

THE PREDICTIVE VALIDITY OF THE SELECTION BATTERY USED FOR JUNIOR LEADER TRAINING WITHIN THE SOUTH AFRICAN NATIONAL DEFENCE FORCE

JOHANNES MULLER
JOHANN SCHEPERS

*Programme in Industrial Psychology
Department of Human Resource Management
Rand Afrikaans University*

ABSTRACT

The principal objective of the study was to determine the predictive validity of the test battery used for the selection of junior leaders in the South African National Defence Force. A sample of 96 respondents completed certain indices of the SPEEX-Battery as well as the Advanced Ravens Progressive Matrices test. The test results were compared with the course results. Using canonical correlation analysis, a highly significant relationship was found between the independent variables and the dependent variables ($r = 0,787$; $p < 0,00005$). The predictors with the highest loadings were cognitive ability, conceptualisation, reading comprehension, listening potential, physical stress, and mental stress.

OPSOMMING

Die hoofdoelwit van die studie was om die voorspellingsgeldigheid van die toetsbattery vir keuring van junior leiers in die Suid Afrikaanse Nasionale Weermag te evalueer. 'n Steekproef van 96 respondente het sekere indekse van die SPEEX-Battery asook die Advanced Ravens Progressive Matrices toets voltooi. Die toetsresultate is vervolgens vergelyk met die kursusuitslae. Die veranderlikes is aan kanoniese korrelasie-ontleding onderwerp wat 'n betekenisvolle verwantskap opgelewer het tussen die onafhanklike veranderlikes en die afhanklike veranderlikes ($r = 0,787$; $p < 0,00005$). Die voorspellers met die hoogste ladings was kognitiewe vermoë, konseptualisering, leesbegrip, luisterpotensiaal, fisieke stres en psigiese stres.

Since the First World War psychologists have been studying and devising methods of measuring human attributes with the purpose of predicting future job performance. The goal of measuring human attributes in the work situation is to identify the potential of individuals and to fit them with the right job (Anastasi & Urbina, 1997). The reasons why this has become so important is due to the cost implications of placing individuals in the wrong positions, or indiscriminately selecting candidates for training programmes which in turn could lead to a high failure rate and high costs to the organisation. This is a challenge that psychologists are faced with in the South African National Defence Force (SANDF) where they need to select trainees from a large population to undergo junior leadership training.

Labour costs are often the largest single cost in many organisations, which have resulted in a lot more attention being focused on the selection process. The selection process aims to contribute towards organisational objectives through the acquisition of a competent and motivated workforce. Gatewood and Field (1990) define selection as a process of collecting and evaluating information about individuals in order to extend offers of employment to them. Such employment could be either a first position for a new employee or a new position for an existing employee.

According to Wood and Payne (1998) selection is focused at the point where a decision has to be made about whom to select. As such, it is most concerned with the instruments and methods used to assess candidates. Psychological tests are commonly employed as aids in making a variety of decisions regarding employees, including the selection of employees. According to Anastasi and Urbina (1997) psychological tests have proven to be helpful in such matters as hiring, job assignment, transfers, promotion, and termination of services.

The rationale for using psychometric tests in the selection process lies in the purported ability of the testing instruments to

accurately and objectively assess an applicant's ability to perform the work required by the job (Ritson, 1999). When examining psychometric testing, the focus should thus be on the quality of decision making it allows, and not only on the psychometric properties of the tests, although this should not be interpreted to mean that measurement and test theory are to be regarded as irrelevant or obsolete.

The guidelines of the Society for Industrial Psychology of South Africa on the use of personnel selection procedures state that: "The underlying assumption of any personnel selection procedure is that the procedures used can predict one or another important and relevant behavioural requirement or job performance aspect of the position" (Society for Industrial Psychology, 1992, p. 6). Therefore if an organisation uses psychometric assessment in its selection of employees, it should be because it assists in accurately predicting whether the applicant possesses the behavioural requirements and competencies necessary to perform the required job. "If there is any doubt regarding the ability of a test to provide an accurate idea of an applicant's future performance in the job, then the test itself should be analysed for suitability of purpose" (Ritson, 1999, p. 35).

The purpose of the current study was to ascertain whether the instruments used in selecting candidates for junior leader training in the SANDF are valid. This is in line with the guidelines of the Society of Industrial Psychology which state that the onus is on the practitioners to validate the tests they use, and to give concrete empirical evidence that their selection practices are fair (Society for Industrial Psychology, 1992).

Selection represents a relatively visible mechanism through which access to employment opportunities are regulated. Because of this aspect of selection it has been singled out for intense scrutiny from the perspective of fairness and affirmative action (Milkovich & Boudreau, 1994). Test fairness is in the interest of both the employer and the employee, as both benefit from a fair and meaningful selection process in which the best employment decisions are made.

Advocates of fairness in selection feel that a more concerted effort has to be made in order to gain more information about the comparability of psychometric tests for different ethnic groups. The fundamental principle underlying the empirical fairness of selection instruments used as predictors is that selection techniques should not have an adverse impact on employment opportunities for individuals of different race, age, gender, religion or national origin (Byars & Rue, 1994). Cascio (1997) claims that if an individual from a specific population group does not have an equal opportunity of being selected for a specific post, but has an equal probability of succeeding in the job, test bias exists which could result in unfair discrimination.

It is important to acknowledge that a test battery is designed to discriminate between candidates with higher and lower abilities on certain criteria. A valid selection measure accurately discriminates between those with higher, and those with lower probabilities of job success (Cascio, 1987). The issue is whether the test discriminates fairly.

Applicants in most countries enjoy protection against unfair discrimination. In South Africa Reference to psychometric testing is made in the Employment Equity Act (Act 55 of 1998) Chapter 2 point 8: "Psychometric testing of an employee is prohibited unless the test –

- a) Has been scientifically validated as providing reliable results which are appropriate for the intended purpose;
- b) Can be applied fairly to employees irrespective of their culture, and
- c) Is not biased against people from designated groups."

The issue of discrimination has also received much attention in the United States of America. In this respect it is worthwhile to note the conclusion of the US National Academy of Sciences, as quoted by Schneider and Schmitt (1986, p. 45): "The committee has seen no evidence of alternatives to testing that are equally informative, equally technically adequate, and also economically and politically viable ... and little evidence that well-constructed and competently administered tests are more valid predictors for one population subgroup than for another; individuals with higher scores tend to perform better on the job regardless of group identity". Hough and Oswald (2000) investigated affirmative action programmes and found that they achieved a slight improvement in employment conditions for women and racial minorities but appeared to have virtually no effect on organisational effectiveness. In the USA, reverse-discrimination court cases have also shown that race and other job-irrelevant class membership cannot be used when making employment related decisions.

Anastasi (1982) argues that all behaviour is affected by culture and that cultural influences will always be reflected in test performances. According to Hunter and Hunter (1984) there is an accumulation of evidence that shows results that are consistently opposite of that predicted by the test bias hypothesis, when testing for cultural bias. They argue that if test scores for blacks were lower than their true ability scores, then their job performance will be higher than what is predicted by their test scores.

Regression lines for black applicants, however, were found to be either below or equal to the regression lines for white applicants. This shows that a difference in mean test scores reflects a real difference in mean developed ability.

Huysamen (1996, p. 129) discusses the terms predictive bias and test bias, and describes them using the following example: "if the present test is used to predict future performance as a motor mechanic, men may indeed outperform women in the test. If this is the case, applying the test does not result in predictive bias." This suggests that the instrument is not biased, but that the situation to which it has been applied may be. In addition, this does not necessarily mean that women would not be able

to perform well as motor mechanics. Taking into consideration the history of black versus white education in South Africa as well as language, cultural background and common background of experience of subjects, it could very well turn out that a test is not biased but that the situation to which it is applied is indeed biased.

Hughes (1989, p.12) states that, "At the heart of the question of test fairness is the question of validity. Of particular relevance in personnel selection is criterion related validity." The importance of the validation of any instrument to be used for assessment purposes is highlighted by recent and ongoing developments in South African Labour Legislation, and especially the implications of the Employment Equity Act (Eckstein, 1998).

Before continuing, let us first define validity. "Validity refers to the degree to which available evidence supports inferences made from scores on selection measures" (Gatewood & Field, 1990, p. 302). In the context of human resource selection we want to know how well predictors (such as psychometric tests) are related to job criteria (e.g. training results). If a test is used for more than one purpose, then it has a separate validity for each of them. It is possible that a test may be highly valid for one purpose (e.g. predicting success in selling insurance policies), and at the same time be highly invalid for another (e.g. predicting success in selling men's clothing). Validation must thus be done in relation to the purpose for which the test is used (Kaplan & Saccuzzo, 1997). Validity studies attempt to develop a theory of performance that explains how an individual can meet the demands of a particular job. The most important definitions of validity are those related to content, construct, and criterion-referenced validity, each of which is an evaluative standard in its own right.

It must be recognised, however, that a test or selection method should possess all three types of validity (Anastasi & Urbina, 1997). The reason for assessing criterion validity is that the test or measure is to serve as a stand-in for the measure we are really interested in (Anastasi & Urbina, 1997). The correlation coefficient obtained is known as the validity coefficient. The higher such a correlation, the better the criterion-referenced validity of a test. Criterion related validity relates to both concurrent and predictive validity.

Predictive validity refers to the degree to which a current measure (the predictor) can predict the variable of real interest (the criterion), which is not observed until sometime in the future (Ghiselli, Campbell & Zedeck, 1981; Huysamen, 1996). It involves the collection of data over a period of time. Job applicants, rather than job incumbents are used as the source of the data. Predictor scores are collected from job applicants and the results are then filed. After passage of a suitable period of time, criterion data are then collected. Predictive validation is most relevant for aptitude and interest tests, which are used for selection and classification of job applicants, or applicants for specialised training courses (Huysamen, 1996). If a test possesses predictive validity, it improves the decision making process, which is of particular importance in the selection process.

According to Campbell (1991) a perfect correlation between the predictor and criterion would imply that the decision-maker has a flawless understanding of the predictor-criterion latent structure. It would also mean that he obtained psychometrically flawless measures of all relevant constructs and can thus with perfect precision, and complete certainty, infer values on the intermediate criterion from the combined substitute measure.

The decision-maker would then have a relatively simple selection problem to contend with, because it would imply that he could anticipate the actual outcomes for any applicant with complete certainty, should such an applicant be accepted. Such a situation would, however, be very difficult to create. More often than not the decision-maker's lack of perfect

understanding and complete certainty regarding selection outcomes coupled with his/her reliance on fallible information denies him/her the possibility of anticipating selection outcomes with complete certainty.

Cognitive ability testing

Regarding ability tests, Wood and Payne (1998), found that the proportion of organisations using them to select staff has risen from just under 50% in 1991 to around 75% in 1996, making them as popular as curriculum vitae's. Most ability tests measure maximum performance, in other words what an applicant can do. In terms of contributing to good selection decisions, and leaving aside all other considerations, ability tests deliver the best results, far in excess of personality measures, interviews, or educational attainments. This does not mean that ability tests are excellent at prediction – at best they predict something like 25% of the variance in job performance – but they are the best single personnel selection measure (Wood & Payne, 1998). In a recent meta analysis Schmidt and Hunter (1998) came to the conclusion that general mental ability should be considered as the primary personnel measure for hiring decisions, and that other personnel measures, such as integrity tests, conscientiousness tests, employment interviews, peer ratings, reference checks, job experience, biographical data, years of education, and interests, should only be seen as supplements to general mental ability measures (Schmidt & Hunter, 1998).

Schmidt and Hunter (1998) found that employers reasoned that they used general mental ability tests to select employees who will have the highest level of performance on the job. What they did not realise was that they were at the same time selecting those who would learn the most from job training programmes, and those who would acquire job knowledge faster from experience on the job. Hunter (1986) showed that much of the predictive power of cognitive ability tests is explained by the relationships between cognitive ability, job knowledge, and job performance. According to him these relationships can be explained by the fact that general cognitive ability predicts who will master job knowledge and who will not.

According to Hunter (1986) general cognitive ability predicts performance ratings in all lines of work, although the validity is much higher for complex jobs than for simple jobs. He also found that general cognitive ability predicts training success, at a slightly lower level than performance ratings, uniformly high for all job levels. Therefore it makes sense to use training success as an indicator of potential job success rather than performance ratings on the job. Schmidt and Hunter (1998) found that the major direct causal impact of mental ability on job performance was due to the acquisition of job knowledge.

They also found a smaller direct causal impact of mental ability on job performance independent of job knowledge. For non-supervisory jobs the direct causal effect was found to be about 20% as large as the indirect causal effect, and for supervisory jobs the direct causal effect was found to be about 50% as large as the indirect causal effect (Borman, White, Pulakos & Oppler, 1991).

Learning in a formal training programme means absorbing knowledge, which is presented directly to the student. Hunter (1986) found that cognitive ability predicts job performance largely because it predicts learning and job mastery. He argued that this may be because high ability workers are faster at cognitive operations on the job, are better able to prioritise conflicting rules, are better able to adapt old procedures to altered situations, are better able to innovate to meet unexpected problems, and are better able to learn new procedures quickly as the job changes over time. The most common argument against the apparent importance of general mental ability testing in employment is that differences in job performance stem primarily from differences in specific learned

skills, therefore mental tests predict job performance only because they either measure those relevant skills or predict who is most likely to acquire them (Hunter, 1986).

Followers of this school of thought therefore believe that differences in job performance depend more on training than on intelligence. A meta analysis done by McDaniel and Schmidt (1985) provides evidence that this belief is not true. They found that the predictive value of relevant training and experience yielded low to moderate correlations with later job performance, but the predictive value of training and experience drop among workers with increasingly higher levels of experience. In contrast to this the predictive value of cognitive ability remains high even for experienced workers (McDaniel, 1986).

If differences in job experience are controlled for, the direct impact of cognitive ability on job knowledge rises, as does the indirect impact on work sample performance. It appears likely then that more extensive training or experience in relevant job skills can temporarily render less intelligent workers equally productive as more intelligent but less experienced workers, but that the latter will eventually outperform the former. The reason for this is probably because the more able workers develop expertise more quickly from the same increment in experience.

Research on predictors other than general cognitive ability, reviewed by Hunter and Hunter (1984), shows that if people are to be trained for their job after being hired, then there is no other predictor with validity nearly as high as that of general cognitive ability. Hunter and Hunter's (1984) meta analysis led them to conclude that if general cognitive ability alone was used as a predictor, the average validity across all jobs was $r = 0,54$ for a training success criterion and $r = 0,45$ for a job proficiency criterion. This has major implications for personnel selection practices. It implies that a test of general mental ability should be considered for inclusion in virtually all selection procedures (Cooper & Robertson, 1995).

Sometimes, however, there are perfectly good reasons for discounting the use of a test of general mental ability, for example, when the candidates are all likely to have similar levels of ability (e.g. when graduates are being selected). Another situation where the use of a general mental ability test needs to be considered carefully is when there is the possibility of bias against members of ethnic or other specific subgroups of the population. Research evidence, however, shows that cognitive ability tests do not produce significantly more errors of prediction for one group than another (Schmitt & Noe, 1986). Hunter, Schmidt, and Rauschenberger (1984) found that general cognitive ability tests predict performance equally well for Blacks, Hispanics, and Whites. One could, however, question the relevance of these findings for application in the South African context due to huge educational differences between race groups.

A survey of 1020 experts in psychometrics and behaviour genetics reported that 53% believe that genes and environment are both involved in the mean Black-White IQ difference, compared to 17% who attribute the cause only to the environment, with the remaining 30% stating that there is insufficient evidence for any conclusion (Snyderman & Rothman, 1986).

Many psychologists question the use of cognitive tests in selection and feel that it is necessary to conduct an organisation specific criterion-related validation study before using cognitive tests. This view is based on the fact that the size of the validity coefficients obtained for cognitive tests, differs across many studies, often by some large margin.

Hunter and Schmidt, however, showed that when adjustments were made for various factors such as sample size and restriction in the range of scores available, the studies of cognitive tests gave

remarkably consistent results (Murphy, 1988). These results showed that cognitive tests are valid in a wide range of settings and can be used to predict people's performance in most jobs. For this reason it was decided to include a cognitive ability test in the selection battery under discussion.

Studies investigating the relationship between cognitive ability and managerial potential have found consistent support for a positive correlation between verbal intelligence and managerial potential, although the strength of the relationship has been questioned. The work of Ghiselli et al. (1981) suggests a relatively weak correlation (average $r = 0,30$). It has, however, been argued that the correlation is actually much higher once one controls the statistical artifacts that distort many of the studies (Hunter & Hunter, 1984). Hunter and Hunter (1984) reported correlations in excess of 0,50 using validity generalisation techniques to correct for some of these artifacts (sample size and unreliability of measures).

Training versus Operational or Career Success

One question that needs to be addressed is exactly what one should be assessing or attempt to predict when selecting candidates to be trained as junior leaders in the military. Many psychologists would argue that the task of selectors is to decide who is worth training. Others would counter argue that this objective might result in simply accepting a group of superior trainees. The two sides of this argument tend to line up intellectual measures (educational attainment and aptitude tests) on the one side, and personality assessments (group exercises, personality inventories and interviews) on the other side. The advocates of operational or career success are not saying that initial selection measures should be such strong predictors of operational efficiency and promotability that training and appraisal of potential and proficiency are no longer needed. Selection measures would then be all that are needed for personal planning and development. What they are saying is that optimising prediction against training criteria may result in the exclusion of individuals who would be adequate or better suited in operational units and/or the inclusion of individuals who can cope with training demands, but not necessarily with their operational duties.

The training success advocates on the other hand examine the prediction validity of available predictors and conclude that longer term forecasting is difficult and attempts at it may do more harm than good. They argue that it is the trainers' job to develop and equip trainees to be effective in operational situations. The training course content and its measures of success should be well related to later requirements, or else it should be changed as well as the selection criteria with it. They might also point out that the collection of operational performance data is very difficult and that the criteria used in such validity studies usually turn out to be superiors' ratings, which may exhibit poor reliability (Zeidner & Drucker, 1988). Obviously both these arguments have some strength although both sides may overstate their case. This is clear from the fact that not many individuals are successful in training and then proceed to show poor operational performance. There is also some evidence that aggregated assessments can predict operational and career success even when selection and training effects have considerably reduced the amount of variation in the initial selection measure (Gardner & Williams, 1973).

Training scores are most commonly used as interim criteria in the process of validating selection measures for many jobs. A notion that encourages the use of training scores is that they represent necessary conditions for initial commissioning. Identifying those individuals who have no chance of completing the conditional stage of training is useful and necessary. However, there needs to be an awareness of what the selection score actually predicts whenever training scores are used as criteria. The danger embedded in selecting applicants

according to their prospects of completing training, emphasises the need to clearly establish how well training is related to job performance.

In order to become a junior leader in the SANDF a candidate must successfully complete a 19-week training course. Due to the fact that only a certain number of people can undergo junior leadership training, the organisation has to be selective in its choice of candidates for training. The organisation needs to determine which candidates have the potential to reach a certain level of success at the end of the junior leader-training programme. Only the best candidates will therefore be allowed to continue with the junior leader-training programme. Training scores will be used as interim criteria in the process of validating the selection battery.

Research objectives/Aim of the research

In view of the issues raised in the previous discussion, the following objectives are set:

- To examine the predictive validities of the measuring instruments used in the selection of junior leader trainees for the SANDF
- To determine cut-off points on the measuring instruments used for selection purposes
- To determine the structure of the measuring instruments and estimate the reliabilities of each of the composite scores.

METHOD

Research design

The SANDF is currently using the Advanced Ravens Progressive Matrices test to select trainees for junior leader training. The SANDF has recently acquired the Situation-specific Evaluation Expert test battery and wanted to validate certain of the SPEEX indices for selection of trainees for junior leader training. Due to time constraints during the 19-week course the applicable SPEEX indices could only be administered to the trainees during the last week of their course.

Sample

The study was conducted using a sample of employees in the service of the SANDF who applied for training as junior leaders. The sample consisted of soldiers who had already completed at least one-year initial military service as well as soldiers nominated by their respective units for officer formative training.

The Advanced Ravens Progressive Matrices test was administered to this group. Their scores were then sorted from highest to lowest according to race and gender. The best 100 trainees were then chosen according to a quota of 70% Black, 20% White, and 10% Coloured. The male: female ratio sought was 80: 20. Four trainees did not complete the 19-week course, which further reduced the sample to 96.

Measuring instruments

The criteria to be predicted are the results of the training course. Trainees write an exam at the end of each of the modules that are presented to them over the 19 week period, and need a 60% pass mark. The modules that have to be successfully completed are: (1) Leadership, Command and Management Skills, (2) Communication Skills, (3) Officership, (4) Art of War, (5) Military Studies, and (6) Environmental Studies.

The independent variables or predictors represent those attributes that have been identified through a job analysis as being important for job success. The Advanced Raven's Progressive Matrices test, hereafter referred to as ARPM, was used to measure general cognitive ability. The SANDF recently also started using the Situation-specific Evaluation Expert (SPEEX) for selection purposes and would therefore like to validate

certain of the indices of the SPEEX for this purpose. The SPEEX indices that were identified form part of a generic leadership test battery (SPEEX 100 - Conceptualisation, SPEEX 1600 - Reading Comprehension, SPEEX 1700 - Listening Potential and SPEEX 2200 - Humanising).

Advanced Raven's Progressive Matrices (ARPM)

John Raven developed the Raven Progressive Matrices (RPM) test. It is a non-verbal group ability test (Raven, 1960). One may administer the Raven Progressive Matrices test to groups or individuals. The popularity of the test can probably be ascribed to the ease of administration and interpretation to which it lends itself. The ARPM consists of 36 multiple-choice items of progressively increasing complexity. It is expected of the respondent to complete a matrix pattern given eight possible alternative answers.

The ARPM is designed to assess "a person's capacity, (at the time of the test), to apprehend meaningless figures presented for observation, see the relations between them, conceive the nature of the figure completing each system of relations, and by so doing, develop a systematic method of reasoning" (Raven, 1960 p. 45). According to Raven (1960) the Standard Progressive Matrices does not differentiate clearly amongst adults of superior intellectual capacity, and therefore the Advanced Progressive Matrices test was specifically developed for use with superior adults (Raven, 1965).

Carpenter, Just and Shell (1990) found that the ability to educate relations, and manage a large set of problem-solving goals, tends to distinguish between high and low scoring subjects on the RPM. They mention several reasons why the RPM is appropriate in the study of cognitive, analytical abilities:

- The large number of items included in the RPM lends itself to experimental analysis of problem solving behaviour.
- Correlations between RPM scores and other measures of intellectual achievement suggest a general underlying construct similar to Spearman's g-factor rather than specific aspects of cognitive functioning.
- The RPM is commonly used in research which requires that language processing be minimised.
- Several studies have concluded that the RPM measure processes central to analytical intelligence.

With regard to construct validity, the RPM is generally seen as measuring fluid intelligence and it provides a particularly pure measure of general intelligence, or Spearman's 'g' factor (Paul, 1985). In fact, the RPM may be the best single measure of 'g' available as shown through multidimensional scaling by Marshalek, Lohman and Snow (1983). The RPM was originally designed to assess military recruits irrespective of their educational background (Kaplan & Saccuzzo, 1997).

A survey of reliability studies shows a wide range of coefficients, from the high 0,70's to the low 0,90's. Early studies revealed a fairly high correlation between the RPM and the Stanford-Binet of $r = 0,60$ (Keir, 1949), Wechsler performance IQ of $r = 0,70$ (Hall, 1957), and Wechsler verbal IQ of $r = 0,58$ (Hall, 1957).

Situation- Specific Evaluation Expert (SPEEX)

The aim of the SPEEX is to provide a comprehensive assessment package suitable for the assessment and development of human potential in the workplace. The various SPEEX indices assess human potential relating to the following dimensions or basic competencies and are identified in the SPEEX battery manual (Erasmus, 2001) as follows:

SPEEX 100: Conceptualisation

A normative scale consisting of 30 items, where the respondent must complete a pattern. There is a time limit of 18 minutes to complete the test. It assesses the potential to reason in spatial terms; to see the relationship between parts; to complete the picture; to envisage the whole or end result; and to anticipate the outcome. Cronbach coefficient alpha of the scale $r_{xx} = 0,90$ (Schaap, 2001).

SPEEX 1600: Reading Comprehension

A normative scale consisting of 20 items. Respondents get five minutes to read through a couple of paragraphs and must then answer 20 questions on the content of those paragraphs within eight minutes. It assesses the potential or capacity to read and understand what has been read clearly and objectively. Cronbach coefficient alpha of the scale $r_{xx} = 0,85$ (Schaap, 2001).

SPEEX 1700: Listening Potential

It is a normative scale. The respondents must listen to a recording for five minutes and then answer 20 specific questions on the content of the recording within eight minutes. It assesses the potential or capacity to listen and to understand what has been heard clearly and objectively. The Cronbach coefficient alpha of the scale $r_{xx} = 0,72$ (Schaap, 2001).

SPEEX 2200: Humanising

A normative scale that consists of 96 items, which needs to be completed within 25 minutes. The respondents must indicate on a seven point Likert scale the extent to which certain statements apply to their own behaviour, attitudes, or beliefs. It aims to determine whether the respondent is more task or people orientated in his/her outlook and application. The humanising scale gives an indication of the respondents' orientation in respect of the following:

- Empathy (Speex2201) – refers to the disposition of a person to show concern, tolerance, sympathy and understanding for the needs, concerns, values, views, attitudes, behaviour, beliefs etc. of other people. The Cronbach coefficient alpha of the scale $r_{xx} = 0,80$ (Schaap, 2001).
- Emotional sensitivity (Speex2202) – refers to the capacity to understand and appreciate why people feel as they do when they are intolerant, concerned, downhearted, moody, angry etc. The Cronbach coefficient alpha of the scale $r_{xx} = 0,83$ (Schaap, 2001).
- Tact (Speex2203) – refers to the disposition of a person to be courteous, diplomatic, comforting, respectful, accommodating etc. when attending to the problems or difficulties people experience. The Cronbach coefficient alpha of the scale $r_{xx} = 0,83$ (Schaap, 2001).
- People development (Speex2204) – refers to a person's appreciation of the developmental needs of workers in the workplace and concern with the effective implementation of development procedures. It also relates to what quality time and attention is devoted to development as a very important and integral part of daily activities in the workplace. The Cronbach coefficient alpha of the scale $r_{xx} = 0,86$ (Schaap, 2001).
- Mental stress (Speex2205) – refers to the capacity of a person to cope with emotional stress and pressure. The Cronbach coefficient alpha of the scale $r_{xx} = 0,80$ (Schaap, 2001).
- Interpersonal objectivity (Speex2206) – refers to the inclination of a person to understand interpersonal matters for what they really mean. The Cronbach coefficient alpha of the scale $r_{xx} = 0,76$ (Schaap, 2001).
- Physical stress (Speex2207) – refers to the physical capacity of a person to cope with social and emotional stress, as reflected by the absence of psychosomatic symptoms i.e., the physical manifestation of symptoms of stress such as ulcers, headaches, extreme sweating etc. The Cronbach coefficient alpha of the scale $r_{xx} = 0,82$ (Schaap, 2001).
- Diversity facilitation (Speex2208) – refers to the capacity of a person to relate positively to teams or groups whose composition reflects diversity in gender, culture, language, beliefs, attitude, behaviour etc. The Cronbach coefficient alpha of the scale $r_{xx} = 0,63$ (Schaap, 2001).

The SPEEX-battery consists of two types of scales namely cognitive and behavioural scales (Erasmus, 2001). Scales 100 (Conceptualisation), 1600 (Reading Comprehension), and 1700 (Listening Potential) are cognitive scales, which means that they assess intellectual potential.

Speex100 (Conceptualisation) is a visual scale and “because it comprise visual items, it could therefore be administered in any language whatsoever” (Erasmus, 2001, p. 98). The Speex100 has furthermore been designed to measure participants ranging from the lowest to the highest levels of sophistication, indeed from levels with virtually no formal education to levels of high educational or formal development (Erasmus, 2001). “This scale can therefore be used to establish a person’s functional cognitive potential” (Erasmus, 2001, p. 98). Speex2200 (Humanising) is a behavioural scale.

Kriel (2001), has conducted an item bias analysis on the Speex1600 (Reading Comprehension) and found correlations between z-scores in the range of 0,903 to 0,993 between language groups Afrikaans, English, Northern Sotho, Zulu, Southern Sotho, Xhosa, and Tswana.

Procedure

The Advanced Ravens Progressive Matrices test was administered to a large number of applicants who applied to be trained as junior leaders. Their scores were then sorted from highest to lowest according to race and gender. A quota of 70% Black, 20% White and 10% Coloured and Asian was sought. The ratio of males to females sought was 80:20.

There was also a number of chaplain candidates that had to be accommodated on the course irrespective of their ARPM scores. The reason for this is that they need to complete officers training before they can be appointed as chaplains. During the last week of the 19-week course the Speex100 – conceptualisation, Speex1600 – reading comprehension, Speex1700 – listening potential, and Speex2200 – humanising were administered to the trainees. Due to time constraints during the 19-week course the SPEEX indices could unfortunately only be administered to the trainees during the last week on course.

Statistical analysis

A descriptive and exploratory design was used. Descriptive statistics were used for the ARPM, Speex100, Speex1600, Speex1700, Speex2201 to 2208, and the six respective course modules.

For regression analysis there should ideally be ten observations for every measuring instrument used to ensure sufficient degrees of freedom (Tabachnick & Fidell, 1996). “ R^2 is not an unbiased estimate of the corresponding parameter in the population. The extent of this bias depends on the relative size of N and p . When $N = p + 1$, prediction is perfect and $R = 1$, regardless of the true relationship between Y and X_1, X_2, \dots, X_p in the population” (Howell, 1997, p. 521). In the present study there are 12 independent variables and 6 dependent variables, which implies that ideally there should be at least 180 observations or respondents. To overcome the problem of too few observations a principal components analysis was done that yielded four

components, therefore requiring a minimum of 40 observations for a regression analysis.

Apart from principal components analysis a canonical correlation analysis was also done. Since canonical correlation analysis is not as well known as other correlational techniques a short description of the technique follows. According to Hair, Anderson, Tatham and Black (1995) canonical correlation analysis can be viewed as a logical extension of multiple regression analysis. Whereas multiple regression analysis involves a single dependent variable and several independent variables, canonical correlation analysis involves several dependent variables and several independent variables. The underlying principle of canonical correlation analysis is to develop a linear combination of two sets of variables (the dependent and independent), and to maximise the correlation between the two sets.

RESULTS

Descriptive statistics in respect of the independent and dependent variables are presented in Table 1. From the analysis of the data it can be seen that the mean and median of the variables are relatively close to one another for all the variables.

Usually the mean is the best measure for describing a set of data except in the case of extreme values when the median will be a better alternative to use.

For all parametric statistics the assumption of normality of distributions must be met. The two indicators of normality are skewness and kurtosis. The skewness of a distribution refers to whether the scores are equally distributed on both sides of the mean or not. Kurtosis on the other hand relates to the peakedness of the distribution. Leptokurtosis results in a reduction of the variance of a variable. If a distribution is normal the coefficients of skewness and kurtosis are zero or close to it. Table 1 shows that the Speex100 is negatively skewed, while the ARPM and the Speex1600 are platykurtic, which means that both of them have a wide dispersion of scores and should therefore be very reliable.

The pass rate of all the modules is 60%. Of the 96 trainees none failed the Communication and Military Studies modules. Only two failed Environmental Studies, while eight failed the Officership module. As far as the Art of War and Leadership and Command modules are concerned, nine and ten trainees respectively failed. In total, 18 of the 96 trainees failed one or more modules. The means of the different modules are 73% for Leadership and Command, 74% for Officership, 77% for Art Of War, 80% for Environmental Studies, 83% for Military Studies, and 85% for Communication. The mean of the six modules is 79%.

With canonical correlation analysis in mind the suitability of the ARPM and the Speex100 – Conceptualisation, 1600 – Reading

TABLE 1
DESCRIPTIVE STATISTICS IN RESPECT OF DEPENDENT
AND INDEPENDENT VARIABLES

	Mean	Median	SD	Skewness	Kurtosis	Minimum	Maximum
ARPM	16,46	16,00	6,987	0,015	<u>-1,001</u>	4	31
Speex100	19,98	21,00	4,314	<u>-1,079</u>	0,629	7	26
Speex1600	10,19	11,00	3,495	-0,131	<u>-0,944</u>	3	17
Speex1700	10,74	11,00	3,020	-0,185	-0,079	3	18
Speex 2200 average	4,01	4,00	0,414	-0,267	0,209	2	5
Communication	85,38	86,00	7,783	-0,605	-0,476	67	97
Leadership and Command	73,92	75,00	11,981	-0,334	-0,191	43	98
Officership	74,51	74,50	10,572	-0,341	-0,272	46	94
Art of War	77,11	78,00	11,829	-0,687	0,502	41	99
Military Studies	83,28	85,00	7,241	-0,615	-0,102	64	95
Environmental Studies	80,80	82,00	8,853	-0,739	0,724	52	97

TABLE 2
MATRIX OF INTERCORRELATIONS OF SPEEX-INDICES, ARPM, AND COURSE RESULTS

	1	2	3	4	5	6	7	8	9
1 Communication	1,000								
2 Leadership and Command	0,592**	1,000							
3 Officership	0,688**	0,776**	1,000						
4 Art of War	0,590**	0,767**	0,711**	1,000					
5 Military Studies	0,549**	0,550**	0,500**	0,567**	1,000				
6 Environmental Studies	0,502**	0,773**	0,741**	0,773**	0,554**	1,000			
7 ARPM	0,508**	0,451**	0,598**	0,594**	0,291**	0,450**	1,000		
8 Speex100	0,411**	0,282**	0,399**	0,408**	0,147	0,261*	0,643**	1,000	
9 Speex1600	0,380**	0,345**	0,545**	0,491**	0,235*	0,384**	0,485**	0,367**	1,000
10 Speex1700	0,353**	0,193	0,318**	0,305**	0,302**	0,226*	0,350**	0,344**	0,385**
11 Speex2201	0,030	0,081	0,095	-0,031	0,085	0,067	-0,044	-0,023	0,014
12 Speex2202	-0,011	0,055	0,024	-0,040	0,119	0,026	0,006	-0,018	-0,004
13 Speex2203	0,071	0,129	0,160	0,048	0,196	0,134	0,038	0,004	0,008
14 Speex2204	0,008	0,077	0,025	0,035	0,061	-0,014	0,142	0,187	-0,039
15 Speex2205	0,328**	0,326**	0,375**	0,361**	0,402**	0,334**	0,326**	0,166	0,267**
16 Speex2206	0,054	0,041	0,053	-0,049	0,189	-0,043	0,043	0,122	-0,017
17 Speex2207	0,388**	0,415**	0,407**	0,434**	0,467**	0,340**	0,360**	0,215*	0,221*
18 Speex2208	0,180	0,072	0,145	0,170	0,121	0,072	0,251*	0,100	0,165
10 Speex1700	1,000								
11 Speex2201	0,133	1,000							
12 Speex2202	0,078	0,766**	1,000						
13 Speex2203	0,049	0,625**	0,723**	1,000					
14 Speex2204	0,005	0,448**	0,579**	0,590**	1,000				
15 Speex2205	0,298**	0,133	0,095	0,223*	0,121	1,000			
16 Speex2206	0,044	0,582**	0,638**	0,593**	0,549**	0,146	1,000		
17 Speex2207	0,249*	0,083	0,073	0,144	0,132	0,601**	0,120	1,000	
18 Speex2208	0,144	-0,177	-0,286**	-0,238*	-0,136	0,316**	-0,144	0,291**	1,000

Note:
 ** Correlation is statistically significant at the 0,01 level (2-tailed)
 * Correlation is statistically significant at the 0,05 level (2-tailed)

TABLE 3
STATISTICAL SIGNIFICANCE OF CANONICAL CORRELATIONS: BARTLETT'S TEST

Eigenvalues	Canonical R	Root Removed	Significance of Remaining Roots		
			Chi-Square	df	p
0,619565	0,787	0	128,7890	72	<0,00005
0,237346	0,487	1	46,1583	55	0,796
0,107796	0,328	2	22,9921	40	0,986
0,076677	0,277	3	13,2399	27	0,988
0,045424	0,213	4	6,4191	16	0,983
0,028184	0,168	5	2,4443	7	0,931

Comprehension, 1700 – Listening Potential, and 2200 – Humanising as independent variables and the course results as dependent variables were investigated by scrutiny of the intercorrelations between the variables (Table 2).

The intercorrelations between the six course results vary between 0,500 and 0,776 significant at the 0,01 level. The ARPM, Speex100 (Conceptualisation), Speex1600 (Reading Comprehension), and the Speex1700 (Listening Potential) intercorrelate substantially with one another (r = 0,344 to r= 0,643; p<0,01) and quite highly with most of the course results.

Exceptions to this are the Speex100 (Conceptualisation) that is uncorrelated with Military Studies and Environmental Studies, the Speex1600 (Reading Comprehension) that is lowly correlated with Military Studies, and the Speex1700 (Listening Potential)

that is uncorrelated with Leadership and Command as well as with Environmental Studies.

Only the Speex2205 (Mental Stress) and Speex2207 (Physical Stress) of the Speex2200 (Humanising) indices are statistically significantly correlated with the course results, ARPM, Speex1600 (Reading Comprehension), and Speex1700 (Listening Potential). The Speex2207 (Physical Stress) is also statistically significantly correlated with the Speex100 (Conceptualisation) while the Speex2205 (Mental Stress) is uncorrelated with the Speex100 (Conceptualisation).

Before proceeding with canonical correlation two statistical assumptions must first be met. The first is that there must be a linear relationship between the x and y canonical variates, and the second is that the distributions of all the variables must be

TABLE 4
CANONICAL CORRELATIONS OF TEST BATTERY (12 IV'S) WITH COURSE RESULTS (6 DV'S)

Canonical Correlations	Correlations of Original Measures with Canonical Variates						
	Variate 1	Variate 2	Variate 3	Variate 4	Variate 5	Variate 6	
Independent variables							
1 ARPM	<u>0,858</u>	-0,121	0,185	-0,073	0,037	-0,108	
2 Speex100	<u>0,478</u>	0,281	0,192	0,435	0,313	0,098	
3 Speex1600	<u>0,764</u>	-0,131	-0,087	0,163	-0,354	0,283	
4 Speex1700	<u>0,613</u>	-0,213	0,423	0,139	0,134	-0,410	
5 Speex2201	0,018	0,230	-0,490	-0,034	-0,121	-0,543	
6 Speex2202	-0,050	0,357	-0,308	-0,145	-0,119	-0,225	
7 Speex2203	0,113	0,394	-0,523	-0,029	-0,131	-0,171	
8 Speex2204	0,015	0,174	0,148	-0,313	-0,401	-0,191	
9 Speex2205	<u>0,487</u>	0,477	-0,110	-0,138	0,236	0,158	
10 Speex2206	-0,004	0,572	-0,087	0,170	-0,469	-0,476	
11 Speex2207	<u>0,540</u>	0,598	0,182	-0,426	0,100	-0,075	
12 Speex2208	0,261	0,064	0,332	0,288	0,007	0,150	
Average % variance accounted for	21,15%	11,98%	8,75%	5,62%	6,12%	8,00%	Total: 61,62%
Average % redundancy	13,10%	2,84%	0,94%	0,43%	0,28%	0,23%	Total: 17,82%
Canonical Correlations							
	Correlations of Original Measures with Canonical Variates						
	Variate 1	Variate 2	Variate 3	Variate 4	Variate 5	Variate 6	
Dependent Variables							
1 Communication	<u>0,740</u>	0,211	0,211	0,117	0,382	-0,453	
2 Leadership and Command	<u>0,668</u>	0,195	-0,122	-0,643	0,072	-0,287	
3 Officership	<u>0,914</u>	0,090	-0,304	-0,082	-0,089	-0,221	
4 Art of War	<u>0,860</u>	0,036	0,164	-0,400	0,153	0,220	
5 Military Studies	<u>0,502</u>	0,818	-0,020	-0,116	0,217	0,132	
6 Environmental Studies	<u>0,696</u>	0,057	-0,423	-0,384	0,416	0,110	
Average % variance accounted for	55,10%	12,73%	5,97%	12,58%	6,71%	6,91%	Total: 100%
Average % redundancy	34,14%	3,02%	0,64%	0,97%	0,31%	0,19%	Total: 39,27%
Canonical Correlation	<u>0,787</u>	0,487	0,328	0,277	0,213	0,168	
Variance	<u>0,620</u>	0,237	0,108	0,077	0,045	0,028	

normal to allow for maximum correlations amongst the variables. Canonical analysis can be used with non-normal variables if the distribution form (e.g., highly skewed) does not decrease the correlation with other variables.

Canonical correlation was used to quantify the strength of the relationship between the two sets of variables (12 independent and 6 dependent variables). The results are shown in Tables 3 and 4. Bartlett's test of the statistical significance of the canonical correlation is shown in Table 3 and the canonical correlation analysis is given in Table 4.

From Table 3 it can be seen that there is only one statistically significant canonical correlation: $\chi^2(72) = 128,789$; $p < 0,00005$ and the canonical correlation is 0,787.

From Table 4 it can be seen that the first variate has high loadings on the ARPM (0,858), Speex100 (0,478), Speex1600 (0,764), Speex1700 (0,613), Speex2205 (0,487), and Speex2207 (0,540) in respect of the independent variables.

The first variate also has high loadings on all six dependent variables: Communication (0,740); Leadership and Command (0,668); Officership (0,914); Art of War (0,860); Military Studies (0,502); and Environmental Studies (0,696). The independent variables account for 21,15% of the variance of the x-component of the first variate and the dependent variables account for 55,10% of the variance of the y-component of the first variate.

The correlation between the x- and y-components is 0,787. The redundancy index of 13,10%, in respect of the first variate

indicates the amount of variance of the dependent variables accounted for by the independent variables. Sixty two percent of the variance of the y-variate is accounted for by the x-variate, with high loadings on the ARPM, Speex100 (Conceptualisation), Speex1600 (Reading Comprehension), Speex1700 (Listening Potential), Speex2205 (Mental Stress), and Speex2207 (Physical Stress).

According to Cliff (1987), classical factor analysis is an approach that attempts to explain the relationship between observed variables in terms of latent variables, and conceivably uses the manifest variables to measure the latent variables. Components analysis on the other hand uses a composite of the manifest variables as a summary of those variables.

TABLE 5
EIGENVALUES OF UNREDUCED INTERCORRELATION MATRIX IN RESPECT OF COURSE RESULTS

Root	Eigenvalue	% of Variance	Cumulative %
1	<u>4,233</u>	<u>70,551</u>	<u>70,551</u>
2	0,603	10,055	80,606
3	0,513	8,554	89,160
4	0,262	4,361	93,521
5	0,216	3,596	97,117
6	0,173	2,883	100,000
Trace	6,000		

TABLE 6
ROTATED PRINCIPAL COMPONENT MATRIX OF DEPENDENT VARIABLES (DIRECT OBLIMIN ROTATION)

Variables	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
Leadership and Command	<u>0,894</u>	-0,206	0,015	0,102	-0,370	0,107
Officership	<u>0,884</u>	-0,130	-0,256	0,250	0,094	-0,257
Art of War	<u>0,882</u>	-0,160	0,103	-0,405	-0,003	-0,148
Environmental Studies	<u>0,871</u>	-0,284	0,201	0,070	0,261	0,217
Communication	<u>0,771</u>	0,391	-0,466	-0,108	0,039	0,151
Military Studies	<u>0,724</u>	0,534	0,423	0,094	-0,010	-0,060

First a principal components analysis was done in respect of the dependent variables. Table 5 shows the eigenvalues of the unreduced intercorrelation matrix in respect of the course results. Table 6 shows the rotated principal components matrix of the dependent variables.

From Table 5 it can be seen that only the first component has an eigenvalue greater than unity. This component accounts for more than 70% of the total variance of the measures.

Table 6 shows that all six dependent variables have high loadings on the first component. Accordingly a total score was formed by adding the scores (percentages) of the various course results together. Secondly a principal components analysis was done in respect of the independent variables.

Table 7 shows the eigenvalues of the unreduced intercorrelation matrix in respect of ARPM, and SPEEX-indices. Table 8 gives the rotated principal components matrix of the independent variables.

TABLE 7
EIGENVALUES OF UNREDUCED INTERCORRELATION MATRIX IN RESPECT OF ARPM, AND SPEEX-INDICES

Root	Eigenvalue	% of Variance	Cumulative %
1	<u>3,640</u>	<u>30,331</u>	30,331
2	<u>2,859</u>	<u>23,827</u>	54,158
3	<u>1,253</u>	<u>10,440</u>	<u>64,598</u>
4	0,900	7,497	72,095
5	0,677	5,643	77,738
6	0,619	5,160	82,898
7	0,456	3,799	86,697
8	0,417	3,473	90,170
9	0,391	3,255	93,425
10	0,318	2,650	96,075
11	0,282	2,353	98,428
12	0,189	1,572	100,000
Trace	12,000		

From Table 7 it can be seen that three components have eigenvalues greater than unity. The three components account for 30,33%, 23,83%, and 10,44% of the total variance respectively. The cumulative percentage of variance explained by the three components is 64,60%.

Table 8 indicates that the Speex2200 subscales 1 (Empathy),2 (Emotional Sensitivity),3 (Tact),4 (People Development), and 6 (Interpersonal Objectivity) have high loadings on the first component (from here on called emotional sensitivity), ranging from 0,853 to 0,738. The ARPM, Speex100 (Conceptualisation), Speex1600 (Reading Comprehension), and Speex1700 (Listening Potential) have high loadings on the second component (from here on called cognitive ability), ranging from 0,757 to 0,559.

The Speex2200 subscales 5 (Mental Stress),7 (Physical Stress), and 8 (Diversity Facilitation) have high loadings on the third component (from here on called stress tolerance), ranging from 0,545 to 0,448. In forming composite scores, in respect of the various components, the variables with high loadings on each component were given equal weights.

TABLE 8
ROTATED PRINCIPAL COMPONENTS IN RESPECTS OF ARPM, AND SPEEX-INDICES (DIRECT OBLIMIN ROTATION)

Variables	Component 1	Component 2	Component 3
Speex2202	<u>0,853</u>	-0,294	-0,030
Speex2203	<u>0,832</u>	-0,212	0,069
Speex2201	<u>0,788</u>	-0,243	0,049
Speex2206	<u>0,781</u>	-0,176	-0,006
Speex2204	<u>0,738</u>	-0,123	-0,080
ARPM	0,240	<u>0,757</u>	-0,309
Speex1600	0,152	<u>0,636</u>	-0,287
Speex100	0,229	<u>0,619</u>	-0,530
Speex2207	0,315	0,578	<u>0,504</u>
Speex2205	0,356	0,568	<u>0,545</u>
Speex1700	0,239	<u>0,559</u>	-0,166
Speex2208	-0,170	0,514	<u>0,448</u>

The coefficient of internal consistency of the first composite (course results/criterion scores), corresponding to component 1 (total score) was assessed using Cronbach's coefficient alpha (Table 9).

TABLE 9
VARIANCE-COVARIANCE MATRIX OF DEPENDENT VARIABLES

	1	2	3	4	5	6
1	78,371	82,036	80,981	69,344	35,541	34,580
2	82,036	143,551	108,683	98,233	47,697	55,190
3	80,981	108,683	139,934	88,857	48,578	54,336
4	69,344	98,233	88,857	111,768	38,266	56,586
5	35,541	47,697	48,578	38,266	52,436	30,957
6	34,580	55,190	54,336	56,586	30,957	60,574

Cronbach alpha = 0,912

The obtained coefficient alpha of 0,912 for the total score indicates a high reliability. Cronbach's coefficient alpha was also calculated in the same fashion in respect of the independent variables.

It gave the following results: Component 1 (emotional sensitivity) gave a coefficient alpha of 0,886 which indicates a high reliability; Component 2 (cognitive ability) gave a coefficient alpha of 0,724 which indicates an acceptable reliability; and Component 3 (stress tolerance) gave a coefficient alpha of 0,675 which indicates borderline reliability.

The intercorrelations of the dependent (total score) and independent (emotional sensitivity, cognitive ability, and stress tolerance) composites are given in Table 10.

TABLE 10
MATRIX OF INTERCORRELATIONS OF DEPENDENT AND INDEPENDENT COMPOSITES

		1	2	3	4
Dependent variable	1 total score	1,000			
Independent variables	2 emotional sensitivity	0,081	1,000		
	3 cognitive ability	<u>0,584*</u>	0,058	1,000	
	4 stress tolerance	<u>0,456*</u>	0,030	<u>0,393*</u>	1,000

Note:

* Correlation is significant at the 0,01 level (2-tailed)

The total score is highly correlated with cognitive ability (0,584) and stress tolerance (0,456) at a significance level of 0,01. The total score is uncorrelated with emotional sensitivity (0,081). Cognitive ability and stress tolerance is also positively correlated (0,393) at a significance level of 0,01.

A stepwise multiple regression analysis was done of the three independent variables (emotional sensitivity, cognitive ability, and stress tolerance) on the dependent variable (total score). The stepwise multiple regression shows that the best model is found if component 1 (emotional sensitivity) which is essentially uncorrelated with the total score is removed, and only components 2 (cognitive ability) and 3 (stress tolerance) are used. This model is shown in Table 11.

The multiple correlation found between the total score (DV), and cognitive ability and stress tolerance (IV's) is 0,634. This correlation is statistically highly significant and explains 40,2% of the variance of the total score $\{F(2,93) = 31,266; p(F) < 0,001\}$. Table 11 also shows that the beta-weights of cognitive ability (0,479) and stress tolerance (0,268) are statistically significant. The beta-weight of cognitive ability is almost double that of stress tolerance.

DISCUSSION

The primary objective of this study was to validate a selection battery for junior leader training within the South African National Defence Force. The results indicate that the test battery is indeed valid for predicting which candidates would be

successful in the training programme for junior leaders. Descriptive statistics show that trainees scored extremely high on the course modules. They had an overall mean of 79% across the six modules. These extremely high scores might indicate that the tests for the various course modules were very easy and this should be investigated.

The intercorrelations between the course results were found to be generally high, which indicates that to a certain degree the same construct was measured, possibly the ability to learn and assimilate new information. The intercorrelations between the ARPM and SPEEX-indices indicate that the measuring instruments intercorrelated statistically significantly high with one another. The only exception to this was Speex2200 (Humanising) where certain of its indices were uncorrelated with the other measuring instruments. This was confirmed through a canonical correlation analysis that yielded a statistically significant relationship between the six course results on the one hand and the ARPM, Speex100 (Conceptualisation), Speex1600 (Reading Comprehension), Speex1700 (Listening Potential), Speex2205 (Mental Stress), and Speex2207 (Physical Stress) on the other hand.

A principal components analysis further confirmed that the six course results loaded highly on a single component, labelled total score, in respect of the dependent variables, while three components were found in respect of the independent variables. The first component had high loadings on the Speex2201 – Level of Empathy, Speex2202 – Emotional Sensitivity, Speex2203 – Tact, Speex2204 – People Development in the workplace, and Speex2206 – Interpersonal Objectivity, and was labelled emotional sensitivity. The second component had high loadings on the ARPM – learning potential, Speex100 – Conceptualisation, Speex1600 – Reading Comprehension, and Speex1700 – Listening Potential, and was labelled cognitive ability. The third component had high loadings on the Speex2205 – Mental Stress, Speex2207 – Physical Stress, and Speex2208 – Diversity Facilitation, and was labelled stress tolerance.

The four components were intercorrelated and yielded statistically significant correlations with total score, cognitive ability, and stress tolerance. The regression analysis indicated that the inclusion of cognitive ability and stress tolerance yielded the best regression equation. A multiple correlation of 0,634 was obtained, indicating that cognitive ability and stress tolerance account for 40,2 percent of the variance in total score. Cognitive ability makes almost double the contribution that stress tolerance does towards predicting the total score.

One could argue that the dependent variables are all of an academic nature, and that other aspects that are indeed very important for junior leaders is not currently being measured.

TABLE 11
STEPWISE REGRESSION: DEPENDENT VARIABLE (COURSE RESULTS)

		ANALYSIS OF VARIANCE			
		Source of variation	DF	Sum of squares	Mean square
Multiple R:	0,634	Regression	2	2680,534	1340,27
R Square:	0,402	Residual	93	3986,612	42,867
Adjusted R Square:	0,389	F = 31,266; p(F) < 0,001			
Standard Error:	6,547				
VARIABLES IN THE EQUATION					
INDEPENDENT VARIABLES	B	SE B	Beta	t-value	p
cognitive ability	0,530	0,097	<u>0,479</u>	5,492	0,000
stress tolerance	0,290	0,094	<u>0,268</u>	3,077	0,003
(Constant)	8,992	5,299		1,697	0,093

The focus seems to be on measuring whether the trainees are able to cope with the amount of theory and are able to integrate it into practical solutions for every day life.

Some of the indices of the Speex2200 were included in the regression model that yielded the best regression equation (2205 – Mental Stress, 2207 – Physical Stress, and 2208 – Diversity Facilitation), while the rest (2201 – Empathy, 2202 – Emotional Sensitivity, 2203 – Tact, 2204 – People Development, and 2206 – Interpersonal Objectivity) was excluded. The Speex2200 needs to be included or excluded as a whole and therefore a decision has to be made. It can be excluded and replaced by another test that measures stress tolerance (both physical and mental stress), or it can be included on the grounds that any leader without empathy, tact, emotional sensitivity, people development skills and objectivity are destined to fail. One should also remember that the trainees experience a lot of stress while on course and this could inflate the stress tolerance importance compared to normal circumstances where stress tolerance is less important.

The reliabilities of the ARPM, and SPEEX-indices range from 0,63 to 0,90. The variance-covariance formula was used to calculate Cronbach's coefficient alpha for the four components. This yielded reliabilities between 0,68 and 0,91, which is better than for the individual measuring instruments.

To a certain extent, the results of this study could have been anticipated due to the fact that it has been proven before that there is a strong causal relationship between cognitive ability and training outcome.

The reason for this is that people with higher cognitive ability learn faster and master job knowledge quicker. This is not to say that selection should now only concentrate on finding superior trainees. The ideal would be to study the test results in respect of the cognitive ability component and then determine a cut-off point on each of them depending on who passes and who fails.

The fact that the data have limitations (restriction of range) has far reaching implications with regard to the calculation of cut-off points. The group that did not pass the training course is underrepresented and there is not enough data to discriminate statistically between them for the calculation of a cut-off point on the psychometric tests. The current cut-off points for the ARPM and SPEEX-indices were found to be suitable since the average scores for the different course modules are very high.

The extremely high scores might, however, indicate that the tests for the various course modules are too easy and this should be investigated. The data should be updated to allow for bigger datasets, which would have enough data in respect of unsuccessful candidates as well as to overcome the effects of restriction of range. In the interim the regression equation obtained from the multiple regression analysis can be used to calculate cut-off points. $Y = 0,530 \text{ cognitive ability} + 0,290 \text{ stress tolerance} + 8,992$

The findings of this study are in line with expectations that leaders should have a certain level of cognitive ability. However, one cannot assume that trainees that were successful on the junior leader training course will actually go on to be the best leaders. Training success is merely an indication of the trainee's ability to cope with the demands of junior leadership training. It is recommended that one should follow up on this group by measuring job performance sometime in the future. This would enable one to find the correlation between the selection battery and job performance. One could then also calculate the correlation between training outcome and job performance. This would indicate whether the training programme itself is valid.

A limitation of this study is that currently we assume that all six modules are equally important to become a junior leader.

It could, however, turn out that some of them are much more important and therefore should be weighted. Such an analysis will only be possible once job performance data become available.

REFERENCES

- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: MacMillan.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Borman, W.C., White, L.A., Pulakos, E.D. & Oppler, S.H. (1991). Models evaluating the effects of ratee ability, knowledge, proficiency, temperament, awards, and problem behavior on supervisory ratings. *Journal of Applied Psychology*, *76*, 863-872.
- Byars, L.L. & Rue, L.W. (1994). *Human resource management* (4th ed.) Burr Ridge, IL: Irwin.
- Campbell, J.P. (1991). Modelling the performance prediction problem in industrial and organisational psychology. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organisational psychology* (2nd ed. Vol. 1). Palo Alto, CA: Consulting Psychologists Press.
- Carpenter, P.A., Just, M.A. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*, 404-431.
- Cascio, W.F. (1987). *Applied psychology in personnel management* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Cascio, W.F. (1997). *Applied psychology in human resource management*. Englewood Cliffs, NJ: Prentice-Hall.
- Cliff, N. (1987). *Analyzing multivariate data*. Orlando, FL: Harcourt Brace Jovanovich Publishers.
- Cooper, D. & Robertson, I.T. (1995). *The psychology of personnel selection: A quality approach*. London: Routledge.
- Eckstein, S. (1998). Testing times. *People Dynamics*, *17* (6), 54-55.
- Employment Equity Act (55 of 1998). *Government Gazette*, *400*, No. 19370: Republic of South Africa, Pretoria.
- Erasmus, P.F. (2001). *Situation Specific Job Profiling and Assessment in the workplace of the 21st century*. Johannesburg: Potential Index Associates cc.
- Gardner, K.E. & Williams, A. P. O. (1973). A twenty-five year follow-up of an extended interview procedure in the Royal Navy. *Occupational Psychology*, *47*, 1-13 and 147-161.
- Gatewood, R.D. & Field, H.S. (1990). *Human resource selection* (2nd ed.). Orlando, FL: Holt, Rinehart & Winston.
- Ghiselli, E.E., Campbell, J.P. & Zedeck, S. (1981). *Measurement theory for the behavioural sciences*. San Francisco, CA: Freeman.
- Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. (1995). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Hall, J.C. (1957). Correlation of a modified form of Ravens Progressive Matrices (1938) with the Wechsler Adult Intelligence Scale. *Journal of Consulting Psychology*, *21*, 23-26.
- Hough, L.M. & Oswald, F.L. (2000). Personnel selection: Looking toward the future – Remembering the past. *Annual Review of Psychology*, *51*, 631-664.
- Howell, D.C. (1997). *Statistical Methods for Psychology* (4th ed.). London: Duxbury Press.
- Hughes, C.E. (1989). *The comparison and evaluation of various models of test fairness on samples of white and black employees*. Unpublished Masters thesis, University of Port Elizabeth, Port Elizabeth.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340-362.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98.

- Hunter, J.E., Schmidt, F.L. & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on bias in mental testing*, 41-99. New York: Plenum.
- Huysamen, G.K. (1996). *Psychological measurement. An introduction with South African examples*. Pretoria: Van Schaik.
- Kaplan, R.M. & Saccuzzo, D.P. (1997). *Psychological testing: principles, applications, and issues*. (4th ed.). Pacific Grove, CA: Brooks/Cole.
- Keir, G. (1949). The Progressive Matrices as applied to school children. *British Journal of Psychology*, 2, 140-150.
- Kriel, H. (2001) Horses for Courses: Situation Specific Selection Revisited. In P.F. Erasmus (Ed.), *Situation Specific Job Profiling and Assessment in the workplace of the 21st century*. Johannesburg: Potential Index Associates cc.
- Marshalek, B., Lohman, D.F. & Snow, R.E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Journal of Intelligence*, 7, 107-127.
- McDaniel, M.A. (1986). *The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance*. Doctoral dissertation, Washington, DC: George Washington University.
- McDaniel, M.A. & Schmidt, F.L. (1985). *A meta-analysis of the validity of training and experience ratings in personnel selection*. Washington, DC: U.S. Office of Personnel Management, Office of Staffing Policy.
- Milkovich, G.T. & Boudreau, J.W. (1994). *Human resource management* (7th ed.). Homewood, IL: Irwin.
- Murphy, K.R. (1988). Psychological measurement: Abilities and skills. In C.L. Cooper and I.T. Robertson (Eds.), *International Review of Industrial and Organisational Psychology* (Vol 3). Chichester: Wiley.
- Paul, S.M. (1985). The Advanced Ravens Progressive Matrices: Normative data for an American university population and an examination of the relationship with Spearman's "g". *Journal of Experimental Education*, 54, 95-100.
- Raven, J.C. (1960). *Guide to the Standard Progressive Matrices*, Sets A, B, C, D, and E. London: H.K. Lewis.
- Raven, J.C. (1965). *Advanced Progressive Matrices, Sets I, and II*. London: H.K. Lewis.
- Ritson, N.A. (1999). *The validity of a test battery used in the selection of apprentice electricians*. Master of Arts in Industrial Psychology, Durban: University of Natal.
- Schaap, P. (2001) The Reliability and Validity of the PIB. In P.F. Erasmus (Ed.), *Situation Specific Job Profiling and Assessment in the workplace of the 21st century*. Johannesburg: Potential Index Associates cc.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmitt, N. & Noe, R.A. (1986). Personnel selection and equal employment opportunity. In C.L. Cooper and I.T. Robertson (Eds.), *International Review of Industrial and Organisational Psychology*, (Vol 3). Chichester: Wiley.
- Schneider, B. & Schmitt, N. (1986). *Staffing organisations*. Glenview: Scott, Foresman.
- Snyderman, M. & Rothman, S. (1986). Science, politics, and the IQ controversy. *The Public Interest*, 83, 79-97.
- Society for Industrial Psychology (1992). *Guidelines for the validation and use of personnel selection procedures*. Psychological Association of South Africa.
- Tabachnick, B.G. & Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
- Wood, R. & Payne, T. (1998). *Competency based recruitment and selection: A practical guide*. Chichester: Wiley.
- Zeidner, J. & Drucker, A. J. (1988). *Behavioural science in the Army: a corporate history of the Army Research Institute*. Alexandria, VA: US Army Research Institute for the Behavioural and Social Sciences.