

CHAPTER 5

CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

In this chapter a summarised interpretation of the empirical study and the research results are provided. Conclusions that can be drawn from these findings, as well as limitations of the research are highlighted. The chapter concludes with recommendations for future research purposes.

5.1 INTERPRETATION OF RESEARCH RESULTS

The primary objective of the research was to validate a test battery for the selection of first-line supervisors in a mining company. The aim was to conceptualise relevant concepts and to provide a context which enabled an understanding of the research as well as the dependent and independent variables. From an empirical view the research aimed to determine the correlation between the LPCAT, NT6.1, VC1.1, AccuVision and Assessment Centre scores, and supervisors' job performance as rated by their immediate supervisors or managers.

In line with the research aims, a concurrent validation study was conducted, correlating supervisors' scores on the LPCAT, NT6.1, VC1.1, AccuVision and Assessment Centre with a measure of performance. The company's annual performance review, where employees' job performance are rated by their immediate supervisors or managers was used as the performance measure.

Internal reliability of the instruments was cited to check for internal consistency. The reliability scales of the independent variables, namely LPCAT, NT6.1, VC1.1 and AccuVision were satisfactory and in line with reported instrument properties.

Intercorrelations were then calculated between the instruments to check for overlap. These intercorrelations were acceptable and in line with meta-analytical studies performed to date, except for the lower than expected correlations between the AccuVision and the two cognitive ability tests. However, intercorrelations with the NT6.1 and VC1.1 highlighted an overall dilemma. Both instruments are suitable for administration to persons with a minimum education level of grade 12. The sample included 75 individuals with a lower education level. When the analysis of the intercorrelations was repeated, excluding the scores of these 75 individuals, lower intercorrelations were obtained.

Correlations were calculated between the predictors and the criterion. A statistically significant correlation was found between the LPCAT and supervisory ratings of work performance ($r=0,22$; $p\leq 0,01$) when all education levels were included in the analysis. The analysis excluding the 75 individuals with education levels lower than grade 12 showed no statistically significant relationship between the LPCAT and supervisory ratings of job performance ($r=0,04$). However, the limitations of the study should be taken into account before rejecting the first hypothesis, "There is a statistically significant relationship between the learning potential test scores and job performance".

Statistically significant correlations were found between the ability test predictors and the criterion ($r=0,18$; $p\leq 0,01$ for the NT6.1 and $r=0,19$; $p\leq 0,01$ for the VC1.1). However, when the 75 individuals with education levels lower than grade 12 were omitted from the sample, the correlation analysis reported no positive correlations for both the tests ($r=0,04$ for the NT6.1 and $r=0,03$ for the VC1.1). The limitations of the study should be taken into account before rejecting the second and third hypotheses, "There is a statistically significant relationship between the numerical test scores and job performance" and "There is a statistically significant relationship between the verbal ability test scores and job performance".

No statistically significant correlation was found between the AccuVision test scores and supervisor ratings when all education levels were included in the analysis ($r=0,14$). The analysis excluding the 75 individuals with education levels below grade 12 also reported no statistically significant relationship between the AccuVision test results and supervisory ratings of job performance ($r=0,06$). However, the limitations of the study should be taken into account before rejecting the fourth hypothesis, "There is a statistically significant relationship between the situational judgment test scores and job performance".

A positive correlation, although very weak, was found between the Assessment Centre results and job performance when all education levels were included in the analysis ($r=0,18$; $p\leq 0,05$). The exclusion of the 75 individuals with education levels below grade 12 reported no statistically significant correlation between the Assessment Centre results and supervisory ratings of job performance ($r=0,06$).

The weaker reported correlation could, apart from the other limitations listed, also be the result of only the overall assessment centre score for each individual available for the analyses, and not the scores of the various competency clusters measured. However, the limitations of the study should be taken into account before rejecting the fifth hypothesis, "There is a statistically significant relationship between the Assessment Centre test scores and job performance".

The extraneous variables race, age, gender and education level were considered in the research to determine their moderating effect. Correlations were calculated and reported on between the biographical data and the criterion data. None of the extraneous variables correlated significantly.

Multiple Regression showed the combined predictive power of the predictors. It was evident that none of the predictors correlate with supervisory ratings of work performance. The stepwise regression analysis performed, indicated that only

the LPCAT scores contribute significantly to predicting job performance ($p \leq 0,05$). However, the various limitations of the research should be taken into account before the sixth hypothesis, “There is a significant relationship between the test battery and job performance” is rejected.

As highlighted in the results chapter (Chapter 4) and the aforementioned research summary, no statistically significant correlations between the criterion and predictors were found when the individuals with education levels below grade 12 were omitted in the analyses. The aims of the research as listed in Section 1.3.2 have thus been met.

However, various difficulties were experienced in conducting the study. Instead of predictor raw scores, only the converted sten scores, as outlined in Section 4.2.1 were available by the company. The scores of the separate Key Job Performance/Results Areas per individual could not be obtained, as was the case with the individual competency ratings of the assessment centre.

5.2 CONCLUSIONS

One of the factors that increasingly determine the market value of corporations is the quality of their leadership talent (Lawler & Worley, 2006). Organisations therefore have to ensure that they select people with the required competencies to enter their leadership positions. This need to select leaders, according to the kinds of skills and qualities required of leadership in the organisation, was highlighted at various stages of this research.

The starting point of the selection process, as discussed in Section 2.1.2, is the identification of the knowledge, skills and abilities required to perform the job. Through the process of job analysis a competency framework is determined against which to assess individuals. Defining the job in terms of outputs and

performance is helpful in determining the appropriate use of tests in selection. It is also important to build an understanding of the organisation's needs so that sound hypotheses can be formulated about the relationships between predictors and criteria.

The accuracy, with which psychometric assessments predict future job performance in the selection context, is a crucial factor for the quality of the selection decision. It is therefore important that valid predictors or assessment instruments are included, that is, assessments that correlate significantly with performance are seen to be contributing to productivity. The validation of assessment instruments was discussed in Section 2.3 and their importance were emphasised.

Concurrent validity as a measure of validity was discussed in Section 2.3.1.1. Although a concurrent approach has its shortcomings (refer to Section 5.3.2), this approach was followed due to practical considerations. The job performance of current supervisors within a mining company was compared to their test performance in an attempt to determine whether the instruments in the test battery present a relationship to job performance in the form of supervisory ratings. Through validating the battery in this manner, the organisation is able to determine whether the test battery will add value to the selection of first-line supervisors in the technical disciplines.

Another factor that influenced the choice of a concurrent validity design was that the battery is also used for determining the competency gap with the aim of future training and development of current supervisors.

The results of the research reflect no statistically significant correlations between supervisory ratings of job performance and learning potential, numerical and verbal critical reasoning, situational judgment and assessment centre test scores. The findings did not support the research hypotheses and the regression analysis

suggested that only the LPCAT would add value in the prediction of first-line supervisor job performance.

However, attention should be drawn to the fact that it cannot be concluded that the test battery is not a valid predictor of work performance. The reason for this is that the limitations in the data made available by the company resulted in an underestimation of the results that could be expected. These limitations include the use of an average and not separate KPI score(s) per individual, and the fact that not all individuals were evaluated against the same number of subcategories of the criterion. More detail on the limitations is provided in Section 5.3.2.

As a result, recommendations will be made for future research to ensure that the organisation has a measure of the concurrent validity of the test battery and current and future employees can feel more confident knowing that the tests included in the battery would measure what they intend to. In addition to this, it is suggested that the findings be substantiated via a predictive study, a larger sample and within alternate environments in order to aid generalisability of the research results.

5.3 LIMITATIONS OF THE RESEARCH

A number of limitations were present in the research. These limitations need to be noted and considered in interpreting the research results. The limitations of the literature and empirical study are discussed in the following sections.

5.3.1 Limitations of the literature review

Limitations in the literature review relate largely to the absence of published journal articles dealing with some of the predictors (NT6.1, VC1.1 and AccuVision). Most of the literature sourced dealt with the assessment of mental ability and situational judgement tests in general. Most of the studies therefore

cited with regard to ability assessment in the text are validation studies facilitated by SHL.

5.3.2 Limitations of the empirical study

Due to time constraints and practical consideration a concurrent validity study was conducted as opposed to a predictive validity study. Four major criticisms of concurrent validity studies are assumed to distort the validity coefficient (Barrett et al., 1981). These include the possibility that present employees are not a representative sample of potential applicants, where less effective employees may have been removed from the system as a result of the selection process, and highly effective persons promoted. Restriction of range is a further potential factor where a homogeneous group can be poor at predicting individual differences. Predictive validity studies are also seen as a superior methodology as it is assumed that present employees and job applicants would differ on a variety of physical and motivational characteristics, and the effects of job experience and training are also believed to 'invalidate' concurrent designs. The correlation results from a concurrent validity study therefore need to be interpreted with caution.

There were also limitations which arose from the format in which the various data was made available by the company. The use of an averaged KPI score (job performance) per individual was not ideal and the use of the separate Key Job Performance/Result Areas as outlined in appendix 8, would have added more value to the study. In such an instance the predictors could have been correlated not only with the average KPI score, but also with the scores for each of the subcategories.

Schmidt and Kaplan as cited in Levy (2006) argue that the use of separate criteria versus the use of a composite criterion is an issue of purpose. According

to them, multiple criteria should be examined in the light of predictor variables without combining them into a composite score if the primary goal is psychological understanding. If the primary goal is decision making with an economic focus, they propose that criteria be combined into one measure. Advocates of a composite criterion, as cited in Cascio and Aguinis (2005, p. 76), state that it 'provides a yardstick or overall measure of success or value to the organisation of each individual'. They hold the assumption that the criterion should represent an economic rather than a behaviour construct, that is, overall employee efficiency should be measured in monetary terms. The validation process is therefore carried out only for practical and economic reasons, and not to promote greater understanding of the psychological and behavioural processes involved in various jobs.

It could be argued that the combination of multiple criteria into a composite criterion is often done subjectively. In this regard a quantitative weighting scheme contributes to objectivity in terms of the importance placed on each of the criteria that was used to form the composite (Cascio & Aguinis, 2005).

In this research, the company combined twelve subcategories into one composite score of job performance, without communicating the weighting of each of the subcategories. It was also reported that the final supervisory rating of work performance was obtained by adding the scores of the individual KRAs and then dividing the sum by the number of KRAs that each individual obtained a score for. This implies that all individuals were not evaluated against the same number of subcategories of the criterion. This poses a major limitation when considering a study by Murphy and Shiarella (Cascio & Aguinis, 2005). They found that thirty-four percent of the variance in the validity of a battery of selection tests was explained by the way in which measures of task and contextual performance was combined to form a composite score. It should be emphasised that the relative importance of each of the criterion measures be carefully considered when forming a composite criterion.

In contrast to the arguments in favour of a composite criterion, advocates of multiple criteria also view increased understanding as an important goal of the validation process. 'The goal of the search for understanding is a theory (or theories) of work behaviour; theories of human behaviour are cast in terms of psychological and behavioural, not economic constructs (Schmidt & Kaplan as cited in Cascio & Aguinis, 2005, p. 78).

A further limitation of the study was the fact that only the overall score for the Assessment Centre, and not the separate ratings for the individual exercises (in-basket, performance management interview and group exercise) were made available by the company. The individual exercises could therefore not be correlated with supervisory ratings of job performance. Similarly, the scores for the various competency clusters measured by the assessment centre were not known to the researcher. The researcher was therefore unable to establish which competencies correlate with, for example, learning potential, verbal and numerical ability and supervisory effectiveness.

As discussed in Section 5.1, the NT6.1 and VC1.1 are only suitable for administration to persons with a minimum education level of grade 12. 75 individuals with a qualification below grade 12 were included in the sample. This goes against fair test administration.

Selection decision-making is a comprehensive, integrated process of gathering information about an individual in an attempt to predict future job success. Although the initial research has shown that probably only the LPCAT has some role to play as one part of the first-line supervisor selection decision-making process, various limitations of the available data were highlighted that could have affected the analysis. However, limitations may also be found in the manner in which the performance review is structured or carried out.

In this study, the criterion measure is quite skewed, with only one of the four scores indicating unacceptable performance. The ratings of 2, 3 and 4 all indicate acceptable and above performance. This could also be part of the reason for the poor correlations shown.

Viswesvaran & Ones (2005) indicate that criterion-related validity coefficients can be significantly increased by providing raters with a training programme to ensure they have a common frame of reference as to what each scale point of the performance measure refers to. Such a training programme, which teaches people behaviour observation skills in recording accurately the frequency with which they have seen the employee perform the requirements of the job, has shown to decrease rating error (Latham, Wexly & Pursell as cited in Viswesvaran, 2005). They propose three basic principles of learning to be incorporated in the training, namely, active participation, knowledge of results and practice. The training stresses the necessity of basing a numerical rating on the frequency with which the rater has observed the employee engage in critical job behaviours and requires participants to brainstorm as a group in order to find specific ways of ensuring that a rating is not contaminated by factors irrelevant to job performance.

Generalisability of the research findings was affected in a number of ways. In terms of the data, processing regression weightings can vary depending on the sample. This impacts on the generalisability of the findings to other samples (Anastasi & Urbina, 1997). Cross validation of the results to a new sample is therefore needed. The research was further conducted specifically within the technical disciplines of a mining company. Generalisability of the research results to other environments, industries and types of mining companies is therefore limited.

5.4 RECOMMENDATIONS FOR FUTURE RESEARCH

In the light of the limitations with regard to the data available to the researcher, it is recommended that future research use multiple criteria as opposed to a composite criterion. It is recommended that future studies correlate predictor raw scores with the separate KPI scores, as opposed to a composite criterion or KPI score as outlined in Section 5.3.2. The research was conducted using the overall assessment centre score, and not the scores for the individual exercises, or the individual competency ratings of the assessment centre. It would be of interest to conduct similar research utilising these scores.

The potential disadvantages of a concurrent validation study have been presented in the research limitations above. It would thus be of value to repeat the research from a predictive validity perspective. This would further be beneficial with a larger research sample.

It should be remembered that test results represent only one source of information (Foxcroft & Roodt, 2001; PAI, 2005). It would therefore be of value to conduct further, more comprehensive validation studies including other aspects of the selection battery in an effort to improve the end-to-end process.

The sample included only 5% women, which illustrates the dilemma faced by mining companies with regard to the representivity of women in the technical disciplines. The Women in Mining initiative aims to address this imbalance and it would be of interest and value to conduct a similar study within an environment where there is a greater representivity of females (Rebelo, 2005).

5.5 CHAPTER SUMMARY

In this chapter the main findings of the research were summarised and discussed. Limitations of the research were highlighted and recommendations for future research were suggested.