with construct validity of assessment center dimensions: The number of dimensions per simulation was limited, and the observers and administrators were given thorough training and exposure. Behavior checklists can also be used by observers to record whether something occurs or not, instead of making inferences about participants' performance on underlying constructs (Reilly et al., 1990).

The intervention in this study include three components, namely the developmental assessment center with its simulations, the feedback session, and the on-the-job development. This multicomponential intervention raises another research question of which component produced the observed effects, or whether all three components are necessary for a successful intervention.

The feedback session is of primary importance in the developmental assessment center process. Future research should therefore focus on which person or combination of persons as source of the feedback will make it most beneficial. Further areas to be investigated are those of content, timing, and consistency of feedback.

It is also recommended that future studies utilize more sophisticated statistical techniques and research designs (Kerlinger, 1986). A multivariate procedure should be used to take into account the correlations among the dependent measures and to control experimentwise error rate.

Research emphasis should also be placed on post-center activities. After the center and subsequent feedback, a variety of on-the-job activities take place which influence the transfer of skills and ultimately managerial job performance. The influence and nature of these activities (e.g., mentorship, coaching) need to be investigated further.

Finally, given the evolution of the assessment center method, it also becomes evident that research should shift its focus considerably. The developmental applications (developmental assessment and collaborative centers) of the assessment center methodology need to be studied more, especially in the light of large numbers of organizations using assessment center technology for development. Especially the evaluation of the behavioral and utility/value added outcomes are of vital importance to organizations using center methodology.

REFERENCES

- ALLIGER, G. M., & JANAK, E. A. Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 1989, 42, 331-342.
- BAIRD, L. S., SCHNEIDER, C. E., & LAIRD, D. *The training and development sourcebook*. Amhurst: Human Resources Development Press, 1983.
- BALDWIN, T. T., & PADGETT, M. Y. Management development: A review and commentary. In C. L. Cooper and I. T. Robertson (Eds.), *International review of industrial and organizational psychology 1993*. Chichester: John Wiley & Sons, 1993.
- BEARDSLEY, S. A. *Unanticipated outcomes experienced by assessors in a management skills assessment center: An exploratory analysis*. Unpublished PhD dissertation, University of California, Berkeley, 1985.
- BERRY, J. K. Linking management development to business strategies. *Training and Development Journal*, August, 1990, 20-21.
- BOEHM, V. R. Using assessment centers for management development—Five applications. *Journal of Management Development*, 1985, 4(4), 40-51.
- BOYATZIS, R. E. *The competent manager*. New York: Wiley, 1982.
- BURKE, M., & DAY, R. A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology*, 1986, 71, 232-245.
- BUSHNELL, D. S. Input, process, output: A model for evaluating training. *Training and Development Journal*, 1990, 44(3), 41-43.
- CAMP, R. R., BLANCHARD, P. N., & HUSZCZO, G. E. *Towards a more organizationally effective training strategy and practice*. Englewood Cliffs: Prentice Hall, 1986.
- CARNEVALE, A. P., & SCHULTZ, E. R. Return on investment: Accounting for training. *Training and Development Journal*, July 1990, 515-517.
- CASCIO, W. F. *Costing human resources: The financial impact of behavior in organizations*. Boston: PWS-KENT, 1991.
- COHEN, J. *Statistical power analysis for the behavioral science*. New York: Academic Press, 1977.
- DULEWITZ, V. Improving assessment centers. *Personnel Management*, June, 1991, 50-55.
- FISCHER, A. H. *An evaluation of the impact of a developmental assessment center on managerial performance*. Unpublished Master thesis, University of Stellenbosch, 1992.
- FLEENOR, J. W. *The Utility of Assessment Centers for Career Development*. Unpublished PhD dissertation, North Carolina State University, Raleigh, 1988.
- GATEWOOD, R. D., & FEILD, H. S. *Human resources selection* (2nd ed.). Orlando: Dryden Press, 1990.
- GAUGLER, B. B., ROSENTHAL, D. B., THORNTON, G. C., & BENTSON, C. Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 1987, 27(3), 403-511.
- GIBSON, J. L., IVANCEVICH, J. M., & DONNELLY, J. H. *Organizations. Behavior, structure and processes*. Boston, MA: Irwin, 1994.
- GRIFFITHS, P., & ALLEN, B. Assessment centers: Breaking with tradition. *Journal of Management Development*, 1987, 6(1), 18-29.
- HARRISON, R. *Training and development*. London: Institute of Personnel Management, 1989.
- ILGEN, D. R., & MOORE, C. F. Types and Choices of performance feedback. *Journal of Applied Psychology*, 1987, 72(3), 401-406.
- KERLINGER, F. N. *Foundations of behavioral research* (3rd ed.). New York: CBS Publishing, 1986.
- KINGSTROM, P. O., & BASS, A. R. A critical analysis of studies comparing behaviorally anchored rating scales with other rating formats. *Personnel Psychology*, 1981, 34, 263-289.
- KIRKPATRICK, D. L. Techniques for evaluating training programs. *Training and Development Journal*, 1979, 6, 78-92.
- KRIEK, H. J. Die bruikbaarheid van die takseersentrum: 'n Oorsig van resente literatuur (the utility of the assessment center: A review of current literature). *Journal of Industrial Psychology*, 1991, 17(3), 34-37.

- LORENZO, R. V. Effects of assessorship on managers. *Personnel Psychology*, 1984, 37, 617-634.
- MARSH, N., MacCORMICK, A., & ROBINSON, P. Management development and strategic management change. *Journal of Management Development*, 1989, 5(1), 29-30.
- McCLOSKEY, L., & SLIVINSKY, L. W. Maximizing assessment centers. *Manpower Journal*, May, 1983, 24-29.
- NOE, R. A., & SCHMITT, N. The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, 1986, 39, 497-520.
- ORPEN, C., & KING, G. Effects of superior's feedback, credibility and expertise on subordinate's reactions: An experimental study. *Psychological Reports*, 1989, 64, 645-646.
- PHILLIPS, J. J. *The handbook of training evaluation and measurement methods*. London: Gulf, 1991.
- REILLEY, R. R., HENRY, S., & SMITHER, J. W. An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology*, 1990, 43, 71-83.
- ROBERTSON, P. I., & ROUT, U. Assessment based development centers. *Journal of Managerial Psychology*, 1989, 4(3), 11-15.
- SACKETT, P. R. Assessment centers and content validity: Some neglected issues. *Personnel Psychology*, 1987, 40, 13-25.
- SCHMITT, N., FORD, J. K., & STULTS, D. M. Changes in self-perceived ability as a function of performance in an assessment center. *Journal of Occupational Psychology*, 1986, 59, 327-335.
- SHAVELSON, R. J. *Statistical reasoning for the behavior sciences*. Boston: Allyn and Bacon, 1981.
- SLIVINSKI, L. W., McDONALD, U. S., & GOURGEOUS, R. P. Immediate and long-term reactions to an assessment center. *Journal of Assessment Center Technology*, 1979, 2(22), 13-18.
- SMITH, P. C., & KENDALL, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 1963, 47(2), 149-155.
- SPANGENBERG, H. H. *Assessing managerial competence*. Cape Town: Juta, 1990.
- SPANGENBERG, H. H., ESTERHUYZE, J. J., VISSER, J. H., BRIEDENHANN, J. E., & CALITZ, C. J. Validation of an assessment center against BARS: An experience with performance related criteria. *Journal of Industrial Psychology*, 1989, 15(2), 1-10.
- TANNENBAUM, S. I., & YUKL, G. Training and development in work organizations. *Annual Review of Psychology*, 1992, 43, 399-441.
- THORNTON, G. C., & BYHAM, W. C. *Assessment centers and managerial performance*. London: Academic Press, 1982.
- WOODRUFFE, C. *Assessment centers—identifying and developing competence*. London: Institute of Personnel Management, 1990.

BIOGRAPHICAL NOTES

AMOS ENGELBRECHT received his Ph.D from the University of Stellenbosch, South Africa. He is a Senior Lecturer in Industrial Psychology at the University of Stellenbosch. His current research interests include organizational leadership, cultural diversity, managerial motivation, and transformational leadership.

HERMANN FISCHER received his M. Econ in Industrial Psychology from the University of Stellenbosch. He has extensive experience in the application of the developmental assessment center process in organizations. He is currently a consultant in human resources development.