

Assessment Centres for Screening and Development of Skill: A Critical Analysis

Presented at the 1st International Conference on Applied Psychology (2016), Gurgaon, India

Abstract

Assessment centres have been used globally for more than 50 years now. The purposes of using assessment centres ranges for external selection, internal promotion, rewarding exemplary work to leadership identification. The scope of assessment centres has grown multi-fold with the growth of business and diversity in industry in the global market. To identify the right talent has always been an area of concern for all sectors of the industry and is a matter of high priority for HR professionals globally. With the increasing popularity of assessment centres, it is imperative that the process is able to yield effective results. The current paper intends to critically analyse the process, virtues, and concerns of an assessment centre. The paper intends to bring into focus facets of assessment centres such as its relationship with competencies specific to roles and culture of organisations; generalizability and specificity of results; relationship with effectiveness at work; and validity of tools. Propositions have also been offered to improve the success factor of assessment centres.

Key Words: Assessment centres, competencies, concerns, effectiveness, propositions

Introduction

We know it well that none of us acting alone can achieve success' (Mandela, 1994). For anything and everything we do it is very essential to have political, social and economic support, human beings cannot thrive in isolation. The same is true with respect to Assessment Centres (AC).

AC primarily tries to assess an individual's communication skills and team building ability. It has been observed that in the modern day globalized organizations these two skills are very vital and indispensable. They are very significant for growth of the individual as well as the organization. The success or failure of an organization to quite an extent depends on these two basic skills.

ACs were first used to select the right recruits to be OSS (Office of Strategic Services; later to become the CIA) agents for the United States (Thornton & Rupp, 2006). Using this method helped the agency to select new agents for its ranks, and now it has grown beyond the OSS/CIA and into the ranks of private and public business.

What is an Assessment centre?

“Assessment centers” are not brick and mortar research centers or buildings. They are rather an abstract concept that exists in practice and refer to standardized procedures used for assessing *behavior-based or performance-based* dimensions whereby participants are assessed using multiple exercises and/or simulations (Thornton, 1992). Common simulation exercises used in an assessment center include oral presentation, leaderless group discussions, role playing, in basket, oral fact-finding, business game, and integrated simulations (Thornton, & Mueller-Hanson, 2004). Dimensions for assessment (equated to competencies) are usually identified through job analysis. It should be noted that though job analysis is often used as interchangeably with competency modeling, the two differ in terms of assessment of reliability, strategic focus, and expected outcome (Shippmann, et al., 2000)

The assessment center method (ACM) has been used for many purposes in human resource management including selection, diagnosis, and development since its introduction over 50 years ago (Thornton & Rupp, 2006).

Industrial Psychology personnel and Human resource managers have been using this term since more than 50 years. They have used assessment centers as a combination of unique essential elements codified in the Guidelines and Ethical Considerations of Assessment Center Operations (henceforth Guidelines; International Task Force, 2008).

In an AC there are multiple assessors trained in various dimensions who observe and assess the behavior of the candidates in multifarious situations and then rate the performance of the participants on areas that are deemed to be significant for effective performance in specific positions.

In order to identify the tasks and competencies of a particular job or job group, job analysis is done (the analysis of tasks and responsibilities in a particular job or job group). Many organizations also go for competency modeling in order to accomplish the premeditated goals of the organization.

ACs have the ability to assess all observable behavior. Some of these are managerial skills, interpersonal skills, leadership, team dynamics, decision making, critical thinking, sales abilities and many more. All the tasks that an individual performs and the responsibilities that entail as part of the job too come under the purview of the ACs.

The assessors need not necessarily be from an outsourced agency, they can be any employee who is in a target position, human resource managers or psychologists too. Quite often a combination of internal and external members forms the AC team.

Irrespective of who is doing the assessment they have to follow a systematic pattern of recording the behavioral measures. Not only this the team also needs to have their own independent observations and evaluations and record it systematically.

ACs use the 360degree assessment pattern. Therefore a thorough check into the background of the person (case history), information and interview from peer group, supervisors, and subordinates, scores on performance tests, personality tests and cognitive ability have to be taken into consideration.

Once all the data has been gathered the ACs need to collate all the data either statistically (if applicable) or through the classic process of sharing and discussing the observations and notes with the relevant people in order to achieve some definite conclusion. The qualitative and quantitative data is then collectively analyzed to have an overall personality assessment.

The basic idea of collating information from all the relevant sources is in order to have a holistic assessment of an individual so that he is successful in whatever he /she is assigned to do. The international task force for AC guidelines presents ten features that must be present for a process to be called an "Assessment centre (Joiner, 2000)." These ten features are as follows: (1) Job Analysis, (2) Behavioural Classification, (3) Assessment Techniques, (4) Multiple Assessments, (5) Simulations, (6) Assessors, (7) Assessor Training, (8) Recording Behaviour, (9) Reports, and (10) Data Integration. While all of these features are necessary for an AC to bear that moniker, one should not simply checklist these items and move on. An AC is a very expensive engagement and poor planning can be responsible for a lack of return-on-investment (Howard, 1997). Additionally, even though a number of the best practices that will be presented here have been the focus of research, which presents promising evidence of efficacy, many companies have failed to implement them (Spychalski, Quinones, Gaugler, & Pohley, 1997).

The essential features of an AC are therefore: (International Task Force, 2009) include:

- multiple assessors including line managers, HR staff, industrial/organizational psychologists, or other persons trained in behavioral observation procedures,
- multiple simulation exercises with high fidelity to the target job(s) including such activities as group discussion, presentation, written case analysis, and interviews with a subordinate, supervisor, or client,
- a systematic process for observing and recording behaviors, classifying behaviors into relevant performance-related dimensions or other categories, and
- Integration of behavioural observations across assessors and exercises via consensus discussion or arithmetic combination.

The primary uses of assessment centers are in the areas of selection. Selection herein is a very broad term with multiple manifestations. Selection refers to the ratings that assist the AC in the following areas.

- For Recruitment of candidates
- For promotion of internal candidates into supervisory and managerial ranks,
- For classifying individuals into a pool of high potentials who will get special training,
- For identification of exemplary staff members who are likely to receive certification of competence in job skills,
- For identifying employees for retention when there is a need for reduction in force and reorganization.

Assessment centres used for selection

External selection

Assessment centers are regularly used for assortment of applicants at multiple levels. Despite the fact that assessment centres are at times quite costly and there are a lot of arguments and counter arguments about the economic viability of the ACs at the entry level. Many people and organizations have been raising questions about the economic feasibility of the ACs at the entry level.

Coulton and Feild (1995) argued that the potential benefits exceed the costs of an AC for police selection. In fact, the Israeli police force has used a variety of performance exercises simulating challenging physical and leadership situations to screen candidates and found the 2-day behavioral assessments added unique predictive accuracy over cognitive ability tests (Dayan, Kasten, & Fox, 2002).

The police department of Fort Collins, Colorado used an assessment center as one step in screening candidates for patrol officer (Gavin & Hamilton, 1975). The exercises simulated interpersonal situations officers may encounter on the job, such as domestic disputes and disorderly persons on the street. In the United Kingdom, assessment centers have been used by many organizations to select management trainees. With the rising expectations to have manufacturing employees contribute to teamwork and continuous quality improvement, organizations have turned to the assessment centre method to assess interpersonal and decision making skills. One of the areas of great expansion of applications in the 1980s and 1990s came in large organizations such as Diamond Star Motors (Henry, 1988), Cessna (Hiatt, 2000), Coors (Thornton, 1993), and BASF (Howard & McNelly, 2000) and even small organizations such as Wilkerson Manufacturing (Thornton, personal communication, June 15, 2008). Assessment centers have also been used to select pilots (Damitz, Manzey, Kleinmann, & Severin, 2003), automotive manufacturing trainers (Franks, Ferguson, Rolls, & Henderson, 1999), and salesmen (Bray & Campbell, 1968).

Certification of competence in many organizations, especially teaching and the IT industry, formal certification of individual employees carries considerable weight. Thus, Connecticut has used assessment centers to certify teachers (Jacobson, 2000) and Sun Microsystems developed behavioral assessments to certify the competence of consultants providing services in the design of selection and training programs for clients and of customer service representatives providing telephone support (Howard & Metzger, 2002; Rupp & Thornton, 2003). Sackett (1998) reported the use of related methods with lawyers.

Promotion of internal candidates

Promotion of candidates is one area where the inputs of the AC have been very religiously considered. Many early practical applications of the method in large organizations such as AT&T, Standard Oil, Sears, and IBM were designed to screen manufacturing or sales persons into first level management (Thornton & Byham, 1982). Because of the multidimensional nature of assessment that AC does the organizations get an opportunity to observe candidates in various situations gives an insight into the competencies and behavioral quality of the candidates which have been hitherto undiscovered. Mostly when people are working on the assigned job they hardly get an opportunity to display the innate potential.

The ACs have also been used to aid promotion decisions into middle and upper management levels. In fact, the purpose of the original AT&T Management Progress Study involving men (Bray & Grant, 1966) and Management Continuity Study involving women (Howard & Bray, 1988) was to identify persons with potential for success in a wide array of middle management positions, some of which may not be known at the time of assessment. These studies provided evidence that Overall Assessment Ratings (OARs) predicted management progress and performance over several years. At the top executive levels, Howard (1997) reported that the Assessment Centre Method has been used to assess both internal and external candidates. With the increasing pressure to evaluate top leaders, the ACM may be used even more frequently (Howard, 2001).

Identification of high potentials

People are multitalented and multifaceted. It is for the organization and the bosses to realize the potential of the employees and make the best out of them. Attrition is one fear that every organization faces. With attrition it is not only the loss of the employee but starting all over again, training a new employee which involves spending a considerable amount of money and time resulting in stunted growth at times.

Besides attrition, AC method also assists in succession planning. Because of the systematic and multidimensional assessment it unearths hidden talent of the employees. Keen and observant managers can hope to identify the potential talent and groom them according to their needs and the demands of the dynamic organization.

The assessment center method has become a core feature of many succession planning programs involving systematic procedures to identify staff members, early in their careers, with high potential for long range leadership success. These “hi pos” are then moved through a set of training, mentoring, and experiential assignments (Byham, Smith, & Paese, 2000).

Process of ACs

During development of an AC that has the purpose of attempting to predict leadership, a 6-step recursive process has been used that has effectively predicted leadership (Brownell, 2005). This process (see Figure 1) which takes the development team through all of the necessary stages of AC construction and implementation and then advises that information from use of the AC be used to evaluate its effectiveness and retool the AC for its next implementation.

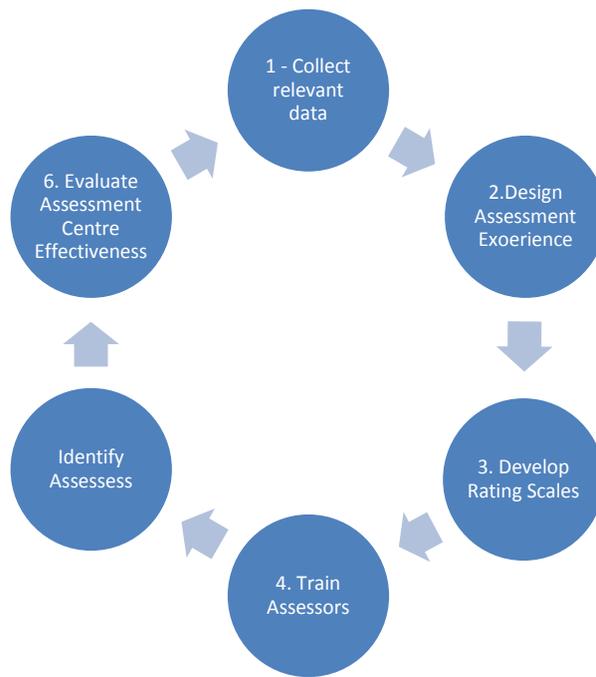


Figure 1. Assessment centre development process (adapted from Brownell, 2005)

Benefits of Assessment Centres

Organizations have used ACs for recruitment, selection, placement, performance appraisal, organization development, HR planning, promotions, and even layoffs (Thornton, et al., 2006). In a meta-analysis of AC usage by organizations Spsychalski, et al. (1997) found that the three most popular reasons for implementing an AC were selection, promotion, and development planning.

Selection and promotion are usually grouped together for the purpose of examining best practices because this gives an opportunity to assessees to look into the various dimensions of human behaviour. They do not only get a glimpse into the sample rather the behavioural processes of the individuals which is very vital for the growth and development of an organization. Despite concerns with the construct validity, there has been substantial research which has found that ACs following the guidelines have good predictive validity (.37 to .41) in comparison to other selections devices (Arthur, et al., 2003; Howard, 1997).

Also, ACs have been found to be linked to job retention in both male and female applicants (Anderson & Thacker, 1985). In a study by Macan, Avedon, Paese and Smith (1994), it was found that applicant tend to find that ACs are more face valid than other assessment methods and applicants believe that ACs give them a fair and better opportunity to display their skills than other assessment forms.

Criticism of Assessment Centres

There have been several studies which have been able to identify the errors in conducting assessment centres. The errors suggested by Caldwell et al. (2003) include poor planning, inadequate job analysis, weakly defined dimensions, poor exercises, lack of pretest evaluation, unqualified assessors, inadequate assessor training, inadequate candidate preparation, sloppy behavior documentation and scoring, and misuse of results.

Lievens and Thornton III (2005) have also explained two disturbing trends in the implementation of assessment centres. They have suggested that in many cases, modifications of essential steps in the development and implementation have led to several short cuts. This has greatly impacted the accuracy and effectiveness of the program. Secondly, they also specified that assessment centre is being used to refer to many other methods which do not comply with the essential elements and benchmarks set for the assessment centre method to be effective. While there are no patent registrations and copyright claims they suggest there are strong reasons to restrict the use of the term assessment centre.

Validity of Assessment centres.

An indicator of the similarity between the subject matter of a measure and the domain that it purportedly represents is indicative of the content validity. While, construct-related validity is an indicator of the relationship between a measure and the theoretical concept it is intended to measure. An indicator of the relationship between a measure and a relevant behavioural indicator of performance represents the criterion-related validity.

Content validity

Thornton and Mueller-Hanson (2004) have suggested that assessment centers demonstrate substantial content-related validity in that they are high-fidelity simulations of managerial work that are typically based on a job analysis that generated critical work-related performance dimensions (Thornton, 1992).

Criterion related validity

A test has criterion-related validity when the test measures relate significantly to a criterion, for example, to job performance. In the AC domain, various studies have demonstrated evidence for AC criterion-related validity. Meta-analyses found correlations between the overall AC rating and job performance ranging from .26 to .40, indicating that the overall AC rating is predictive of job performance (Arthur et al., 2003; Becker, Höft, Holzenkamp, & Spinath, 2011). In addition, meta-analytic findings suggest that ratings of specific dimensions are also criterion valid (Arthur et al., 2003; Meriac, Hoffman, Woehr, & Fleisher, 2008).

Moreover, ACs contribute to the prediction of job performance beyond cognitive ability tests or personality inventories (Dilchert & Ones, 2009; Melchers & Annen, 2010; Meriac et al., 2008).

Construct validity

One of the biggest strengths of the AC method is its predictive validity. While there are concerns with construct validity many pieces of research have found that ACs utilizing The Guidelines have good predictive validity (.37 to .41) in comparison to other selections devices (Arthur, et al., 2003). Assessment centres seem to present good criterion-related validity, however the construct validity has been found to be poor (Howard, 1997; Woehr & Arthur, 2003). Some researchers have also conducted studies and found evidence that seems to point to clear results of weak construct validity (Joyce, Thayer, & Pond, 1994) For instance, some researchers have examined exercises and dimensions side-by-side demonstrating that the majority of variance in ratings is attributable to exercises, not dimensions (Lance, Lambert, Gewin, Lievens, & Conway, 2004).

Sackett and Dreher (1982) studied three organisations and in each of these organizations, they found low correlations among ratings of a single dimension across exercises (i.e., weak convergent validity) and high correlations among ratings of various dimensions within one exercise (i.e., weak discriminant validity).

Lievens and Conway (2001) have studied three perspectives which seem relevant while describing how design characteristics can increase dimension variance. They suggest that the limited cognitive capacity is an important factor that could lead to dimension variance of assessment centres. Assessors possess limited information-processing capacities and meeting the cognitive demands of the AC becomes challenging. Cognitive overload could be a result of the many inferential leaps that assessors make in order to provide essential ratings for dimensions. Differences between expert and novices also lead to dimension variance as result of using different schemas when rating based on their experiences. The study also reflected upon the interactionist perspective based on the premise that the candidates' performance on same dimension but in different exercises would vary resulting into a lack of cross-functional consistency.

Some others have also found by using generalizability theory; that is, dimensions, not exercises, are responsible for the majority of variance in ACs (Arthur, Woehr, & Maldegen, 2000). Even

others have found evidence that both dimensions and exercises have significant influence on AC ratings (Bowler & Woehr, 2006).

Cahoon et al. (2012) in their study have reflected that ACs have eventually displayed both content and criterion related validity while failing to demonstrate construct related validity.

Some other studies have been able to show evidence of construct validity using Confirmatory factor analysis and Multi Trait Multi Method designs (Kolk, Born, & van der Flier, 2004). While some others have attempted using the same methods and found that AC ratings do not display construct validity (Sackett & Dreher, 1982) However, it has also been argued that possibly that techniques used to measure the construct validity (MTMM) of ACs are being improperly applied leading to the inability to compute the construct validity (Kleinmann & Koller, 1997).

It has also been noted that convergent validity of ACs have been found problematic due to low correlations between ratings of the same dimension from different exercises. However, a lack of discriminant validity has also been found as ratings of different dimension from the same exercise tend to correlate substantially (Melchers et al., 2007; Woehr & Arthur, 2000). Also, the CFA of ACs usually yield exercise factors in the latent factor structure of AC dimension ratings accounting for a substantial proportion of variance in ratings. While, dimension factors if identified seem to be a less significant source of variance in AC dimension ratings leading to the result that ACs do not have a significant internal construct validity (Bowler & Woehr, 2006; Lance, Lambert, Gewin, Lievens, & Conway, 2004; Lievens, Dilchert, & Ones, 2009).

There are some others who have also questioned why it is necessary to worry about construct-validity as long as ACs can demonstrate predictive and content-validity (Norton, 1977). This has eventually led to a huge debate, displaying the need for more empirical evidence (Dreher & Sackett, 1981). And, even to this day, the debate about what to do continues as a recent issue of *Industrial and Organizational Psychology: Perspectives on Science and Practice* (SIOP's official publication) highlights the current feelings of the community (Connelly, Ones, Ramesh, & Goff, 2008).

Explanations for the Findings on Internal Construct-Related Validity of ACs

Several researchers have conducted studies and attempted to find the relationship between internal construct related validity of ACs with the biases on the side of assessors. Rater bias has been referred to as rating inaccuracies because of difficulties during the AC process (Zedeck, 1986). It has been studied that assessors provide inaccurate and unreliable ratings which leads to low construct validity of ACs. According to the Limited Cognitive Capacity model ((Lievens & Klimoski, 2001), the assessors are unable to meet the high cognitive demands of their task due to limited information processing capacities (Bycio, Alvares, & Hahn, 1987). AC dimension ratings are of impaired reliability and accuracy which leads to poor internal construct validity.

In line with the limited cognitive capacity model, some studies have provided support for the idea that AC design interventions that are assumed to reduce cognitive demands placed on assessors improve internal construct-related validity. Assessors who lack expertise do not possess well – established cognitive structures. They tend to provide less reliable and less accurate ratings and also insignificant construct valid ratings than expert assessors as suggested by the Expert model (Lievens & Klimoski, 2001).

Lower exercise variance in AC dimension ratings which reflects situational specificity of candidates' behaviour has been studied to attribute to insignificant internal construct-related validity of ACs (Lance et al., 2004).

ACs have been studied to comprise of different exercises which are intended to represent the variability of contextual demands of the target position (Neidig & Neidig, 1984). Different exercises require different behaviours and candidate's behaviour tends to be different in different exercises (Howard, 2008; Lievens & Conway, 2001; Neidig & Neidig, 1984). It has also been studied that cross-situationally inconsistent candidate's performance in different exercises has been found to lead to convergence between ratings of a specific dimension across exercises and resulting in substantial exercise effects.

It has been demonstrated that using exercises that pose similar demands on candidates' behaviour would lead to convergence between ratings leading to establishing of substantial dimension variance (Schneider & Schmitt, 1992).

Sackett and Dreher (1982) have also presented the following three explanations for poor internal construct validity: (a) poor assessment centre design might lead to assessor biases, (b) exercise variance confounds exercise and assessor variance, and (c) assessees might behave cross-situationally inconsistently across the exercises. Poorly designed assessment centres (inadequate training of assessors, asking assessors to rate a large number of dimensions) might result in assessors being prone to halo bias when rating the candidates. This in turn leads to strong exercise factors. The main problem has been studied to be associated with the assessors' limited information processing capacity. The second perspective also intended to signify that assessors might use incorrect schemas for categorizing the information observed, which might lead to assessor biases. Exercise variance has basically been studied to be a confounding of exercise and assessor variance. This is primarily due to the common practice of assessors rotating through the various exercises. However, taking care of this explanation would not be a viable condition as to use a large number of assessors would be very costly. Also, it was studied that as exercises place different psychological demands on the assessees leading to inconsistent performance and different behaviour (in a one-to-one role-play as compared to in a group discussion). Whereas the first two explanations put the "blame" on assessment centre design or lack of inter-rater reliability among assessors, this third explanation focuses on candidate performances. Assessors are viewed as relatively accurate. Theoretical, methodological, and practical implications have been suggested.

Theoretical Implications

After understanding that candidates' performances was one of the main reasons behind the poor construct-validity of ACs, it was imperative to understand the reasons behind such inconsistency. This has been studied in reference to the person- situation debate in personality and social psychology (Highhouse & Harris, 1993; Lievens, 2002b).

Lievens (2009) studied that the trait activation theory could provide a more comprehensive theoretical explanation for the variability in candidate performances across different exercises in the assessment centre. Trait activation theory is an interactionist theory to explain behaviour based on responses to trait-relevant cues found in situations (Tett & Guterman, 2000). The trait activation theory explains that situational similarity is described through the trait activation potential (the capacity to observe differences in trait-relevant behaviour within a given situation). A situation is considered to be relevant to a trait if it provides cues for the expression of trait-relevant behaviour.

Mischel (1973) has also studied that situational strength also impacts the variability and consistency of behaviour. While, strong situations involve unambiguous behavioural demands leading to less differences in reactions to the situation, weak situations are characterised by ambiguous expectations leading to more variability in behavioural responses. Therefore, it has been suggested that exercises should not be projecting too strong situations with clearly defined tasks. This would eventually lead to few options left open for the assessees. It has been suggested that to ensure enough scope for observation of differences in how assessees tackle the situation there should be certain amount of ambiguity in situations.

Methodological Implications

Most studies have recognised the MTMM approach and CFA to examine construct related validity of assessment centres. However, the CFA model has been studied to have serious estimation problems such as factor correlations greater than one (Bagozzi & Yi, 1991). However, recent research (Conway, Lievens, Scullen, & Lance, 2004) has suggested that studies with large samples would help in overcoming challenges with the MTMM matrix.

Practical Implications

Lievens (2009) have suggested that Frame-of-reference training (Bernardin & Buckley, 1981) which aims to impose a shared performance theory on assessors, providing them with common standards as a reference for evaluating assessee performance has an impact on ACs. Research demonstrated that this training approach, in comparison to other assessor training formats, resulted in higher discriminant validity, higher interrater reliability, and higher rating accuracy (Lievens, 2001a).

A study by Lievens (2009) suggests that once job analysis has identified the dimensions to be measured, trait activation theory might be used to eliminate or combine dimensions within an exercise that seem to capture the same underlying trait (e.g., “innovation” and “adaptability” are based on behaviours that might all be expressions of Openness). With reference to this, trait activation might be helpful in dimension selection and exercise design.

Propositions

It is suggested that trait activation theory should be used proactively. It is imperative that assessments centres are supported with a stronger theoretical background. Using the trait activation in a more prescriptive way, the selection of dimensions, the design of exercises, and the development of feedback reports are some the components that should be modified.

However, it is important that construct-related and criterion related validity should be examined (Woehr & Arthur, 2003). This would enable to ascertain whether specific assessment centre design interventions positively affect both construct-related and criterion-related validity.

It has also been suggested that to improve internal construct-related validity, assessors have to simultaneously observe a lower compared to a higher number of dimensions during the exercises (Gaugler & Thornton, 1989) or when they observe a lower compared to a higher number of candidates in group exercises (Melchers et al., 2010).

Also, it has been suggested that using behavioural checklists, also lead to improvements in internal construct-related validity of ACs (Reilly et al., 1990). Lievens (2001) through study suggest that assessors training and assessor background improve rating accuracy and construct related validity of ACs. Also, a meta-analytic finding by Woehr and Arthur (2003) suggest that psychologists provide more construct valid ratings than managers. It has been suggested that the assessment centre be designed in such a way that competency model to be addressed be clearly defined in terms of sub-competencies and behavioural indicators specific to the job role. The current research also intends to promote the value of confirmatory factor analysis to ensure that the assessment centre designed has significant construct validity. It has also been recommended that the trait activation theory be considered while designing and developing exercises for the assessment centre. This would lead to the capacity to observe differences in trait-relevant behaviour within a given situation.

Based on the findings on the reasons behind the insignificant construct-validity of assessment centres, it is also proposed that assessment of limited number of competencies by each assessor would lead to higher accuracy of ratings. This would eventually lead to improving the construct validity.

Also, it has been suggested in the current research that the assessment centre comprises of psychologists. This would ensure assessment of explicit and implicit behavioural traits. The paper also recommends using the Frame of reference training for all assessors to improve the validity of assessment centre. This would lead to a model with shared performance theory to the assessors, providing them with common standards as a reference for evaluating assessee performance has an impact on ACs.

Future Scope of Research and Limitations

Several empirical research has been able to recognise only two sources of systematics variance, namely dimensions and exercises. More research needs to be conducted to focus on decomposing variance according to three main sources of variance in AC ratings: dimensions, exercises, and assessors. This would be possible only when a fully crossed design is available. This means that multiple assessors evaluate each assessee in each exercise.

It is also recommended that studies be conducted to understand the validity index of ACs in the Indian industries. This would require large number of organisations participating in the study leading to a significant cost and amount of time associated with it. Also, one of the other major limitations of establishing the validity of each AC would be the customised nature of each assessment centre.

References:

- Anderson, L. R., & Thacker, J. (1985). Self-monitoring and sex as related to assessment centre ratings and job performance. *Basic and Applied Social Psychology*, 6, 345–361.
- Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the Criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125-154.
- Arthur, W. J., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management*, 26, 813-835. [http://dx.doi.org/10.1016/S0149-2063\(00\)00057-X](http://dx.doi.org/10.1016/S0149-2063(00)00057-X)
- Bagozzi, R. P., & Yi, Y. J. (1991). Multitrait–multimethod matrices in consumer research. *Journal of Consumer Research*, 17, 426–439.
- Becker, N., Höft, S., Holzenkamp, M. & Spinath, F.M. (2011) The Predictive Validity of Assessment Centers in German-Speaking Regions: A Meta-Analysis. *Journal of Personnel Psychology*, 10(2), 61–69.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review*, 6, 205–212.
- Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, 91, 1114-1124. <http://dx.doi.org/10.1037/0021-9010.91.5.1114>
- Bray, D. W., & Campbell, R. J. (1968). Selection of salesmen by means of an assessment center. *Journal of Applied Psychology*, 52, 36–41.
- Bray, D.W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, 80(17), 1–27.
- Brownell, J. (2005). Predicting leadership: The assessment center's extended role. *International Journal of Contemporary Hospitality Management*, 17(1), 7.
- Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463-474. [doi:10.1037/0021-9010.72.3.463](https://doi.org/10.1037/0021-9010.72.3.463)
- Byham, W. C., Smith, A. B., & Paese, M. J. (2000). *Grow your own leaders. Acceleration pools: A new method of succession management*. Pittsburgh, PA: DDI Press.

- Cahoon, M. V., Bowler, M. C., & Bowler, J. L. (2012). A reevaluation of assessment center construct-related validity. *International Journal of Business and Management*, 7, 3-19. doi: 10.5539/ijbm.v7n9p3
- Caldwell, C. Thornton, G. C. & Gruys, M. L. (2003). Ten classic assessment center errors: Challenges to selection validity. *Public Personnel Management*, 32, 73-88.
- Connelly, B. S., Ones, D. S., Ramesh, A., & Goff, M. (2008). A pragmatic view of assessment center exercises and dimensions. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1(1), 121-124.
- Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling*, 11, 535–559.
- Coulton, G. F., & Feild, H. S. (1995). Using assessment centers in selecting entry-level police officers: Extravagance or justified expense? *Public Personnel Management*, 24, 223–254.
- International Task Force on Assessment Center Guidelines (2008). Guidelines and ethical considerations for assessment center operations. Available at www.assessmentcenters.org
- Damitz, M., Manzey, D., Kleinmann, M., & Severin, K. (2003). Assessment center for pilot selection: Construct and criterion validity and the impact of assessor type. *Applied Psychology: An International Review*, 52, 193–212.
- Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center: Efficient tool or a hammer to kill a fly? *Personnel Psychology*, 55, 827–849.
- Dilchert, S., & Ones, D. S. (2009). Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment*, 17, 254-270.
- Dreher, G. F., & Sackett, P. R. (1981). Some problems with applying content validity evidence to assessment center procedures. *Academy of Management Review*, 6(4), 551-560.
- Franks, D., Ferguson, E., Rolls, S., & Henderson, F. (1999). Self-assessments in HRM: An example of an assessment centre. *Personnel Review*, 28, 124–133.
- Gavin, J. F., & Hamilton, J. W. (1975). Selecting police using assessment center methodology. *Journal of Police Science and Administration*, 3, 166–176.
- Gaugler, B.B. & Thornton, G.C. (1989). Number assessment centre dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, 74, 611-618.
- Hiatt, J. (2000, May). Selection for positions in a manufacturing startup. Paper presented at the 28th International Congress on Assessment Center Methods. SanFrancisco, CA.

Highhouse, S., & Harris, M. M. (1993). The measurement of assessment center situations: Bem template matching technique for examining exercise similarity. *Journal of Applied Social Psychology*, 23, 140–155.

Henry, S. E. (1988). Nontraditional applications of assessment centers. Assessment in staffing plant start-ups. Paper presented at the meeting of the American Psychological Association, Atlanta, GA.

Howard, A. (1997). A reassessment of assessment centers: Challenges for the 21st century. *Journal of Social Behavior and Personality*, 12, 13–52.

Howard, A. (2001). Identifying, assessing, and selecting senior leaders. In S. J. Zaccaro & R. Klimoski (Eds.), *The nature of organizational leadership* (pp. 305–346). San Francisco, CA: Jossey-Bass

Howard, A. (2008). Making assessment centers work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 98-104. doi:10.1111/j.1754-9434.2007.00018.x

Howard, A., & Bray, D. W. (1988). *Managerial lives in transition: Advancing age and changing times*. New York: Guilford Press.

Howard, L. & McNelly, T. (2000, May). Assessment center for team member level and supervisory development. Paper presented at the 28th International Congress on Assessment Center Methods. San Francisco, CA.

Howard, A. & Metzger, J. (2002, October). Assessment of complex, consultative sales performance. Paper presented at the 30th International Congress on Assessment Center Methods, Pittsburgh, PA.

Jacobson, L. (2000, May). Portfolio assessment: Off the drawing board into the fire. Paper presented at the 28th International Congress on Assessment Center Methods, San Francisco, CA.

Joiner, D. A. (2000). Guidelines and ethical considerations for assessment center operations: International task force on assessment center guidelines. *Public Personnel Management*, 29(3), 315.

Joyce, L.W., Thayer, P.W., & Pond, III. (1994). Managerial Functions: An Alternative to Traditional Assessment Center Dimensions? *Personnel Psychology*, 47 (1), 109-121.

Kleinmann, M., & Koller, O. (1997). Construct validity of assessment centers: Appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behavior & Personality*, 12(5), 65-84.

- Kolk, N. J., Born, M. P., & van der Flier, H. (2004). A triadic approach to the construct validity of the assessment center: The effect of categorizing dimensions into a feeling, thinking, and power taxonomy. *European Journal of Psychological Assessment*, 20(3), 149-156.
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center post exercise dimension ratings. *Journal of Applied Psychology*, 89, 377-385.
- Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, 86, 255-264. doi:10.1037/0021-9010.86.2.255
- Lievens, F. (2001a). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, 86, 255-264.
- Lievens, F. (2002b). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, 87, 675-686.
- Lievens, F. (2009). Assessment centres: A tale about dimensions, exercises, and dancing bears. *European Journal of Work and Organizational Psychology*, 18:1, 102-12.
- Lievens, F., & Conway, J.M. (2001). Dimension and Exercise variance in Assessment Centre Scores: A Large Scale Evaluation of Multitrait-Multimethod Studies. *Journal of Applied Psychology*, 86 (2), 1202-1222.
- Lievens, F., Dilchert, S., & Ones, D. S. (2009). The importance of exercise and dimension factors in assessment centres: Simultaneous examinations of construct-related and criterion-related validity. *Human Performance*, 22, 375-390. doi:10.1080/08959280903248310.
- Lievens, F., & Klimoski, R. J. (2001). Understanding the assessment center process: Where are we now? In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 16, pp. 245-286). Chichester, UK: Wiley.
- Lievens, F., & Thornton, G. C. III (2005). Assessment centers: recent developments in practice and research. In A. Evers, O. Smit-Voskuil, & N. Anderson (Eds.) *Handbook of Selection* (pp. 243-264). Blackwell Publishing.
- Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47, 715-738.
- Melchers, K. G., & Annen, H. (2010). Officer selection for the Swiss armed forces: An evaluation of validity and fairness issues. *Swiss Journal of Psychology*, 69, 105-115. doi:10.1024/1421-0185/a000012.

- Melchers, K. G., Henggeler, C., & Kleinmann, M. (2007). Do within-dimension ratings in assessment centers really lead to improved construct validity? A meta-analytic reassessment. *Zeitschrift für Personalpsychologie*, 6, 141-149. doi:10.1026/1617 6391.6.4.141.
- Melchers, K. G., Wirz, A., & Kleinmann, M. (2010). Dimensions AND exercises: Theoretical background of mixed-model assessment centers. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers*. New York: Routledge.
- Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93, 1042–1052.
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80, 252–283.
- Neidig, R. D., & Neidig, P. J. (1984). Multiple assessment center exercises and job relatedness. *Journal of Applied Psychology*, 69, 182-186. doi:10.1037/0021 9010.69.1.182
- Norton, S. D. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. *Academy of Management Review*, 2(3), 442-453.
- Reilly, R. R., Henry, S., & Smither, J. W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology*, 43, 71-84. doi:10.1111/j.1744-6570.1990.tb02006.x
- Rupp, D. E., & Thornton, G. C. III. (2003). Development of simulations for certification of competence of IT consultants. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- 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. (1998). Performance measurement in education and professional certification: Lessons for personnel selection? In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 113–129). Mahwah, NJ: Lawrence Erlbaum.
- Schneider, J. R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology*, 77, 32-41. doi:10.1037/0021-9010.77.1.32
- Shippmann, J. S., Ash, R. A., Pattista, M., Carr, L., Eyde, L. D., Hesketh, B., et al. (2000). The practice of competency modeling. *Personnel Psychology*, 53, 703-740.
- Spychalski, A. C., Quinones, 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.

- Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and crosssituational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34, 397–423.
- Thornton, G. C., III, (1992). *Assessment centers in human resource management*. Reading, MA: Addison-Wesley.
- Thornton, G. C. III, (1993, March). Selecting entry-level brewery workers at Adolph Coors Company. Paper presented at the 21st International Congress on the Assessment Center Method, Atlanta, GA.
- Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Thornton, G. C. III, & Johnson, R. (2006, May). Employment discrimination litigation involving assessment center practices. Presentation in M.M. Harris, *Recent Developments in Employment Discrimination Law and I–O Psychology* at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Thornton, G. C., III, & Mueller-Hanson, R. A. (2004). *Developing organizational simulations: A guide for practitioners and students*. Mahwah, NJ: Lawrence Erlbaum.
- Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, 29, 231-258. doi:10.1177/014920630302900206
- Zedeck, S. (1986). A process analysis of the assessment center method. *Research in Organizational Behavior*, 8, 259-296.