

DO ASSESSMENT CENTRES WORK, AND IF SO WHY?

Theresa Feathers
January 2001

Assessment centres are used by organisations for making selection decisions for appointments to jobs, for promotion, and for decisions about the potential development of individuals over a longer period (Fletcher, 1991). Their use is relatively common both in large UK organisations and those of other industrialised countries (Shakleton & Newell, 1997), particularly for graduate recruitment (Keenan, 1997). If assessment centres work in terms of ensuring that organisations appoint, develop and promote only people that are effective in their jobs, this will bring benefits to both the organisation and to the individuals concerned (Robertson, 1996). For the individual such benefits may include greater job satisfaction, good career prospects, and enhanced earnings. For the organisation, the effectiveness of job-holders will be reflected in significant organisational effectiveness and, hence may bring considerable financial gain. Although the use of assessment centres continues to grow in popularity (Shakleton & Newell, 1991), they have also been dogged with evidence that content and criterion measures are lacking in construct validity (Robertson, 1996). In considering whether or not assessment centres really work as selection processes, the following paragraphs will examine the evidence for and against their use, and will consider this question from the perspectives of both organisations and individuals.

The term "assessment centre" refers to a process of assessment of a group of individuals, which can typically last up to three days. The Task Force on Assessment Centre Guidelines (1989) defined assessment centres as "*a standardised evaluation of behaviour based on multiple inputs*" (Baron & Janman, 1996). There is no uniform format for assessment centre design (Shackleton et al, 1997), but they typically include traditional assessment procedures such as interviews, paper and pencil tests of cognitive ability and personality assessment, group discussions, individual presentations and role-plays; an assessment centre can also include work sample exercises which simulate real job tasks such as the in-tray exercises, leaderless group exercises and business games (Woodruff, 1990). The performance of candidates on these exercises, each of which is designed to measure some of a set of defined competencies, is evaluated by a number of assessors (Baron et al, 1996), and the results of the assessment centre are the pooled evaluations and judgements of the assessors. Competencies may be described as "*psychological attributes that determine people's behaviour in a variety of settings*" (Boam & Sparrow, 1992, in Robertson, 1996). Some competencies are similar to different kinds of cognitive ability such as numeracy, whilst others

are more aligned to characteristics of personality, which may include interpersonal skills, persuasiveness, analytical skills (Robertson, 1996).

In the design of an assessment centre, the first stage is to identify, through job analysis, the competencies that are important to successful job performance. It is these competencies that will inform the choice and design of assessment centre exercises and against which candidate behaviour will be assessed (Woodruff, 1990). Thorough job analysis is an effective method for organisations to focus on what they need in terms of competencies and can therefore be seen as a good tool for organisational development (Fletcher, 1991); however, Torrington & Hall (1991) found that less than half of personnel departments use job analysis and its outcomes for assessment and selection purposes. If the competencies are not well specified through such a process, assessment centre assessors will inevitably interpret the requirements themselves (Robertson, 1996). Indeed Sparrow (1997) argues that assessment centres are fraught with difficulties from the perspective of assessors, particularly in relation to their limited ability to handle complex decisions concerning multiple competencies. This may hold significant implications for fairness in the selection process, in that different assessors may hold different views of the required behaviours to meet the competency requirements (Baron et al, 1996). For example, Schein (1975) illustrated that perceptions of successful managers were very close to perceptions of men in general. Similarly, and more recently, Sparrow and Rigg (1993, in Baron et al, 1996) illustrated how men and women see the role of a manager differently – in a job analysis study in which the structured interview method was used to elicit descriptions of the same managerial job from men and women job holders, women emphasised the need for a participative leadership style and the avoidance of snap decision-making, whilst men said that the role needed forcefulness and quick decision makers. Clearly the definition of assessment centre competencies would be very different depending on which set of information was used and, a leadership competency, for example, could be interpreted differently by different assessors, leading to different scores being awarded for the same behaviours. Nevertheless, assessment centres do not seem to create adverse impact on the basis that as many women pass as men (Ritchie & Moses, 1983). On the other hand, although women's participation in the labour market is increasing, only 30 percent of managers in Britain are women (Alimo-Metcalfe, 1993).

Clearly, the extent to which assessment centres work is dependent not only upon the establishment of well-defined criteria in order to generate objective ratings from assessors (Baron et al, 1996), but also on the judgement and skill of assessors in the process. An individual's biases, either conscious or unconscious, could introduce unfairness into otherwise equitable procedures. Indeed, some critics of assessment centres have argued that they can

be prone to perpetuating the status quo, through assessors selecting individuals whose “faces fit,” and filling the organisation with “carbon-copies” of top management (Klimoski & Strickland, 1977, in Klimoski et al, 1987). On the other hand, it can be argued that selecting carbon-copies of existing management is better than selecting at random, which is all that some methods of selection can achieve. Nevertheless, this suggests that there would be considerable benefit in including external, independent assessors in the process in order to bring a greater degree of objective neutrality to the decision making.

The final scores are dependant upon the integration discussion between assessors; the combining of data through debate and discussion involves assessors reporting and challenging observed evidence before reaching a group consensus of the appropriate final rating (Baron et al, 1996). Again, whilst mechanical procedures may be used for combining scores, the tenure of the discussion could introduce bias even at this late stage; indeed, the fact that many integration discussions occur late in the evening may increase the danger of this kind of error (Baron et al, 1996). Wieby, Atlink, Castelijns & Castelijns (1997) emphasise the importance of assessor training as a means of communicating both the criteria and the performance goals. Training sessions can provide opportunities for managers, from potentially diverse organisational cultures in circumstances where they may be drawn from different divisions, to discuss these issues, to identify common ground, and to agree the desired outcomes. Wieby et al (1997) also suggest that this process will lead to managers feeling a greater commitment to the new personnel and to their future development.

Despite the potential for error through ill-defined competencies and assessor bias, assessment centre procedures nevertheless tend to achieve good predictive validity, which may account for their growing popularity and provide justification for their cost (Woodruff, 1990). Validity can be looked at either by examining the validity of the component parts of the assessment centre, or by assessing the validity of the overall scores derived (Robertson, 1996). Most studies, it would seem, have concentrated on the latter. Gaugler, Rosenthal, Thornton & Bentson (1987) carried out a meta-analysis to determine the validity of assessment centres, part of which analysed the correlations between overall assessment centre ratings and subsequent measures of job performance. They found that validity was not consistent across studies, which is hardly surprising given that there is no prescriptive formula for the content of assessment centres and that they cannot be replicated as precisely as, say, an ability test or a structured interview (Woodruff, 1990); but, in general terms, they found corrected mean validities ranges from .30 to .53. In the case of assessment centres used for promotional decisions, Hunter & Hunter (1984) reported a correlation of .43. Sackett & Dreher (1982) analysed ratings on a number of scales across a variety of assessment centre

exercises. They found that ratings of different dimensions within the same exercise correlated very highly, but that ratings of the same dimension in different exercises hardly correlated at all. When the ratings were factor-analysed, the factors clearly identified consistency within exercises, rather than traits. In other words, the assessment centres were measuring general performance on each task rather than competencies across a range of tasks. As such, if a dimension in one task does not generalise well to the same dimension in another task, it is difficult to be sure that it will generalise to performance on the job. Turnage and Munchinsky (1982, in Klimoski & Brickner, 1987) analysed ratings of over 2000 people on eight traits across five exercises and also found that differentiation of ratings were centred on exercises rather than dimension, confirming Sackett et al's (1982) claims. Clearly then, evidence of the extent to which assessment centres produce accurate assessments of the dimensions on which they are built seem problematic.

Whilst findings support the criterion and content validity of assessment centres, the question remains as to what constructs are actually being assessed (Saldago, 1999). Where several different criteria may be important to a particular role it is not always clear which ones should receive greater importance (Robertson, 1996); it would seem that, in exercises designed to assess multiple dimensions, managers will often attach greater importance to a small number of dimension, and frequently those with lower validities, in reaching their overall assessment ratings (Sparrow, 1997). On the basis that the correlation between overall assessment ratings and the collective dimension ratings is often low there may be benefit in designing exercises that are specific to particular dimensions and minimising the number of dimensions overall to those critical to the role (Robertson, 1996). In addition, Kleinmann (1993) found that if dimensions were transparent to participants, they would perform better than in conditions lacking in transparency; he also found that transparency enhanced the construct validity. Whilst the assessment centre may potentially provide a powerful learning experience for participants about their individual competencies (Fletcher, 1991), feedback to participants on dimension that are not rated reliably across exercises cannot easily be justified (Borman, Hanson & Hedge, 1987). And yet, assessment centres hold a high degree of face validity (Iles & Robertson, 1989); they look fair and plausible and include samples of jobs in which people have a chance to prove themselves. And whilst many candidates may be able to bluff their way through other forms of selection procedures such as an interview, acting the part of a good manager for two or three days is more challenging (Woodruff, 1990). However, the ability to perform exercises to the required standard is not the only component of potential job success (Robertson, 1996). The tradition of designing exercises to simulate aspects of the job assumes that candidates already hold the necessary experience to interpret and respond to the situation they are faced with; but, as Keenan (1997) points out, this is precisely what

graduates lack. Whilst such exercises may work well for those that have had relevant exposure to such situations in the workplace, for candidates who have no familiarity of this kind, and yet who may have the necessary potential, such exercises could potentially be demotivating (Keenan, 1997).

There is a substantial body of evidence attesting to the basic validity of assessment centres and, on the face of it therefore, one would have to conclude that the assessment centre is one of the more robust methods of assessment. In addition, group exercises save time and create group dynamics which allow aspects of the individual to be studied that cannot easily be measured by other means (Woodruff, 1990). However, assessment centres are expensive and the difference in value between well designed procedures and poor ones can make the high costs well worthwhile, particularly as the evidence illustrates that well designed assessment centres generally achieve higher levels of validity overall. The fact is though, we don't know why assessment centres work so well (Woodruff, 1990). If further research were to pinpoint precisely which elements of the assessment centre contributed to its success, then their predictive accuracy could be improved still further. In the meantime, the evidence to date would suggest the need for careful job analysis drawn from multiple sources to inform the nature of the key competencies required, careful design of exercises to elicit the behaviours against which individuals' competencies will be evaluated, and thorough training of assessors, in order to maximise success. Where such meticulous procedures are engaged then it can be concluded that assessment centres will work well, and their cost will be justified in terms of ensuring that individuals are well matched to organisations and jobs.

References

- Alimo-Metcalfe, B. (1993). Women in Management: Organisational socialisation and assessment practices that prevent career advancement. *International Journal of Selection and Assessment*, **1**, 68 – 83.
- Baron, H. & Janman, K. (1996). Fairness in the Assessment Centre. In Cooper, C.I. & Robertson, I.T (1996). *International Review of Industrial & Organisational Psychology*. Vol. 11. Chichester: Wiley
- Borman, W.C., Hanson, M.A., & Hedge, J.W. (1997). Personnel Selection. *Annual Review of Psychology*. **48**, 299 – 337.
- Fletcher, C. (1991). Candidates' Reactions to Assessment Centres and their Outcomes: A Longitudinal Study. *Journal of Occupational Psychology*. **64**, 117 – 127.
- Gaugler, B., Rosenthal, D., Thornton, G.C. & Bentson C., (1987). Meta-analysis of Assessment Centre Validity. *Journal of Applied Psychology*. **72**, 493 – 511.

- Hunter, J.E. & Hunter, R.F. (1984). Validity & Utility of Alternative Predictors of Job Performance. *Psychological Bulletin*. **96**(1), 72 – 98.
- Iles, P.A. & Robertson, I.T. (1997) The Impact of Personnel Selection Procedures on Candidates. . In Anderson N. & Herriot, P. (Eds), (1997). International Handbook of Assessment & Selection. Chichester: Wiley.
- Keenan, T. (1997). Selection for Potential: The case of Graduate Recruitment. . In Anderson N. & Herriot, P. (Eds), (1997). International Handbook of Assessment & Selection. Chichester: Wiley.
- Kleinmann, M. (1993). Are rating dimensions in assessment centres transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*. **78**, 988 – 993.
- Klimoski, R. & Brickner, M. (1987). Why do Assessment Centres work? The puzzle of assessment centre validity. *Personnel Psychology*, **40**, 243 – 260.
- Ritchie, R.J. & Moses, J.L. (1983). Assessment Centre correlates of women's advancement into middle management: a 7-year longitudinal analysis. *Journal of Applied Psychology*. **68**, 227 – 231.
- Robertson, I. (1996). Personnel Selection & Assessment. In Warr, P. (1996). (Ed). Psychology at Work. 4th Edn. London: Penguin.
- Sackett, P.R. & Dreher, G.G. (1982). Constructs and assessment centre dimensions: some troubling empirical findings. *Journal of Applied Psychology* **67**, 401 – 410.
- Schein, V. (1975). The Relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*. **57**, 95 – 100.
- Shackleton V. & Newell, S. (1991). Management Selection: A comparative Survey of methods used in top British and French Companies. *Journal of Occupational Psychology*. **64**, 23 – 36.
- Shackleton V. & Newell, S. (1997). International Assessment & Selection. In Anderson N. & Herriot, P. (Eds), (1997). International Handbook of Assessment & Selection. Chichester: Wiley.
- Sparrow, P. (1997). Organisational Competencies: Creating a Strategic Behavioural Framework for Selection and Assessment. . In Anderson N. & Herriot, P. (Eds), (1997). International Handbook of Assessment & Selection. Chichester: Wiley.
- Torrington, D. & Hall, L. (1991). Personnel Management – A New Approach (2nd Edn). Hemel Hempstead: Prentice Hall.
- Wieby, M.M., Atlink, C.F., Castelijns, V. & Castelijns, M. (1997). Criterion Development: The Unknown Power of Criteria as Communication Tools. . In Anderson N. & Herriot, P. (Eds), (1997). International Handbook of Assessment & Selection. Chichester: Wiley.
- Woodruff, C. (1990). Assessment Centres, IPM.