THE CONCURRENT VALIDITY OF LEARNING POTENTIAL AND PSYCHOMOTOR ABILITY MEASURES FOR THE SELECTION OF HAUL TRUCK OPERATORS IN AN OPEN-PIT MINE

by

MARIKIE KAREN PELSER

submitted in part fulfilment of the requirements

for the degree of

MASTER OF COMMERCE

in the subject

INDUSTRIAL AND ORGANISATIONAL PSYCHOLOGY

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: MR ZC BERGH

NOVEMBER 2002

ACKNOWLEDGEMENTS

I wish to thank the following people who made this dissertation possible:

My husband, Dean, who provided the love, support (and difficult research questions!) that kept me going – the eternal catalyst. To Lisa and Christopher, our children, who gave up substantial family time in order to make this research possible and, through being who they are, kept things in perspective. Home is wherever the three of you are.

My mother, Antoinette, for all the practical advice relating to the dissertation and, much more importantly, for providing the hands-on necessities (like grocery shopping) and luxuries (the never-ending supply of soup and home-made bread) that made this dissertation possible.

My father, Philip, for instilling in me the belief that I could do absolutely anything to which I applied myself and making the option of giving up, an unlikely one.

In memory of Maria Eales, my grandmother, who provided much of the impetus for effort and achievement in the family, whose wisdom, exuberance and energy are sorely missed.

Desiree Lawrence, for always being available to listen – mostly on the hands-free kit in poor reception areas!

Cas Coetzee, for living my research results with me. His tremendous insight, both statistically and emotionally, made a tremendous difference to my experience of this research.

Ziel Bergh, my supervisor, for his logical approach, practical advice and being a thoroughly nice person.

Iauma Cooper for language editing and critical reading.

SUMMARY

The purpose of the present study was to determine the concurrent validity of learning potential and psychomotor ability measures for the prediction of haul truck operator (N=128) performance in an open-pit mine. Specific aims were to determine the nature of the relationship between learning potential and psychomotor ability; whether there are higher order cognitive or psychomotor factors present in the combined use of the TRAM 1 and Vienna Test System measures; and the relative contribution of learning potential and psychomotor ability in the prediction of haul truck operator performance. The validity of learning potential and psychomotor ability measures was partially supported. A positive correlation between general (cognitive) ability (*g*) and psychomotor ability was reported. Factor analysis provided relatively consistent evidence for a general (cognitive) ability factor (*g*) underlying performance on all measures. The relative contribution of learning potential and psychomotor ability in the prediction of performance could not be established.

TABLE OF CONTENTS

CHAPTER 4

PSYCHOMOTOR ABILITY AND ITS RELATION TO INTELLIGENCE AS PREDICTOR OF PERFORMANCE

CHAPTER 5

METHODOLOGY

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND MOTIVATION FOR THE RESEARCH

1.1 INTRODUCTION

When skeptical South African line managers ask why they should invest time and effort in the pursual of valid personnel selection procedures, the answer can be put relatively simply:

- Firstly, to remain within the ambits of the law, which will be referred to as the legal incentive (Lopes, Roodt & Mauer, 2001; Mauer, 2000a, 2000b)
- Secondly, to maximize the probability of selecting the potentially most productive candidate(s), which will be referred to as the economic incentive (Schmidt & Hunter, 1981, 1998; Hunter & Hunter, 1984).

There are many sub-sections to both these incentives. The legal incentive has encompassed within it industrial relations, ethical, cultural, fairness and social utility of fairness considerations (Lopes et al, 2001; Mauer, 2000a, 2000b; Wheeler, 1993). The economic incentive encompasses issues such as shortages of skills, the impacts of globalization experienced in the South African economy and the need for increased competitiveness (Wheeler, 1993). These two incentives, however, provide a useful categorization of the benefits of, as well as the motivation for, the use of more scientific selection procedures in current-day South African organisations.

1.1.1 The legal incentive

The validity of psychological tests and other similar assessments used by an organisation to make any decisions affecting an individual's career status is a legal requirement as per the Employment Equity Act (55 of 1998), which stipulates that the use of psychological tests or any similar assessments is prohibited, unless they meet the requirements of being valid, reliable, fair and free from bias.

Although Lopes et al (2001, p. 61) would contend that the following statement involves an "excessively rigid interpretation" of section 8 of the Employment Equity Act, it is the researcher's view there is not much more to be said about validity in the context of the legal incentive than that it is a legal requirement.

Furthermore, this "excessively rigid interpretation" may be the preferred interpretation of those employees or potential employees who feel that the selection measures employed by an organisation, discriminate unfairly against them. Since the onus of proof of the fairness of a selection measure lies with the employer, organisations would be well advised to investigate and, where possible, to validate those tests and other similar assessments that are used to make career decisions about their employees or potential employees. In this way there will be scientific proof of compliance with the requirements set by the Employment Equity Act, which would make selection decisions based on these measures legally defensible.

Beyond the "big stick" of the law, there are distinct economic gains to be obtained by using selection procedures that are valid.

1.1.2 The economic incentive

The use of more valid selection procedures should lead to more valid and reliable predictions of applicant performance on the job, which should translate into increases in the productivity and performance level of the organisation, through the selection of the best candidate for the position (Anastasi, 1988; Hunter & Hunter, 1984; Schmidt & Hunter, 1981, 1998). This conception forms the foundation of studies concerning the utility of selection procedures. Utility studies aim to quantify the value add of selection procedures in economic (dollar or rand) terms.

Schmidt and Hunter (1998) reach the following conclusions pertaining to the utility of selection measures to the organisation:

- Compared to random selection, utility is directly proportional to the validity of the selection measures used.
- The potential economic value gains to be derived by replacing less valid measures with measures that have higher validity as predictors of performance are quite large in financial terms.

Thus the contention is that the higher the validity coefficient of a selection measure, the greater the economic utility in financial terms over the period in which the candidates ultimately selected, work for the organisation in question. The opposite is also true, namely the lower the validity coefficient of the selection measure, the greater the potential economic loss for the organisation in financial terms.

It thus makes economic sense for the organisation to investigate the validity of the various selection methods used in the organisation as well as those methods that are available to be used, in order to maximize the potential economic benefit that can be derived from selecting the candidates most likely to add value to the business of the organisation.

In order to understand the context of this study, it is necessary to move the focus of the discussion from the macro perspective of the need for the use of valid selection procedures in South African organisations to the micro perspective of the need for a valid selection procedure for the selection of haul truck operators in the open-pit platinum mine in which this study was completed.

1.2 MOTIVATION FOR THE RESEARCH

Minerals are South Africa's major source of foreign exchange earnings and it is anticipated that "mining will dictate the pace of economic development for many years to come" (Chamber of Mines, 2002, p.1).

As far as the platinum mining industry is concerned, South Africa holds 94% of the world's known Platinum Group Metals (PGM's), the only other major supplier being Russia (mostly from stockpiles). As from 1999, there have been significant increases in the demand for platinum, which is expected to be sustained in the short to medium term (Anglo American Platinum Corporation Limited, 2000).

Due to the overall shortfall in supply and the strong demand fundamentals for PGM's, there is growing competition between various platinum producers (both nationally and globally) to supply this shortfall at the largest profit margins possible. This translates into coming down the cost curve and increasing productivity at existing operations. It also implies the necessity for undertaking expansion projects (with the same focus on low cost and high productivity rates). This, in turn, translates into the need for valid selection procedures for both existing operations and expansion projects to ensure skilled, productive employees who can add value to the operations.

1.2.1 The role of the haul truck operator in an open-pit mine and the impact on productivity

In an open-pit platinum mine, the haul truck operator plays a major role in the overall productivity of the mine. At the operation in question, haul truck operators make up by far the largest contingent of operators (128 out of a total of 200) on the mine. Compared to the other types of production equipment on the operation, the fleet of haul trucks makes up by far the largest investment in terms of both capital outlay and running costs. A further consideration is that, relatively speaking, the safety risks are higher due to the fact that the haul trucks are mobile and capable of higher speeds than any of the other mobile production equipment in the pit.

1.2.1.1 The production process for haul truck operators

The production process for haul truck operators consists of the following components: loading, transporting and tipping.

The haul truck operators are responsible for transporting material (ore or waste) from the loading face to various tipping areas (crusher, stockpile or waste dumps) depending on the nature of the material to be tipped. They then tip the material into the crusher or onto a stockpile depending on the status of the crusher.

At the loading face, it is the responsibility of the haul truck operator to manoeuvre and place the truck into a position where it is ideal for the face shovel or front end loader to load the material. This procedure is referred to as "spotting in". In the case of the shovel, the haul truck must be positioned so that the shovel has no need to over-swing or overextend the "arm" of the shovel. (Overextending the "arm" leads to losses in productivity, because the time taken to swing is longer and also puts strain on the machine, resulting in potential damage to the shovel). In the case of the front end loader (which is on wheels), the ideal placement of the truck for loading has productivity and safety implications. If the haul truck operator does not place the haul truck optimally, the loader needs to reposition itself to tip with a full load in the bucket, which may result in the machine over-balancing with resultant injuries and property damage. As highlighted earlier, the material is then transported to either the crusher, the waste dump or the ore dump (stockpile).

At the crusher the operators are required to reverse the truck into the ideal position to tip. At the waste and ore dumps they need to reverse towards a safety berm in order to tip over the side of the dump. A safety berm is a heap of rock/sand of 1.5 metres high that is thrown on the edge of the dump. It has two main uses. Firstly the haul truck operator uses it to "spot into" the dump in terms of both location and direction and secondly, it signals to the haul truck operator when he or she has reversed too far, which may result in the truck falling down the side of a 30-40 metre dump resulting in injury (or even a fatality) and property damage.

The process of manoeuvring and placing the truck in the ideal position to tip is also referred to as "spotting in" - hence, "spotting in" is required for both the loading and the tipping components of the production process.

The technology on the haul trucks is advanced. On-board computers are linked to a dispatch system that monitors performance and indicates where and when the operator should load, dump, refuel and report for service. The operators interact with the dispatch system throughout the shift (by pressing various options on the communication touch screen in the haul truck) to indicate their current status, in order for the dispatch system to allocate trucks to the various loading points optimally.

Consequently, training involves not only the actual operation of the truck, but also focuses on safe and efficient "spotting in" in various environmental conditions, safety policy and procedures (which receives a lot of focus, from both an operational perspective and to comply with the legal requirements of the Mines Health and Safety Act (29 of 1996), and the mechanics of the dispatch system. Hence training periods are typically long and relatively expensive.

Bearing in mind that haul trucks are expensive, their running costs are high, their potential for both accidents and impact on overall productivity are significant, and that training costs are high in terms of both time and money, it is suggested that (as with the aviation industry) "even small improvements in identifying potential wash-outs could yield large cost avoidance savings" (Ree & Carretta, 1998, p. 82).

Ree and Carretta's (1998) statement highlights three possible main criteria against which the effectiveness of selection procedures can be measured, namely (1) productivity, (2) safety and (3) length of time taken to complete the training programme. Due to the limited scope of this research, the only criterion that will be covered in this study, is productivity.

1.2.2 The selection measures used to select haul truck operators

As is the case internationally (Campion, Pursell & Brown, 1988; Robertson, Gratton & Rout, 1990), unstructured interviews were traditionally used to select haul truck operators at this operation.

Based on the literature available (mostly in the aviation field), which seemed to indicate the validity of cognitive and psychomotor ability as predictors of both training and job performance criteria for positions requiring the operation and manoeuvring of mobile machinery, the TRAM 1 (as a measure of general cognitive ability, more specifically, learning potential) and the Vienna test system (as an assessment of psychomotor ability) were selected as assessment instruments for the selection of haul truck operators.

1.2.2.1 The predictors

In this study the TRAM 1 and Vienna Test System were the predictor variables.

TRAM 1 Learning Potential Test Battery

The TRAM 1 Learning Potential Test Battery is a learning potential measure which falls within the realms of the learning/dynamic theories of intelligence or cognitive ability. The distinguishing characteristic of a learning potential test is that testees learn a new skill or competency in the process of doing the tasks set out in the test. Some individuals become more competent than others and the differences in competency are captured in the test scores (Taylor, 1999).

Bearing in mind that the test is primarily presented in non-verbal diagrammatical format (except for the instructions), cultural bias that would be prevalent if candidates were to be required to respond to items in a second or third language, is limited to a degree (Lopes et al, 2001; Taylor, 1999).

There are validity studies available for the use of the TRAM 1 in the South African mining industry, but the criterion measures have always been academic outcomes. It has not been validated against job performance criterion measures, as was the case in this study.

The Vienna Test System

The Vienna Test System is a diagnostic instrument developed for the assessment of driving ability in Austria and is now utilised world-wide for the same purpose (Schuhfried, 1996). It is made up of a battery of mostly psychomotor tests, but also includes computerised versions of intelligence and personality tests. The fact that the system requires no prior driving experience to yield predictions of driving ability, makes this test ideal for use in the selection of potential haul truck operators.

Although there are various validity studies available for the different sub-tests of the Vienna Test System, they have been conducted in Europe. For the purposes of this study, five of the psychomotor tests were assessed for their validity as predictors of the performance of truck operators in a South African open-pit mine.

1.2.2.2 The criteria

Two objective and one subjective performance measures were utilised as criterion variables in this study. The objective criteria were the spotting times of the various operators into the loading equipment and the tons per work hour corrected for kilometres and gradient hauled. The subjective performance measure was obtained by the relative ranking per shift of each operator by the supervisor.

1.3 PROBLEM STATEMENT

Bearing in mind the legal incentive for the use of valid selection procedures in the South African context as well as the economic realities of

- the importance of productivity to the expanding platinum mining industry; and
- the potentially significant impact of the haul truck operator on the productivity of an open-pit platinum mine, which jointly constitute a significant economic incentive for the use of valid selection procedures, the problem statement in this research was as follows:

Are learning potential (as assessed by the TRAM 1) and psychomotor performance (as assessed by the Vienna Test System) valid predictors of haul truck operator performance in an open-pit mine?

1.4 AIM

The problem statement translated into the following general aim for this research:

To determine whether learning potential and psychomotor ability measures were valid predictors of haul truck operator performance in an open-pit mine.

1.4.1 Specific aims

The theoretical aims of this study were to

- Gain an understanding of the various theories on general (cognitive) ability and its relevance in personnel selection
- Gain an understanding of the various theories on psychomotor ability and its relevance in personnel selection
- Uncover the relationship between general (cognitive) ability and psychomotor ability according to the literature and previous research
- Determine how general cognitive ability and psychomotor ability interact to predict performance.

The empirical aims of this study were to determine

- Whether learning potential predicted haul truck operator performance
- Whether psychomotor ability predicted haul truck operator performance
- Whether moderator variables influence the relationship between learning potential, psychomotor ability and haul truck operator performance.
- Whether there is a relationship between learning potential and psychomotor ability
- Whether there are higher order factors at work in the combined use of the TRAM 1 and Vienna Test System measures applied in this research
- Whether the relative contribution of learning potential and psychomotor ability in the prediction of haul truck operator performance.

1.5 VALUE OF THE RESEARCH

This study would add value to the mining industry by investigating the validity of learning potential and psychomotor ability (more particularly the TRAM 1 and Vienna Test System measures) to predict haul truck operator performance. If the process were valid, this will constitute proof of such validity and thus confirmation of compliance with section 8 of the Employment Equity Act , (55 of 1998). The study would also provide support for the economic utility of using the TRAM 1 and Vienna Test System as selection measures. If not, it would highlight the need for further research to be completed in this arena in order to ensure both compliance with the Employment Equity Act and the harnessing of the economic gains potentially available, if more valid selection measures were to be used.

1.6 THE PARADIGM PERSPECTIVE

This study was conducted in the sub-disciplines of Personnel Psychology and Psychometrics in the discipline of Industrial Psychology.

The research paradigm perspective was positivism. Pure positivism assumes that only observable behaviour can be studied (Meyer, Moore & Viljoen, 1989). However, Mouton (1997) contends that positivism in the 20th century has relaxed its rigid empiricist criteria. According to him the quantitative methodological approach in twentieth century psychology can be regarded as positivist. Modern positivism differs from pure positivism in that modern positivism accepts that certain theoretical constructs need to be used. These constructs are operationally defined and various measures of the constructs (tests, indices etc.) are employed to measure behaviour. The data gathered is then subjected to statistical analysis. This study employed the modern positivistic paradigm in that measures of various constructs, namely the various sub-tests of the Vienna Test System and the TRAM 1 battery, were employed to measure behaviour and the data was then subjected to statistical analysis.

1.7 SUMMARY

Chapter 1 briefly introduced the study and the motivation for it, formulated the problem statement and stated the aims of the study. The potential value of the research was also discussed.

1.8 PRESENTATION OF THE STUDY

To achieve the theoretical aims of this study, chapter 2 discusses the literature review pertaining to the role of validity in personnel selection, while chapters 3 and 4 discuss the literature pertaining to general (cognitive) ability and psychomotor ability as predictors in personnel selection. In order to achieve the empirical aims of the research, chapter 5 describes the methods used in this study. Chapter 6 covers the results of the study and discusses them by integrating relevant theory and research. The conclusion, a discussion of the limitations of the research and recommendations, then follow.

CHAPTER 2

THE ROLE OF VALIDITY IN PERSONNEL SELECTION

2.1 INTRODUCTION

Although man has been selecting people for different positions, groups and tasks for centuries, personnel selection as a scientific discipline dates back merely 100 years, beginning in the early part of the twentieth century, with the development of assessment methods and procedures, the start of criterion-related validity studies and the gradual development of psychology into the applied science that it is today (Salgado, 2000).

The rationale for scientific personnel selection is to be found in two common sense realities:

- Individuals differ in terms of their abilities, knowledge, interests and personalities (Anastasi, 1988; Cascio, 1991; Aiken, 1997).
- Jobs differ in terms of the skills and human qualities required to get the job done (Cascio, 1991; Dunnette, 1966).

Scientific personnel selection assumes that these differences between individuals, on the one hand, and jobs, on the other, can be measured in some way (Aiken, 1997). The organisation can then capitalise on individual differences by selecting those candidates who possess the greatest number of qualities judged to be important for success in any particular job (Cascio, 1991; Dunnette, 1966). This should lead to a more productive organisation.

It is important to note that personnel selection is successful and hence, useful, only to the extent that the measurements of the differences between individuals and job requirements referred to are accurate or valid. If they are not valid, they are useless as predictors of performance, on the one hand, and as criterion measures on the other - hence, the extreme significance of validity in the selection process.

This chapter discusses the validation process, focusing on its relevance to the organisation, using the incentive framework created in chapter 1, which is

- to remain within the ambits of the law which will be referred to as the legal incentive

- to maximise productivity which will be referred to as the economic incentive

Various predictors (eg. intelligence tests, psychomotor tests, interviews, qualifications or experience) that can be used for the prediction of job success as well as how they differ in terms of their validity for various positions, are also discussed.

2.2 THE VALIDATION PROCESS

As mentioned in chapter 1, in order to determine how valid a selection procedure is, a validation study needs to be undertaken. Such a study involves various activities, which will differ in timing and methodology, depending on whether a predictive or concurrent design is used (Anastasi, 1988; Cascio, 1991). These differences are discussed in more detail in section 2.2.1.2.

The validation of selection measures generally involves the following activities:

- job analysis
- choice and development of predictors and criteria
- obtaining predictor data from all candidates in a sample
- obtaining criterion data from all candidates, preferably, in the same sample
- determining whether there are predictor-criterion relationships (testing the hypotheses)
- implementing the selection procedure if it has statistical validity.

2.2.1 Evaluating the predictor

In essence, the two main ways in which predictors can be evaluated are in terms of their reliability and validity for the purpose for which they are to be utilised.

2.2.1.1 Reliability

Reliability refers to the degree of consistency, dependability or stability of measurement or a measure used in selection research. The reliability coefficient can be seen as the extent to which scores on a particular measuring instrument are due to "true" differences in the attribute measured.

There are different ways in which reliability can be assessed, the main ones being test-retest reliability, parallel/alternate form reliability, split half reliability and Kuder-Richardson reliability/Coefficient alpha (Anastasi, 1988). Latham and Wexley (1981) contend that a reliability coefficient of 0,80 is viewed as acceptable.

Cascio (1991) maintains that high reliability is a necessary but not sufficient condition for validity; in other words a selection measure's validity is limited by the extent to which it is unreliable.

2.2.1.2 Validity

Validity revolves around two issues (Anastasi, 1988; Cascio, 1991):

- what the test measures (the underlying construct)
- how well the test measures it (the relation between the predictor and criterion variables).

These two validity issues were used as a point of reference in discussing validity.

Validity is traditionally seen to be the extent to which a measurement procedure measures what it is designed to measure (Anastasi, 1988; Kaplan & Saccuzzo, 1997; Kerlinger, 1986; Walsh & Betz, 1990).

Cascio (1991) contends that this is inadequate and that numerous different investigations are required to really understand the interrelationships between scores. Similarly, Binning and Barret (1989) maintain that there is no such thing as different types of validity. Instead, they contend, there are different types of evidence for the validity of the same measure. Validity always refers to the degree to which the evidence supports *inferences* that are made from the scores obtained in the measure. It is these inferences rather than the procedure that is validated in a validation exercise (Cascio, 1991). In other words, a validation study is an attempt to determine to what extent it would be correct to make inferences about individuals' anticipated behaviour from the predictors and the job performance criteria used.

It should be noted that job analysis is central to any validation study (Anastasi, 1988; Society for Industrial Psychology, 1992). It is necessary to understand what it is that the job requires a candidate to be able to *do*. In order to determine which predictors and criterion measures are relevant in a validation study, various interdependent strategies available to validate these inferences made from a candidate's scores on a measure:

Content-related and construct-related evidence

As mentioned, validity revolves around two issues (Anastasi, 1988; Cascio, 1991), namely (1) what the test measures (the underlying construct) and (2) how well the test measures the construct (the relation between the predictor and criterion variables).

Content-related and construct-related evidence relate to the first of these issues, namely what the test measures.

Content-related evidence is about assessing whether the measure contains a fair sample of the universe of situations or behaviours it is supposed to represent (Anastasi, 1988; Cascio, 1991). Firstly, it indicates whether the test covers a representative sample of the required behaviours, skills or knowledge and secondly, whether the test performance is relatively free from the influence of variables that are not relevant to the behaviours/situations it is supposed to represent.

Construct-related evidence is about assessing whether the test is measuring the theoretical construct or trait that it professes to measure. Its main objective is to understand the construct which the assessment tool intends to measure. The focus is on the conceptual meaning of the construct, its distinction from other constructs and how the measures of the construct should relate to other variables (Cascio, 1991; Guion, 1987).

Criterion-related evidence

This type of evidence relates to the second validity issue, namely how well the assessment tool used measures the relation between the predictor and criterion variables (Anastasi, 1988; Cascio, 1991).

Guion (1974) defines criterion-related validity as the extent to which scores on the predictor variable may be used to infer performance on a criterion variable.

Whenever measures are used to predict or explain behaviour – as in the case of most measures used for the selection of employees - criterion-related evidence is called for. It is necessary to test the hypothesis that the predictor measure data (for example, test data) is related to performance on some criterion measure (for example, performance appraisal data), in order to determine the practical validity of the predictor for a specific purpose (Anastasi, 1988).

There are two designs available to assess the criterion-related evidence for validity namely, predictive and concurrent designs. These designs differ in three main criteria, namely the timing of criterion data availability, the underlying purpose for the completion of the validity study and the degree of inference required for interpretation.

- *Timing of criterion data availability*

Usually the criterion measure is either currently available, in which case a concurrent validity design is possible, or will become available at some time in the future, in which case a predictive design is called for. But although concurrent validity is often substituted for predictive validity (Anastasi, 1988), the distinction between concurrent and predictive designs is not only a matter of when the criterion data is available (Cascio, 1991).

- *Purpose*

Anastasi (1988) contends that the validation design to be used in a particular validation study depends on the purpose for which the test (predictor) is to be used.

The predictive design is intended for forecasting or predicting some specific future outcome (Aiken, 1997; Walsh, 1989; Walsh & Betz, 1990). The concurrent design, on the other hand, refers to the correlation between predictor data (eg. test scores) and *current* performance measures (Aiken, 1997; Anastasi, 1988; Walsh & Betz, 1990). It is relevant to instruments employed for the diagnosis of existing status rather than for the prediction of future outcomes.

- *Degree of inference required for interpretation*

In predictive designs the relationship between predictor data and criterion data is readily apparent – in other words, the predictor (for example test) performance for the individual predicts x performance and the criterion measure for the individual indicates actual y performance, therefore the predictor is z valid. This is a relatively straightforward deduction.

In concurrent designs, however, the relationship needs to be inferred by the decision maker; that is, in the same way as for predictive designs, the predictor (for example test) performance for the individual predicts *x* performance and the criterion measure for the individual indicates actual *y* performance. In the predictive design, the deduction then is that the predictor is z valid. However, in concurrent designs, it can only be deducted that this may mean that for future applicants the predictor could be *z* valid, if these candidates are similar to those who were involved in the concurrent study in terms of variables such as, amongst others, experience and skill levels (Cascio, 1991). Hence, the deductions that are made from a concurrent study have conditions or "if" statements attached to them and are therefore less direct than the deductions that can be made from predictive designs.

For tests of cognitive ability, studies have found estimates of validity for concurrent and predictive designs to be comparable (Anastasi, 1988; Guion, 1987; Society for Industrial Psychology, 1992). This cannot be assumed for other predictors and criterion measures unless there is research evidence supporting such comparability.

- *Constraints of the concurrent design*

The following are constraints of the concurrent design (leading to greater inference being required from the researcher):

- Candidates would have come into the study with different levels of experience, impacting on their scores on the predictors and hence confounding the relation between predictor and criterion (Anastasi, 1988; Cascio, 1991).
- The motivation levels of the candidates used in the study are more likely to be a moderating variable in the validity exercise.
- Attrition over time shrinks the sample size and thus reduces the variability, effectively lowering validity coefficients (Anastasi, 1988; Cascio, 1991).
- Jennings (1953) as cited in Cascio (1991) states that employees who are secure in their jobs and realise that their test scores will not affect their job standing are not as motivated to perform well in assessments as applicants for the same job. This may also be relevant to performance on the criterion measure.

Finally, the Society for Industrial Psychology (1992) suggests the following guidelines for the selection of predictors over and above the obvious reliability and validity issues:

- Predictors should be chosen on the basis that there is a logical, empirical or theoretical reason for them to be included.
- Predictors should be selected based on scientific knowledge rather than expedience.
- Predictors should be as objective as possible.

2.2.2 Evaluating the criterion

For the reason that the main endeavour of psychologists and line management alike, is the identification of measures or predictors to predict performance, the criterion used to measure this performance is often neglected in validation studies. A predictor measure can be no better than the criterion used to establish its validity (Cascio, 1991; Binning and Barret, 1989). The criterion problem, that is, the availability, reliability and validity of criterion data, is extensively referred to in the literature in the field of validation (Anastasi, 1988; Cascio, 1991; Thorndike, Cunningham, Thorndike & Hagen, 1991; Walsh, 1989).

When using job performance as criterion measure, either objective or subjective measures, both of which have their limitations, may be used (Cascio, 1991).

2.2.2.1 Objective criterion measures

Objective criterion measures (such as sales achieved by a sales representative or tons hauled by a haul truck operator) are very appealing since they eliminate the subjectivity that goes with the territory of supervisor ratings and also eliminates possible bias (Thorndike et al, 1991). Yet, it has its limitations, including performance unreliability (Cascio, 1991; Deary, 2001; Shinar, 1978) and modification of performance by situational variables that are often beyond the control of the candidate being assessed. Furthermore, the focus of objective data is not on behaviour itself, but on the results of behaviour (Cascio, 1991; Thorndike et al, 1991). Cascio (1991) contends that although objective criterion measures can add value as supplementary judgements, correlations between objective criterion measures and subjective ratings are often low.

2.2.2.2 Subjective criterion measures

Subjective measures (such as supervisor rankings and behaviourally anchored rating scales), on the other hand, are equally prone to error (eg, halo effect, central tendency, leniency and severity errors), since they rely on human judgement (Anastasi, 1988; Cascio, 1991; Thorndike et al, 1991). Guion (1987) laments that it would be wonderful to be able to predict a trend away from the use of supervisor ratings in personnel research for this very reason, but holds no such hope. Anastasi (1988), however, maintains that although judgemental errors are a pitfall for the use of performance ratings, when conditions are carefully controlled, they nevertheless represent a valuable source of criterion data.

2.2.2.3 The criterion evaluation process

In essence, it should be borne in mind that all criterion measures measure only part of the success or non-success of candidates in their jobs. The ultimate criterion would be some measure of lifetime success in the profession or job for which the candidate is being assessed (Thorndike et al, 1991). This is not a realistic criterion for day-to-day validity studies and the researcher is ultimately faced with having to select those criteria which are most satisfactory. To a certain extent, it is about selecting the final criterion carefully, being aware of its limitations and making an effort to minimise error.

How can criteria be assessed to determine their suitability? What makes for satisfactory criteria? The qualities desired in a criterion measure are relevance (the most important factor), freedom from bias, reliability, availability (Thorndike et al, 1991) and freedom from contamination (Cascio, 1991; Anastasi, 1988; Society for Industrial Psychology, 1992).

A criterion is deemed to be relevant if candidates' performance on the criterion exemplifies their performance on the job – the determination of relevance is dependent on the researcher's professional judgement. In some way it is similar to its content validity (Thorndike et al, 1991).

It is important that the criterion is free from bias. It should allow for candidates who are equally good at the job to obtain the same score, regardless of the group to which they belong (Thorndike et al, 1991).

A criterion needs to demonstrate a certain degree of reliability in that a measure that is totally unstable cannot be predicted (Thorndike et al, 1991). The Society for Industrial Psychology (1992) suggests that it is desirable (but not essential) for criterion measures to be highly reliable. The reliability of criteria and predictors place a ceiling on the overall validity that can be obtained (Carretta & Ree, 2000; Cascio, 1991; Schmidt, Hunter & Urry, 1976).Thus, the true effect of unreliability in a criterion will be to underestimate the validity of the measure being investigated. Criterion unreliability can be corrected for statistically, also referred to as correction for attenuation (Carretta & Ree, 2000, 2001; Society for Industrial Psychology, 1992).

The fact that a criterion needs to be available seems to be a statement of the obvious. The highlighting of availability as one of the requirements of a criterion emphasizes the practical difficulty that the collection of criterion data often brings to validation research. Criteria are available in as far as they are practically feasible for use in validation studies. Practical feasibility, in turn, is partially dictated by the limits that the organisation places on the time and money it is prepared to invest (Thorndike et al, 1991).

The possibility of contamination should be considered. A measure is seen to be contaminated if it includes unwanted systematic variance. The Society for Industrial Psychology (1992) suggests that it is impossible to avoid (or even to know) completely all sources of contamination, but that the researcher should attempt to minimise its effects. In order to minimise contamination, it is essential for predictor and criterion data to be gathered independently. Individuals involved with the provision of criterion (or performance) data should thus have no access to predictor data, such as test scores (Anastasi, 1988; Cascio, 1991; Society for Industrial Psychology, 1992).

2.2.3 Establishing the relationship between the predictors and the criteria

In order to determine the relationship between the predictors and the criteria, correlation or partial correlation techniques are usually completed to confirm the presence of a relationship as well as to determine the nature of such a relationship.

A correlation coefficient varies in value from -1 (a perfect negative value) to + 1 (a perfect positive correlation). Values close to 0 indicate no linear correlation. Correlations are never perfect so that +1 and -1 are never achieved. Suppose a positive correlation of 0,5 between variable X and Y is found. This means that the higher a person's score on X, the higher that same person's score will be on Y. Or conversely, the lower a person's score on X, the lower that person's score on Y. If the correlation is negative, however, for example -0,5, then the higher a person's score on X, the lower that same person's same score is likely to be on Y. Or, conversely, the lower a person's score on X, the higher that person's score on Y (Huysamen, 1987). Anastasi and Urbina (1997) indicate that correlations as low as 0,20 or

0,30 between predictors and performance criteria are acceptable to warrant their inclusion in selection batteries.

2.2.4 Evaluating the sample and data analysis

The Society for Industrial Psychology (1992) suggests the following guidelines when the choice of sample is being considered:

- The composition of the sample and its relevance to the study need to be considered. Samples are very seldom homogenous and may differ from the ultimate applicant group in specific ways (eg, age, race, etc.). The researcher should rely on the research literature in making professional judgements about whether these differences are relevant. No variable can be assumed to moderate validity unless there is specific evidence to that effect (see also section 2.2.5 referring to moderator variables).
- The sample should be large enough to provide adequate statistical power. Validation exercises also often have as a limitation the constraint that there are not enough candidates doing a particular job in a particular organisation to obtain significant results and scientifically conclude whether there is evidence for predictor-criterion relationships (Anastasi, 1988). This may lead to Type 1 error (i.e, finding a relationship when, in reality, there is none) or a Type 2 error (i.e, finding no relationship when, in reality, there is a relationship). The Society for Industrial Psychology (1992) suggests that this type of error is often underemphasized and that Type 1 and Type 2 errors should receive equal attention. Type 2 errors often occur due to the design not having adequate statistical power (Carretta & Ree, 2001; Schmidt et al, 1976). An extremely large sample or replication is necessary to corroborate validity findings for more complex multivariate studies, involving many predictors and criteria. This often requires cross-validation studies.

Another relevant factor in validation studies relating to the sample is range restriction (Anastasi, 1988; Cascio 1991), also referred to as censored samples (Carretta & Ree, 2000). Since pure predictive designs can very seldom be used practically (Anastasi, 1988) and pre-selection is usually a reality (since all applicants for a position cannot be appointed and followed up on), range restriction is usually prevalent in both the predictor and the criterion, reducing the potential variability. Statistical corrections for range restriction are available and should be applied (Lawley, 1943; Thorndike as cited in Carretta & Ree, 2000).

2.2.5 Evaluating the moderator variables

The characteristics of the group of individuals on whom a test is validated include variables such as sex, age, levels of education, interest or motivation and personality traits. These factors, referred to as moderator variables (Aiken, 1997; Anastasi, 1988; Cascio, 1991; Society for Industrial Psychology, 1992), can affect the correlation between a test and criterion measure. For example, if an applicant has little interest in a job (i.e, demonstrates low motivation levels), he or she may perform poorly either on the predictor or the criterion or both, wreaking havoc with the correlations that are ultimately found. This may lead to the discarding of a predictor which may potentially have added great value to the organisation in question (within both the legal and the economic incentive frameworks referred to in chapter 1).

Moderator variables cannot merely be assumed, but must be supported by the research literature (Anastasi, 1988; Society for Industrial Psychology, 1992). Moderator variables have value in promoting the understanding of the dynamics of individual behaviour and in suggesting hypotheses that should be investigated with proper methodological controls.

Next, the relevance of validity to personnel selection will be examined.

2.3 THE RELEVANCE OF VALIDITY TO PERSONNEL SELECTION.

2.3.1 The legal incentive

The focus on the valid and fair selection of employees has increased considerably in "the New South Africa", which dawned after the country's first democratic elections in 1994 (Lopes et al, 2001; Mauer, 2000a, 2000b). This increased focus has mainly been the result of extensive changes in legislation in the post-1994 era, which were not unanticipated by industry (Taylor, 1992; Veldsman as cited in Wheeler, 1993; Van Wyk as cited in Wheeler, 1993).

Some of the relevant legislation promulgated post-1994 have been the Constitution of the Republic of South Africa (Act 108 of 1996), the Labour Relations Act (66 of 1995) and the Employment Equity Act (55 of 1998). Briefly, the Constitution of the Republic of South Africa (Act 108 of 1996) is the supreme law of the country and contains the Bill of Rights which enshrines specific human rights. Certain human rights have been allocated the additional status of non-derogatory rights. Equality and human dignity are regarded as non-derogatory rights. These have obvious (if indirect) implications for the selection processes utilised by employers (Mauer, 2000b).

The implications of the Labour Relations Act (66 of 1995) for these selection processes is more direct in that the Act elaborates on the principles of unfair discrimination and unfair labour practices within the employment relationship.

Yet, it is through the Employment Equity Act (55 of 1998) that the impact of legislation on selection processes used in the employment context is most directly felt. Although prior to the 1994 election, it was anticipated that unfair selection procedures resulting in discriminatory effects would be more critically scrutinized in "the New South Africa" (Giliomee, 1989; Veldsman as cited in Wheeler, 1993), it is doubtful whether the directness and specificity of the Employment Equity Act was anticipated by industry and the psychological fraternity alike. More particularly, the direct prohibition of the use of psychological testing and other similar devices unless certain very specific scientific criteria were met (Mauer, 2000b) caught employers off-guard, leading to increasing concern as to whether or not their selection processes were, in fact, legitimate (Lopes et al, 2001).

Mauer (2000b) sketches some of the background to the development of the Act. In the earlier drafts, psychological testing was prohibited outright and a large amount of lobbying was apparently necessary by psychological interest groups to convince the Parliamentary Portfolio Committee of Labour that it was not in the best interests of employer, employee or the country as a whole to go the outright prohibition route. Rather, it was argued, the unfair use of psychological testing needed to be eradicated. It was further argued that very precise scientific requirements had to be met to prove that psychological testing was, in fact, being used in a fair manner (commensurate with international best practice in terms of assessment and ethics). Furthermore, it was suggested that stringent guidelines in terms of the psychometric requirements of psychological tests had to be enshrined in law.

Before the finalisation of the Act, it became apparent that some employers merely dropped their psychometric assessment batteries and reverted to unstructured interviews. Another ploy was to assert that certain assessment devices that were clearly psychological were, in fact, not psychological tests, because they were not marketed by psychologists and did not contain terminology pertaining to psychology. The danger at this stage was that the very Act developed to limit unfair discrimination, may actually have led to the situation where unfair discrimination was even more likely to be prevalent (Mauer 2000b). The Portfolio Committee then decided that all instruments used to make decisions about individuals' careers, had to be regulated by the Act, hence the inclusion of the phrase "*and other similar assessments*" in Section 8, which, in the final version, reads:

> *Psychological testing and other similar assessments of an employee are prohibited unless the test or assessment being used*
>
> *(a) has been scientifically shown to be valid and reliable*
>
> *(b) can be applied fairly to all employees*
>
> *(c) is not biased against any employee or group.*

It is in this particular phrase, "*and other similar assessments*" where the enormity of the impact of the Act on the selection processes used in the employment context is felt. Not only do psychological tests need to comply with psychometric theory, but all devices used for the purpose of gaining information about a person and to make any decision about his or her employment status (such as appointment, rejection, promotion, further development, reward systems and bonuses), must comply with the measurement theory requirements of, amongst other things, reliability, validity and freedom from bias (Mauer, 2000a). Fundamentally, this reduces to the fact that all selection devices should comply with the requirements of section 8 of the Employment Equity Act.

Looking at the research on how selection decisions are typically taken, instruments not traditionally viewed as psychological instruments are still the main source of information for selection decisions. This is especially true for the interview as a selection instrument (Campion et al, 1988; Robertson et al, 1990).

Robertson et al (1990, p. 69) expand on the above problem by emphasising that the concerns expressed about the validities of interviews in the literature have had little impact on the extent of their use in the selection process. According to them surveys by ASSPA and Robertson and Makin have revealed "the unsurprising fact" that interviews are the most widely used selection instrument, despite the fact that it has been known for decades that the interview as selection device generally lacks reliability and validity (Guion, 1987; Latham et Saari, 1984). Schmidt and Hunter (1998) also point out that many organisations throughout the world are using sub-optimal selection procedures. Specifically they contend that many organisations in the United States rely solely on the unstructured interview, when there are more valid methods available. Although there is little research available on the topic, it is postulated that the situation in South African organisations is very similar.

Furthermore, it is suggested that a very small percentage of South African employers are developing their selection procedures (mostly interviews) in a scientifically rigorous way that will comply with the Employment Equity Act, section 8 requirements. Bearing in mind that the burden of proof of compliance with the Act lies with the employer, it would be necessary for the organisation to scrutinise its selection procedures and to prove scientifically that the measures are reliable, valid for the purpose for which they are being used and free from bias. This involves the completion of validation studies such as the one to be undertaken in this research.

2.3.2 The economic incentive

Having focused on the law, as is the trend in current-day South African research on the topic, (Lopes et al, 2001, Mauer, 2000a, 2000b), there is a very real risk that the true motivation for the pursual of more scientific selection procedures is lost in the noise of the use of the "big stick" of the law.

The purpose of pursuing fair and valid selection techniques is to maximise the probability of selecting the potentially most productive candidate for the job. The improvement of selection measures and procedures should lead to more valid and reliable predictions of applicant performance on the job. This, in turn, should translate into increases in productivity through the selection of the best candidate for the position (Anastasi, 1988).

The new world of work makes massive demands on both the indivual and the organisation in terms of the ever-increasing pace made possible by new technology, the increased competitiveness brought about by globalisation and the increased complexity of the work to be done and the workplace within which the work needs to be done (Drucker, 1992; Luthans, 1995; Schermerhorn, Hunt & Osborn, 2000). Seen in this light, optimising the employment match through effective selection methods becomes ever more important. As Woodruffe (1990, p. 8) points out,

> *The mistake might be in accepting people who are not suitable, but nowadays the more important mistake for many organisations is in rejecting people who could have made a valuable contribution. In an era of shortages of talent and experience, it is critical to recognise people who could contribute to an organisation's success.*

Studies on the utility of valid selection procedures (Hunter & Hunter, 1984; Schmidt & Hunter 1981; 1998; Schmidt, Hunter, McKenzie & Muldrow, 1979) substantiate the proposal that the use of valid selection tests substantially increases the performance level of the resultant workforce and thus increases productivity. These studies have focused on quantifying the productivity gains derived from the use of valid selection methods in dollar terms. Although the mechanics of the process is beyond the scope of this study, suffice it to say that, compared to random selection, utility (or gains in dollar or in the South African context, rand terms) has been shown to be directly proportional to the validity of the selection methods (Schmidt & Hunter, 1998) – a very clear economic incentive for South African organisations to pursue more valid selection procedures.

In summary thus far, the following arguments has been put forward:

- There are both legal and economic incentives for South African organisations to pursue fair and valid selection procedures.
- In order to meet the legal requirements of the Employment Equity Act (55 of 1998), measures used for personnel selection have to be valid, reliable, fair and free from bias (however, due to its limited scope, only validity will be dealt with in this study).
- It is postulated that the interview is still the main selection device in South Africa.
- It is doubtful whether these interviews are constructed in a scientifically rigorous manner that complies with the Employment Equity Act.
- Selection procedures that are more valid (and hence, more likely to yield results that are correct) need to be pursued.
- In order to investigate the validity of a selection procedure scientifically, validation studies need to be undertaken.

2.4 METHODS OF SELECTION AND THEIR VALIDITY

Many selection methods are available to select candidates for different positions. In this section, the focus is on methods that may lend themselves to the selection of lower level employees at entry level for positions requiring the acquisition of operating skills for mobile machinery.

2.4.1 Intelligence measures

There is extensive support in the literature for the contention that general (cognitive) ability alone predicts performance well across a variety of jobs (Hunter, & Schmidt, 1996; Levine, Spector, Menon, Narayanan & Cannon-Bowers, 1996; Ree & Earles, 1992; Schmidt & Hunter, 1998). In a meta-analytic study covering 19 selection methods spanning over 85 years research on selection procedures, Schmidt and Hunter (1998) report the following findings for the use of general mental ability (GMA), referred to in this study as general (cognitive) ability or *g:*

- General (cognitive) ability or *g,* used on its own, has an average validity coefficient of 0,51.
- General (cognitive) ability or *g* demonstrates the highest validity coefficient and is available at the lowest application cost, due to its measures being amenable to large group testing and its administration and scoring procedures being relatively straightforward.
- Measures of general (cognitive) ability or *g* are flexible in that they can be used for all jobs, inclusive of entry level positions where candidates do not have any experience. Due to the fact that general (cognitive) ability is also the best predictor for job-related learning, it is particularly good for positions where the applicant pool has no experience in the job for which selection is being done and will still have to learn skills.
- The research evidence supporting the validity of general (cognitive) ability or *g* in selection settings by far outweighs the evidence for any of the other selection measures (eg, interviews, work sample tests or personality measures).
- The theoretical foundation for general (cognitive) ability or *g* is stronger than for any other selection measure, bearing in mind that there have been literally thousands of studies investigating its validity over the last century. Due to the volume of research completed in this field, the understanding of what the relevance of intelligence is to the selection process (and what is measured), is much clearer than for some of the other selection measures.

Schmidt and Hunter (1998) are of the opinion that intelligence should be the main selection measure and that there is a need to assess how much additional utility any other additional selection measures will add to the accurate prediction of performance.

2.4.2 Psychomotor ability measures

Psychomotor tests are typically apparatus tests focusing on speed, coordination and other characteristics of movement responses required for job performance, such as, manual dexterity and leg and foot movements required for a specific occupation (Carretta & Ree, 2000; Fleischman, 1988).

Generally speaking, psychomotor tests are relevant to those positions in which motor skills are relevant. They have traditionally been designed for specific occupations and usually rely on the principle of simulation (Duke & Ree, 1996).

Although, most studies in the psychomotor arena involve aviation research (Anastasi, 1988; Griffin & Koonce, 1996), research on the validity of psychomotor ability as predictor of performance has also been undertaken in the industrial and military fields, including automobile mechanics, craft workers, police and fire fighters, sewing machine operators, soap packers and several US army enlisted jobs (Carretta & Ree, 2000). South African studies are also available for train drivers (Schoeman, 1995); operators of mobile machinery at a diamond mine (De Jager & Van der Walt, 1993) and a coal mine (De Jager & Van der Walt, 1997); heavy duty truck drivers (Bouwer, 1985; De Jager, 1997); as well as for drivers in the national defence force (Oosthuizen, 1975).

In general research findings have been supportive of the validity of psychomotor tests (Carretta & Ree, 2000; Wheeler & Ree, 1997; Martinussen, 1996, Schoeman, 1995). The research also seems to point to the fact that as job complexity decreases, the validity of psychomotor skills increases (Levine et al, 1996; Hartigan & Wigdor, 1989; Hunter & Hunter, 1984).

2.4.3 Employment interviews

The interview as selection measure is the measure that is still the most widely used for employee selection, mostly due to its low cost and expedience (Robertson et al, 1990). The validity of interviews in the personnel selection field, is inclined to vary enormously, depending on the quality of both the particular interviewing instrument and the skills of the interviewers.

Hunter and Hunter (1984), report a validity coefficient of only 0,14 in their meta-analysis of validity studies pertaining to interviews. Latham and Saari (1984) contend that low validity findings should not be surprising, since it has been common knowledge since the early fifties that the interview as a selection device generally lacks reliability and validity. They summarize rather well, the reasons for the interview's low validity:

- The interviewees are not asked the same questions.
- When they are asked the same questions, these questions are usually not related to the job.

- When the questions are job-related, the interviewers often cannot agree on what constitutes an acceptable answer.
- When they do agree, the acceptable answer to the question is frequently transparent to the interviewee.

However, in their meta-analysis McDaniel, Whetzel, Schmidt & Mauer (1994) found validity coefficients that were slightly more encouraging, namely, 0,51 for the structured interview and 0,38 for the unstructured interview. Schmidt and Hunter (1998) contend that these validities will definitely be lower in carelessly constructed and conducted interviews. Structured interviews, although more costly, demonstrate significant gains in validity and hence utility over unstructured interviews (Schmidt & Hunter, 1998). In fact, when combined with general (cognitive) ability or *g*, carefully constructed structured interviews yield one of the highest validities for the prediction of job success namely,. a combined validity coefficient of 0,65 (Schmidt & Hunter, 1998).

In summary, the validity of the interview seems to be contingent on the instrument used and the degree to which it is structured, the skill of the interviewers and the requirements of the job for which the interview is being done. Hence, great variability is found in the validity coefficients obtained in various validity studies. Guion (1987, p. 200) summarizes the status quo of interviews in selection research particularly well by stating that "interviews have always been with us, are still with us, and will continue to be with us – even though they have been known for decades to be too unreliable to be valid".

2.4.4 Job try out

It can be argued that it is common sense that the best predictor of future job performance is current job performance. Hunter & Hunter (1984), for example, found validity coefficients of 0,44.

However, trying out candidates on the job for a particular position is very expensive (Schmidt & Hunter, 1998). For example, the training costs for operators of expensive production equipment, are likely to be expensive in terms of direct cost of training as well as the potential cost of errors made (Martinussen, 1996). Furthermore the risk and consequential opportunity cost of not having selected the potentially most productive candidate (Woodruffe, 1990) due to the fact that the selection pool is so much smaller, should not be overlooked.

A further downfall of this method is that very often supervisors are reluctant to get rid of non-ideal performers, thus losing the purpose as well as the potential gains of the selection method (Schmidt & Hunter, 1998). Legally, in the South African context, the situation is exacerbated by the Labour Relations Act (66 of 1995) which affords the applicant in the try out phase, employee status. Although this effect has been diminished slightly by amendments to the Act in 2002, it amounts to near-random appointment of employees, further limiting the practical utility of this method.

2.5 POSSIBLE MODERATOR VARIABLES

In section 2.2.5, it was emphasized that moderator variables cannot merely be assumed. They have to be supported by research findings before they can be included in a validity study. The following section discusses experience, age and years of education as possible moderators of job performance in lower level entry positions requiring operating skills.

2.5.1 Experience

The research on this topic yields different results depending on the job and the number of years of experience prevailing in the position in question. Schmidt and Hunter (1998) report a very small overall validity coefficient (0,18).

In specific situations, though slightly larger validity coefficients have been reported. Martinussen (1996), reported a mean correlation coefficient of 0,25 for training experience with a global pilot training performance criterion in a meta-analysis of research in the aviation field – in fact this is the highest correlation coefficient in the meta-analysis, and is one of the predictors that has been built into the Pilot Candidate Selection Method (PCSM) algorithm that is still in use in the Unites States Air Force today. PCSM scores have been shown to be good predictors of various criteria within the aviation field, namely fighter/non-fighter recommendations (Carretta, 1989), passing/failing (Carretta, 1989, 1992), class rank (Carretta & Ree, 1994) and number of hours taken to complete training (Duke & Ree, 1996). Schmidt et al (1988) also report that experience seems to improve performance of candidates across the board.

Schmidt, Hunter, Outerbridge and Goff (1988) found that if experience does not exceed 5 years, the validity coefficients are as high as 0,33 when measured by supervisor ratings and 0,47 when measured using a work sample test. From these findings it is concluded that for the first five years, job experience predicts job performance – thereafter it has less utility as a predictor.

The above findings seem to point to experience being a relevant potential moderator of job performance in lower level entry positions requiring operating skills.

2.5.2 Age

In their meta-analyses, Schmidt and Hunter (1998) and Hunter and Hunter (1984), report no validity for age as a predictor of job performance. However, there are indications that as age increases, performance on psychomotor measures (more particularly the Cognitrone subtest of the Vienna Test System) decrease (Oehlschlagel & Moosbrugger as cited in Schuhfried, 2000a; Wagner as cited in Schuhfried, 2000a).

Although not directly linked to driver productivity, the literature available on the relationship between age and the performance of drivers measured in terms of safety criteria, provides further evidence supporting the investigation of age as a moderator variable in the age-job performance arena for drivers or operators of machinery. For example, Burg (as cited in Shinar, 1978) states that younger drivers are inclined to more risk behaviour. Hakamies and Henriksson (1999) maintain that age is indicative of accident involvement; in other words, the older the candidate, the more likely he or she is to be involved in an accident.

From the above findings age appears to be a relevant potential moderator of job performance in lower level entry positions requiring operating skills.

2.5.3 Years of education

According to Hunter and Hunter (1984), the amount of formal education to which a candidate has been exposed, does not correlate substantially with job performance, with a validity of only 0,10 indicated in a meta-analysis. When performance in training programmes is used as criterion, the validity increases to a relatively low 0,20. Martinussen (1996) found a mean correlation coefficient of 0,15 for education with a global pilot training performance criterion.

Whether these validity coefficients are relevant in the South African context, where the schooling system for the greater part of the population has been of questionable standard, is not certain due to there being little research available on the topic. Since the schooling system to which the current South African working population was exposed, did not enforce schooling for all learners, it is anticipated that there will be much more variability in the South African context compared to the international situation. It is also anticipated that the number of years of schooling that a candidate claims to have completed may give a very rough indication of the candidate's current literacy and numeracy level. This, in turn, could impact on the ease with which he or she acquires the knowledge and skills to perform the job.

Hence, despite little evidence to support the inclusion of years of formal education as a moderator of job performance, it is proposed that the South African educational situation for the current workforce is sufficiently different to warrant its investigation as a potential moderator of job performance in lower level entry positions requiring operating skills.

2.6 SUMMARY

In this chapter the following aspects of personnel selection methods and validation were discussed:

- It is paramount for organisations to use valid selection methods to comply with the law and to derive the potential economic gains of selecting the potentially most productive candidate for the job.
- There are various selection procedures (predictors) available for use in the selection of candidates for positions in organisations.
- Selection procedures (predictors) differ in terms of their validities to predict job performance (criteria).
- The use of more valid predictors will translate into economic gains for the organisations who utilize them.
- It thus makes economic sense for the organisation to investigate the validity of the various selection methods used in the organisation as well as those methods that are available to be used, in order to maximise the potential economic benefit that can be derived from selecting those candidates most likely to add value to the business of the organisation.
- Validation studies are central to this process
- There are various theoretical and methodological issues that impact on the quality of a validation study.

The above-mentioned aspects of personnel selection methods and validation were relevant to the study and emphasized the legal and economic necessity for validation studies as well as the theoretical and methodological issues that could potentially impact on the quality of a validation study. Potential moderators of performance (in terms of the predictors and the criteria) that may be relevant to the present study were identified.The preliminary investigation of those selection methods that may lend themselves to the selection of lower level employees at entry level for positions requiring the acquisition of operating skills for mobile machinery prompted the need for further investigation in terms of the main predictors of this study, namely general (cognitive) ability or *g* and psychomotor ability. A discussion of these two predictors follows in chapter 3.

CHAPTER 3

INTELLIGENCE

3.1 INTRODUCTION

According to Carretta and Ree, (2000, p. 229), "the measurement and structure of abilities have been a topic of speculation and study since the time of Aristotle, who distinguished ability from emotional and moral faculty".

Within the field of personnel selection, intelligence has been studied scientifically for more than a century and is probably the most researched predictor in the field (Salgado, 2000, Schmidt & Hunter, 1998). Intelligence has been credited with being the most valid indicator of job performance in numerous studies (Hunter & Hunter, 1984; Ree, & Earles, 1992; Ree & Carretta, 1996a; Schmidt & Hunter, 1998). It is also one of the main predictors in this study and hence is discussed in more detail in this chapter.

The theories on intelligence differ markedly in terms of the nature of the construct, definitions, assumptions and the measures used to assess the construct (Suzuki, Meller & Ponteretto, 1996). Taylor (1994) identifies three main schools of thought on the nature of intelligence and the methodology of its assessment, namely the structural approach, the information processing approach and the learning or dynamic approach. These were used as a framework for discussing intelligence in this study.

3.2 THEORIES ON THE NATURE OF INTELLIGENCE

3.2.1 The structural approach

The structural approach is also sometimes referred to as the "individual differences" approach and was the first school of thought on the nature of intelligence and its implications for assessment. This approach attempts to measure performance along those dimensions which are contended to form the fundamental structure of the construct, intelligence. According to Taylor (1994, p. 184), structuralists make extensive use of correlational and factor analytical techniques to resolve theoretical and empirical questions about intelligence.

The major theoretical positions in this approach are the sensory responses theory (Galton & Cattell), intelligence quotient theory (Binet & Simon), Spearman's two-factor theory, primary mental abilities theory (Thurstone) and the hierarchical theories (Taylor, 1994).

3.2.1.1 Sensory responses theory

Galton (1822-1911) (as cited in Walsh and Betz, 1990) was the first to propose the concept of general mental ability – a sphere in which he proposed that individuals differ widely. Galton contended that since all information that reaches the individual, does so through the senses, the differences in general mental ability could be ascribed to differences in sensory function, hence Galton defined intellect as the sum of the simple component parts of sensory functioning. Cattell (1860-1944) joined Galton (1822-1911) and together they developed rudimentary measures for sensory capacity, which Cattell referred to as "mental tests". This involved the measurement of psychomotor ability (encompassing measures such as reaction time) and perception (involving, among other things, the measurement of the ability of candidates to perceive differences in size and colour). Little empirical testing of their hypotheses could, however, take place until the scientific methodology had evolved sufficiently to do so. Galton and Cattell's contribution was the initial identification and propagation of the concept of general mental ability (GMA).

3.2.1.2 Intelligence quotient theory

By the turn of the century, research started indicating that the scores on the various mental tests from the sensory response theory did not correlate sufficiently with one another to be measuring one concept. Furthermore the scores did not seem to correlate with criteria with which they should logically be expected to correlate such as teachers' ratings and school results (Wissler as cited in Walsh & Betz, 1990).

As the popularity of the sensory response theory dwindled, Binet (1857-1911) and Simon (1873-1961) were in the process of developing their theory of intelligence, which was based on the "higher mental processes" of judgment and reasoning as opposed to the lower order sensory-motor capabilities of the Galton-Cattell approach. Over and above their emphasis on judgment and reasoning, Binet and Simon made an important contribution in terms of a hypothesis on the process of the development of these "higher mental processes". They proposed that the capacity to demonstrate "higher mental processes" would increase as a child increased in age and used this hypothesis as the foundation of their intelligence test. Their theory forms the base of the Stanford-Binet intelligence scale, which yields an "intelligence quotient" or, as it is more popularly known, an IQ score. IQ as an expression of intelligence is one of the most widely used intelligence measures in the world (Walsh & Betz, 1990, p. 148).

3.2.1.3 Spearman's g and s factors

Spearman (1904) is credited with being the great pioneer in the development of scientific methodology in the intelligence research field (Jensen, 1986). He devised factor analysis, which made it possible to study the factors that make up intelligence. He confirmed Galton's proposition of the existence of a "general mental ability" component and proposed that all cognitive tests had a general (*g*) and several specific ($s_1, s_2, \ldots s_n$) components. According to Spearman (1904), *g* was a component of all cognitive tests, but specific abilities were test unique (Carretta & Ree, 2000). The general ability referred to by Spearman became much narrower than the general ability referred to by Galton. Galton referred to general ability in relatively broad terms – essentially in biological and evolutionary terms. Spearman derived his conception of *g* exclusively from factor analysis (Jensen, 1986). This gave rise to one of the biggest debates in the field of psychometrics, namely whether *g* is merely a methodological artifact or a consequence of the mathematical manipulation (using factor analysis) of inter-correlations between various tests, or whether it reflects a reality that exists in the real world, independent of psychometric tests and factor analysis (Jensen, 1986). This point is discussed in more detail in section 3.3.

3.2.1.4 Multiple aptitude theories

Since many theorists felt that there were factors of mental ability somewhere between the generality of *g* and the uniqueness of the specific factors, they developed theories in which intelligence was postulated to be constituted of a number of group factors (Jensen, 1986, Taylor 1994) Thurstone (as cited in Carretta & Ree, 2000) identified seven equal "Primary Mental Abilities" such as numerical ability, verbal ability, spatial ability and associative memory ability. This led to the development of several ability taxonomies and many multiple aptitude batteries (Ekstrom, 1973; Fleishman as cited in Carretta & Ree, 2000, Taylor, 1994) such as the General Aptitude Test Battery (GATB) (Carretta & Ree, 2000).

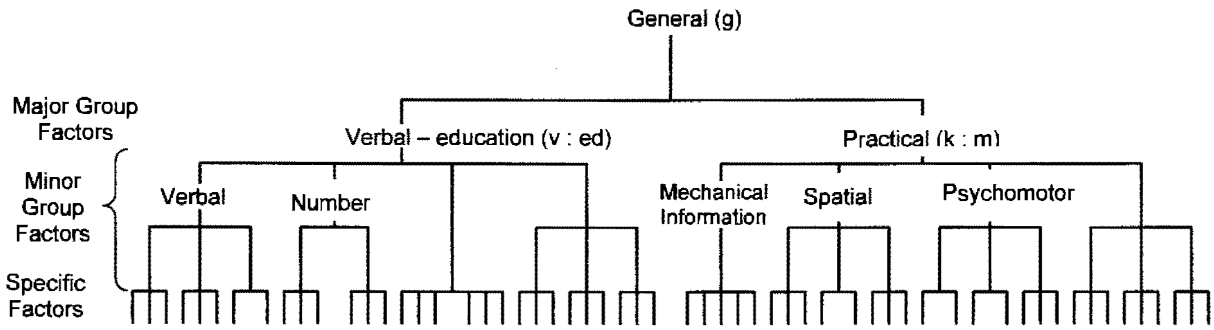
A major contribution of this school of thought was its ability to identify the pattern of abilities for specific individuals. In other words, candidates with equal amounts of *g* could, in effect, differ in terms of their particular strengths and weaknesses as far as, for example, verbal and numerical ability are concerned. This, in turn, could be relevant to the individual's success or non-success in various intellectual pursuits some of which could relate to job performance.

Taylor (1994) contends that most test construction are based on the Thurstonian model. In the South African context, there are numerous tests that measure verbal ability, numerical ability and mechanical ability, such as the Senior Aptitude Test (SAT) and the Intermediate Battery. These are strongly related to Thurstone's Primary Mental Abilities (Taylor, 1994).

3.2.1.5 Hierarchical theories

The advent of multiple aptitude theories was marred by research findings that supported the notion that various abilities are, in fact, not independent of one another and moved research in the cognitive ability field towards hierarchical models. By definition, hierarchical theories imply one or more higher order scores with several lower order scores (Burt as cited in Taylor, 1994; Carretta & Ree, 2000; Carretta and Ree, 1996a; Catell as cited in Carretta & Ree, 2000; Gustafsson, 1993, Ree & Carretta, 1994; Vernon, 1969). To a certain extent, the advent of the hierarchical theories can be seen as a move back to Spearman's model, but with more specific focus being placed on the hierarchical nature of the relationship between *g* and the specific factors (s_1, s_2....s_n) and the dynamics between *g*, specific factors and the prediction of criteria.

Figure 3.1: Vernon's hierarchical model of intelligence



Source: Walsh and Betz (1990, p. 154)

Vernon's model (see figure 3.1), for example, postulates *g* to be at the apex of the generality-to-specificity hierarchy followed by two factors verbal-educational (*v:ed*) and spatial-mechanical (*k:m*). This is followed by minor group factors, namely numerical and verbal ability as components of *v:ed* and spatial and mechanical ability as components of *k:m*. Finally, at the bottom of the hierarchy, follows the specific components of these minor group factors. Vernon's *v:ed* and *k:m* have much in common with

Cattell's (1971) second order factors namely, crystallized and fluid intelligence respectively (Walsh & Betz, 1990).

According to Taylor (1994), Cattell's (1971) theory, which is also a hierarchical theory, is one of the structural theories that is better established theoretically because it also deals with the nature rather than just the structure of intelligence. According to this theory, intelligence consists of two main components, namely fluid and crystallized intelligence. Fluid intelligence is defined as the inherited capacity which is developed by interaction with the environment (which is not culture-specific) and can be seen as a purer content-free reflection of reasoning ability (Walsh & Betz, 1990). Crystallized intelligence is defined as the specialized skill and knowledge required and promoted by a specific culture and is related to acquired knowledge (Taylor, 1994). Whilst crystallized intelligence is usually assessed with tests with informational content which draws on previously acquired knowledge (eg. vocabulary, numerical ability), fluid intelligence is usually assessed by tests with as little as possible informational content, assessing the ability to see relationships, for example pattern or series completion (Walsh & Betz, 1990).

Although crystallized and fluid intelligence have been shown to be highly correlated and often indistinguishable in test populations that are homogenous in terms of culture and education levels (Walsh & Betz, 1990), the implication is that this correlation is not necessarily prevalent in test populations where there is heterogeneity in these variables (Budoff, 1987; Laughton, 1990; Taylor, 1994). For this reason, this theory argues that the assessment of fluid intelligence will give a clearer picture of intellectual potential, particularly in culturally and educationally diverse test populations. This educational diversity is expected to be relevant in a country such as South Africa where educational opportunities were assigned on the basis of race for a great part of its history.

Cattell's culture-fair tests were developed to measure fluid intelligence. Test items are presented in abstract-diagrammatical form and involve universal activities such as odd-men-out, pattern completion, series completion and identification of conceptual relationships. There is evidence for the culture-fairness of this theory (e.g. Hakstian and Vandenberg as cited in Taylor, 1994). Cattell's model also forms the foundation of many of the learning/dynamic theories that will be discussed in the following sections.

Currently the trend in the research in the structural approach to intelligence supports the hierarchical model, postulating that the higher order factor, general ability (*g*), is at the apex, followed by lower order factors such as verbal, mathematics and spatial ability, followed by test scores at the lowest level (Carretta & Ree, 1996a, 1997, 2000, Ree & Carretta, 1994).

The structural theories have several weaknesses: (1) their concepts (eg. *g*) are dependent on factor analysis and are data-driven at the expense of being based on theory (Jensen, 1986); (2) other than Cattell's theory, they are relatively theory-poor (Taylor, 1994); and (3) the risk of possible cultural bias inherent in tests based on the structural theories (Jensen, 1986; Taylor 1994). This led to the investigation of other theories and measures of intelligence. More particularly the information processing and learning/dynamic approaches were developed. Both these approaches focus on the nature (rather than the structure) of intelligence and on the development of the theory that underpins the concept. These two schools of thought are discussed in section 3.2.2 and 3.2.3.

3.2.2 The information processing approach

The information processing approach had its advent in the 1960's as access to the processing capacity of computers became more readily available to researchers (Taylor, 1994). It can be argued that this school of thought is the ultimate operationalisation of Galton's sensory responses theory in that it sees the human being as a general-purpose information processor that obtains information from, and produces output to the environment. In this way people's relationship to the outside world is formed and maintained (Sternberg, 1982).

There are two main schools of thought in this approach namely the limited capacity theory of cognitive competence and the cognitive components approach (Taylor, 1994).

3.2.2.1 Limited capacity theory of cognitive competence

The limited capacity theory of cognitive competence holds that the "human information processing system contains one or more 'bottlenecks' which limit the flow of information, and that individuals who are able to process information faster in these bottlenecks are also more competent at problem solving and other real-life tasks" (Taylor, 1994, p. 186). According to this theory, the measurement of the receiving, processing and retrieval speeds of information links to intelligence. Due to the nature of the stimulus material (diagrammatical objects on computer screen), the measurement will probably be relatively free from the impact of prior knowledge and other environmental variables. There is evidence for working memory to correlate well with fluid intelligence measures (Baddeley, 1986; Larson & Saccuzzo, 1989).

3.2.2.2 Cognitive components approach

The major theories in the cognitive components approach, according to Taylor (1994) are amongst others:

- automatisation theory (Sternberg, 1984);
- radex theory (Guttman, 1965 as cited in Taylor, 1994);
- the circular cognitive model (Snow, Kyllonen & Marshalek, 1984); and
- the cyllindrical cognitive model (Ackerman, 1988).

Automatisation theory (Sternberg, 1984) holds that there are two main cognitive processes fundamental to intelligence, namely the response to novelty and automatisation. The way in which individuals respond to novelty, the process of mastering it, automatising it and moving towards efficiency, which, in turn, frees them to apply their mind to more novel tasks, is seen to be indicative of intelligence.

Radex theory (Guttman as cited in Taylor, 1994) holds that intelligence can be seen to be reflected by a radex with more complex tasks in the centre and less complex ones on the circumference. The actual placement of the tests along the circumference of the circular space will depend on the content of the test (e.g. verbal, numerical, spatial).

The circular cognitive model is an extension of Radex theory (Guttman as cited in Taylor, 1994). In Snow et al's (1984) study based on this model, novel rule-induction items are consistently shown to be more complex and *g*-saturated. Test content seems to have no relevance in terms of *g*-loadings. These *g*-loaded abilities are shown in the centre of the radex clustering around fluid intelligence with specific, more specialized rule applying activities plotted on the external boundaries of the radex.

The cyllindrical cognitive model (Ackerman, 1988) expands even further on Snow et al's (1983) model, by adding the notion of speed on a vertical plane, effectively transforming the circle to a cylinder. Hence the generality-specificity dimension is represented by the horizontal cross-section of the cylinder, while the power-speed dimension is demonstrated by the vertical dimension. It is argued that as the movement outwards from the core and downwards on the vertical plain progresses, previously acquired skills and knowledge play an ever greater role in the acquisition of new skills. This has obvious implications for fairness in South Africa where, as mentioned earlier, there has been great disparity in the provision and quality of education.

Although the information processing approach theory is much more precisely defined and its procedures are so simple that it seems as though there should be little bias, little cross-cultural research has been completed to confirm this empirically. Furthermore, despite having shown promising correlations with measures of fluid intelligence (Jensen, 1982, Vernon, 1990), little research has indicated correlations with criteria indicating differential performance in the real world (Taylor, 1994).

3.2.3 The learning/dynamic approach

Whereas the structural approach (and to a lesser extent the information processing approach) contends that intelligence is stable and possibly innate, in that people seem to differ in terms of the intelligence they have been allocated, the learning approach focuses on the capacity of people to adapt to the demands to the environment (Resnick & Neches, 1984) and seems to be more aligned with the humanistic paradigm. The learning approach seems to lend itself particularly well to cross-cultural assessment (Taylor, 1994).

As mentioned, Sternberg (1984) contends that intelligence comprises of an individual's ability to respond effectively to novelty and to automatize. The learning approach focuses on this adaptation of individuals to the completion of novel tasks – as demonstrated by increased speed and accuracy as a result of repetition, instructions, examples and hints (Taylor, 1994). To a certain extent the learning approach tells the researcher more than just what the individual's cognitive capacity is – there is a certain degree of diagnosis inherent in the assessment. There is a definite effort to understand the mechanics of the individual's cognitive capability and how it operates in terms of the learning processes. This, in turn, improves the understanding of how the individual operates cognitively in certain environments and also allows for individual diagnosis of problem areas, which opens possibilities for focused intervention (Bransford, Delclos, Vye, Burns & Hasselbring, 1987).

Early studies in this field reported modest correlations between learning and *g*. Taylor (1994), however, contends that this is largely due to the fact that the learning tasks used in the learning potential measures were very simple. Snow et al (1984) produced results that led to the conclusion that the more complex the learning tasks in the learning potential measures are, the higher the correlations found with *g*.

There are several approaches to the assessment of learning potential. Laughton (1990) identifies three main approaches namely:

- Budoff (1968, 1974, 1987);
- Feuerstein (1980); and
- Campion, Brown, Ferrara, Jones and Steinberg (1985).

Budoff (1968, 1974, 1987 as cited in Laughton, 1990) uses a Ravens-type task to assess subjects ranging across the intelligence spectrum. This is followed by practice and task-specific training and the assessment is then repeated. Feuerstein (as cited in Laughton, 1990) uses a similar process, but the intervention between the pre- and post-assessments involves the development of thinking skills. Campion et al (as cited in Laughton, 1990) focus on the transfer of learning and assess transfer as the inverse function of the number of hints required to complete novel tasks once the basic principles have been acquired doing other similar novel tasks.

Dynamic assessment has shown considerable variance among individuals (Taylor, 1994). Furthermore, it has been found that candidates who perform poorly on static tests are inclined to perform considerably better on dynamic tests (Budoff, 1987; Laughton, 1990). The implication seems to be that this is the case due to the reality that dynamic assessment minimizes cultural bias. This is seen to be the main advantage of the learning or dynamic approach to assessing intelligence. It is suggested that the minimization of cultural bias occurs due to tests focusing on learning tasks which are unfamiliar to all the candidates regardless of their cultures (Taylor, 1994, Sternberg, 1984). Furthermore, improvement scores are bound to cancel out any bias that remains, because each candidate is assessed against his or her own baseline performance (Taylor, 1994).

Some questions regarding the learning/dynamic approach remain. The construct validity of learning potential tests has as yet not been proved with solid empirical evidence although Taylor (1994) indicates that Campion and Brown, and Embretson have provided some promising results. Furthermore, there have been some statistical concerns regarding difference scores and gain scores which are beyond the scope of this study.

3.2.4 Taylor's (1994) integrated theory of the three schools of thought

Taylor's (1994) theory integrates all three approaches. He disagrees with the statement that processing speed and capacity is the complete foundation of intelligence (Jensen, 1982; Vernon, 1986), but contends that it is one of the two fundamental components. The other component is the ability to infer rules or abstract thinking – a concept analogous to Cattell's fluid intelligence. Taylor contends that both of these fundamental abilities are biologically or genetically determined and set upper bounds to performance in the cognitive sphere. The learning or dynamic approach is integrated into his theory in that intelligence is hypothesized to be a product of learning and other interactions with the environment.

Taylor (1994) uses Ackerman's cyllindrical cognitive (1988) model, which was referred to in the Information processing approach section, to explain the basics of his integrated theory. He suggests that the focus should be on learning potential (with the focus on fluid intelligence or general ability i.e. g which lies in the core of the cylinder) and not on learning performance (crystallized intelligence or specific abilities which are encountered when moving towards the periphery of the cylinder). Hence, the inclusion of novel tasks where the stimulus material is unfamiliar to all testees (and previously acquired skills will be of little assistance in the mastery of the tasks) is essential.

Taylor (1994) furthermore emphasizes that transfer has long been recognized as a primary component of learning, the rationale being that an individual needs to be able to take what he or she has learnt in one context and apply it in a different context, since the circumstances of the new challenge are very unlikely to be identical to previous problems or tasks.

Taylor (1994) contends that although it is unwise to do one-on-one mapping from cognitive abilities to learning processes, his two fundamental components of intelligence relate to learning processes in the following ways: abstract thinking capacity (fluid intelligence) is related to transfer and processing efficiency (information processing) is related to automatisation. Hence, he proposes that intelligence tests should tap the following four domains: fluid intelligence; information processing; transfer in tasks requiring learning potential; and automatisation in such tasks. This theory forms the foundation for his learning potential assessment instruments, namely the APIL, TRAM 2 and TRAM 1. The TRAM 1 is one of the main predictors in this research.

3.3 GENERAL (COGNITIVE) ABILITY OR *g*

The common factor in all the approaches to intelligence discussed above, is *g*. It is central to the structural approaches (Cattell's fluid versus crystallized intelligence, Spearman's two factor theory and the various hierarchical theories); the information processing theories (where all the radex and cylindrical models refer to a central, rule inferring ability analogous to *g* at the core); and the learning or dynamic approaches where more complex learning activities have shown significant correlations with *g* (Snow et al, 1984).

Jensen (1986) contends that all tests that are purported to measure intelligence are highly loaded on *g*, even when they have not been constructed with reference to factor analysis. In general, a test's *g*-loading seems to reflect the amount of complexity of mental manipulation or cognitive processing required for the testee to arrive at the correct response. An example to support this statement is the fact that in the Wechsler intelligence scales, backward digit span test is loaded on *g* just about twice as much as forward digit span test (Jensen & Figueroa as cited in Jensen, 1986). Ree and Carretta (1996a) contend that the measurement of *g* is unavoidable in all measures of ability due to the fact that when responding to test material, regardless of whether the test requires a psychomotor response, specialized knowledge or verbal skills, reasoning is unavoidable and causes *g* to be measured (Carretta & Ree, 1996b). The practical predictive validities of tests are highly dependent on their *g*-loadings. Jensen (as cited in Jensen 1986, p. 311) makes the very bold statement that "test validity would be drastically reduced, usually to a level of practical uselessness, if the *g* factor were partialled out of the reported validity coefficients in all categories of test use".

Jensen (1986, p. 311) maintains that it makes sense to identify the term, "intelligence", with *g* for two main reasons: (1) it is the highest common factor in any large and diverse collection of tests of various abilities and (2) the *g* factor is "more highly correlated than any other factors with individual differences in observable behaviours that are most commonly associated with the word intelligence".

Bearing in mind the question raised in section 3.2.1.3 as to whether *g* is merely a consequence of the mathematical manipulation (using factor analysis) of inter-correlations between various tests or whether it reflects a reality that exists in the real world, independent of psychometric tests and factor analysis, Jensen (1986) attempts to prove the "realness" of *g* by looking for correlates with psychometric *g* outside the realms of psychometrics. Jensen quotes studies on the following factors as further proof that *g* correlates with observable factors in the real world which, in turn, point to the reality that *g* does in fact exist outside of the realms of psychometrics and statistics only.

- The heritability of *g* in monozygotic and dizygotic twins (Block and Tambs, Sundet and Magnus as cited in Jensen, 1986), which supports the hypotheses that monozygotic twins' results on the Wechsler Adult Intelligence scale are more similar than dizygotic twins; and that the results between two sets of twins are more similar for those tests with higher *g*-loadings than those with lower *g*-loadings.
- Inbreeding, which impacts on *g* negatively, and outbreeding, which impacts on *g* positively (Agrawal, Sinha & Jensen; Jensen; Nagoshi & Johnson as cited in Jensen, 1986).

- The difference between white and black Americans in the performance on general cognitive ability measures. The higher the *g*-loadings of the tests, the clearer the discrepancy (Jensen, Naglieri & Johnson, and Spearman as cited in Jensen, 1986).

  This is a very contentious argument for the existence of *g*, which has understandably sparked much controversy and dispute – more particularly, because it generated the notion of "lasting inferiority", implying that black candidates would never be able to reach the levels of complexity and achievement that characterizes western civilization (Klineberg, 1980, p. 32).

  Considerable research has been conducted in this field focusing on the effects of socio-economic conditions (Humphreys, 1984), other environmental issues (Anastasi, 1988, Shobris, 1996, Nell, 1991) and cultural influences (Anastasi, 1988; Murphy & Davidshofer, 1994; Shobris, 1996). Although further discussion of this topic is beyond the scope of this study, it can be concluded that there are many variables that could impact on these differences in performance found between white and black candidates in tests loaded on *g*. It could be argued that this is further support for the common criticism of the structural school of thought of intelligence, namely that it theorises about the location or structure of intelligence, but knows little about its nature.

- Electrophysiological measures of brain activity (from the information processing approach, which is discussed in section 3.2.2), which correlate substantially with the *g*-loadings of various tests (Braden & Williams; Eysenk & Barrett; Haier, Robinson, Schafer as cited in Jensen, 1986).
- Reaction time (also from the information processing approach) which seems to correlate with the *g*-loadings of various tests (Fogarty & Stankoff; Hemmelgarn & Kehle; Jensen; Vernon as cited in Jensen 1986).

How *g* is related to learning should also be discussed, particularly to arrive at some conclusion on how findings in terms of *g* relate to learning potential, since learning potential is one of the main predictors of this study. Early studies in this field reported modest correlations between learning and *g*. Taylor (1994), however, holds that this is largely due to the fact that the learning tasks used in the learning potential measures were very simple. Snow et al (1984) produced results that led to the conclusion that the more complex the learning tasks in the learning potential measures are, the higher the correlations found with *g*.

Jensen (1986, p. 315) questions whether intelligence, or *g*, is the same as learning ability. He states that there are two distinct meanings of "learning", namely comprehension (i.e. grasping concepts, "getting the idea", "catching on") and "improvement with practice"

These two meanings concur with Sternberg's (1984, p. 269) "responding to novelty" and "automatisation". Jensen (1986) concurs with Snow et al (1984) that the first meaning, namely the acquisition of concepts and comprehension of new and progressively more difficult material, is analogous to *g*, not necessarily because they measure the same thing, but because the same brain processes are required.

The "improvement with practice" dimension (Jensen, 1986, p. 315) involves the situation where either skills have already been learnt or the skills are so easy that they do not present any problems in terms of comprehension. Practice merely improves the efficiency with which the skills are performed or adds to the acquisition of very similar material at the same level of complexity. This is analogous to Sternberg's (1984) "automatisation" concept, which, in effect, decreases the load on *g* and frees it up for other more novel learning. The example that Jensen refers to, to clarify the concept, is a musician's skill to read music. When learning the skill of reading music for the first time, the demands on *g* are quite substantial since comprehension (or the first meaning of learning referred to above) is involved. Later, however, this skill becomes automatised and *g* is freed up so that the musician can focus on other aspects, such as his interpretation of the music and expression. This second meaning of learning has relatively little correlation with *g* (Jensen, 1986).

Jensen (1986, p. 329) concludes that there is still much that is not known about the nature of *g*, but summarizes what is known as follows:

> *For the time being, what I think we can rather confidently say about g, in light of present evidence, is that* g *reflects some property or processes of the human brain that are manifested in many forms of adaptive behaviour, and in which people differ, and that increase from birth to maturity, and decline in old age, and show physiological as well as behavioural correlates, and have a hereditary component, and have been subject to natural selection as a fitness character in the course of human evolution, and have important educational, occupational, economic, and social correlates in all industrialized societies, and have behaviour correlates that accord with popular and commonsense notions of "intelligence".*

Having highlighted the centrality of "general (cognitive) ability or *g*" to all the theories of intelligence highlighted in this chapter, the term general (cognitive) ability or *g* will be used throughout this study when referring to intelligence or its measures.

There is overwhelming evidence that general (cognitive) ability or *g* predicts performance in the world of work, in terms of both performance criteria and training criteria. The discussion in the following two sub-sections will revolve around general (cognitive) ability or *g* as a predictor of job performance in personnel selection.

3.4 GENERAL (COGNITIVE) ABILITY OR *g* AS A PREDICTOR OF JOB PERFORMANCE IN PERSONNEL SELECTION

3.4.1 Predictive value in terms of job performance criteria

As early as 1919, Yerkes (as cited in Duke & Ree, 1996) demonstrated general (cognitive) ability or *g* to be a valid predictor of pilot training success. Similarly, Flanagan (as cited in Duke & Ree, 1996) as early as 1942, provided evidence that the aviation selection exam was in fact a general mental ability battery measuring predominantly *g* in terms of comprehension and reasoning.

In a later meta-analysis, Hunter and Hunter (1984) reported very promising validity coefficients for general (cognitive ability) as predictor of job performance. The validity coefficients reported, range from 0,29 to 0,61 with the average across job validity coefficient reported, being 0,45.

The research in this field seems to indicate that the predictive validity of *g* increases as job complexity increases and is highest in those occupations involving the least automatisation of performance and the greatest amount of specialized training. (Gutenberg, Arvey, Osburn & Jeanneret, 1983; Hunter, 1983 as cited in Levine et al, 1996; Hunter, 1986; Jensen, 1986). The magnitude of the validity coefficients found in the Hunter and Hunter (1984) meta-analysis, which involved 32000 employees in the US Department of Labour, are 0,58 for professional managerial positions, 0,56 for high level complex technical jobs, 0,51 for medium complexity jobs, 0,40 for semi-skilled jobs and 0,23 for unskilled jobs, confirms this contention.

The lower validity of this predictor for lower complexity jobs is supported by a study done by Levine et al (1996) focusing on technical skilled and semi-skilled positions in the engineering trades. Findings supported the validity of cognitive tests across jobs, but more specifically increases in the validity of the cognitive ability as a predictor of job performance as the complexity of the job increased and, conversely, decreases in its validity as the complexity of the job decreased.

Various studies highlight the fact that although the validity of cognitive ability varies across jobs, it never approaches zero (Hartigan & Wigdor, 1989; Hunter & Hunter, 1984; Hunter, 1986, Schmidt & Hunter, 1981; Schmidt et al, 1988). Schmidt, Hunter and Pearlman (1981) dispel the popular notion that task differences are moderators in the cognitive ability–job performance relationship. They prove the moderating effects of tasks to be negligible even when tasks differ markedly and to be non-existent if task differences are less extreme. Hunter and Hunter (1984) as a result of their findings, argue very convincingly (especially in the case of entry level jobs), that when used in isolation, any predictor other than general cognitive ability has a validity so much lower, that substitution would mean great economic loss to the organisation in question.

Although experience seems to improve performance of candidates across the board (ie, regardless of their level of intelligence), the performance differences between high general ability and low general ability candidates is inclined to persist (Schmidt et al, 1988).

When reviewing the literature reporting on the validity of *g* as a predictor of job performance, results are mixed but generally encouraging. In two meta-analyses of aviation studies completed by Hunter and Burke (1994) and Martinussen (1996), for example, only modest validity coefficients were reported for *g*. In the Hunter and Burke study, the validity coefficient for general (cognitive) ability or *g*, was reported to be only 0,13. In the Martinussen study, validity coefficients of 0,22 for cognitive (eg. mechanical comprehension, spatial orientation, perceptual speed and attention) and 0,13 for intelligence (specifically focusing on *g*) were reported. It is suggested that these validity coefficients were probably under-reported, due to the fact that no correction for statistical artifacts (range restriction and lack of reliability) was included (Carretta & Ree, 2000). These results are also not in line with the vast amount of research indicating good correlation of *g* with both job performance and training success criteria (Carretta & Ree, 1996b; Hunter, 1986; Hunter & Hunter, 1984; Schmidt & Hunter, 1998). In the case of the Martinussen (1996) study, the question may be raised whether the split between intelligence and cognitive ability tests is relevant. Bearing in mind the research indicating the overlap of *g* with these types of tests and the findings that little additional utility is gained from adding specific cognitive ability factors to *g* (Carretta & Ree, 1996b, 1997, 2000; McHenry et al, 1990; Olea & Ree, 1994; Ree & Earles, 1992; Ree et al, 1994), it is suggested that it is highly likely that the correlation of a joint (i.e. cognitive and intelligence) factor with the criteria of pilot success, would have been higher.

Carretta and Ree (1995) examined the validity of 16 subtests of the Airforce Officer Qualifying Test (AFOQT) for the prediction of five pilot training criteria based on academic grades, performance on daily flights and performance on test flights. The correlation of the *g* factor loadings of the AFOQT subtests with the average validity of the subtests for predicting the five pilot training criteria was 0,62. The higher the sub-test's loading on *g*, the higher its validity for predicting the criteria was reported to be.

Schmidt and Hunter (1998) also report excellent validity coefficients for *g*. They compare various selection methods and their validities for the prediction of job success. Due to the high predictive validity coefficient of 0,51 obtained for GMA (general mental ability) tests, which load on *g*, they suggest that it is used as the main selection measure in all selection exercises and that the additional utility of adding supplementary measures is then determined. This is done for each of the 19 measures in the meta-analysis. The combinations that provide the highest validity coefficients for the prediction of job success are *g* plus an integrity test (validity of 0,65), *g* plus a structured interview (validity of 0,63) and *g* plus a work sample test (validity of 0,63).

Jensen (1986) refers in more general terms to the relevance of cognitive ability to success in the world of work. His definition of success is very generally "attained occupational status and all its socio-economic correlates" (p. 317). It is reported that the IQ's of school going children are substantially correlated with their adult occupational level (Jensen, 1986; Austin & Hanisch; 1990). This is purported to be the case, since in the same way as tests make different demands on *g*, different occupations also make different demands on *g* (either due to the educational requirements for the position or differing *g* demands of the positions themselves). Individuals are inclined to settle into jobs with *g* demands commensurate with their own *g* levels, since these are the occupations within which they are most likely to be successful. Jensen (1986) contends that these correlations are in the 0,50 to 0,70 range (higher than the correlations found by Avolio and Waldman (1990) who report a correlation of 0,36 for the complexity of individuals' jobs and their general cognitive ability). This would have been higher were it not for certain influencing factors, namely other traits, interests and special talents also being related to success in various occupations; factors irrelevant to the occupation also moderating the *g*-job success relationship, such as background, opportunity, chance (i.e. "right place right time"); as well as the fact that most occupations are made up of various activities – all of which have different *g*-loadings. This, again, impacts on the *g*-job success relationship. Jensen (1986) concludes that although *g* cannot account for all the variance in occupational level, it accounts for more than any other source of variance that has been discovered.

In general, the validity coefficients relating to job performance criteria are lower than those relating to training criteria. This may be partially ascribed to the fact that the criterion is more indirect with more moderating variables. It is also more difficult to define than training criteria (eg. quality of performance as an artisan, versus performance of the same artisan in the N2 examination and trade test where the latter is more easily defined and operationalised than the former).

3.4.2 Predictive value in terms of training criteria

Research findings support the notion that intelligence correlates equally well or better with performance in training programmes (Levine et al, 1996; Schmidt & Hunter, 1998). Jensen (1986) contends that IQ, or rather the *g* abilities it measures, acts as an entrance variable with respect to educational attainments, with higher levels of IQ being necessary, but not sufficient, for passing higher educational hurdles.

Koonce (as cited in Griffin & Koonce, 1996) hypothesizes, for example, that a criterion that is relevant to successful performance in military flight is what he calls rate of acquisition (i.e. of learning). This would relate to the dynamic/learning theories of intelligence. The argument is that there are specific blocks of time allocated to the acquisition of specific skills, such as acrobatics, formation, instrument flying, and that pilots in training need to meet stringent performance criteria within specified time limits, hence the importance of rate of acquisition in military flight training. This would relate to learning theory concepts such as automatisation (Sternberg, 1984) and transfer (Taylor, 1994).

The relevance of relating general (cognitive) ability or *g* to training criteria, when the focus of the current study is job performance criteria, is that training performance can be seen as a precursor to job performance in that training imparts job knowledge which, in turn, facilitates the performance of job tasks (Carretta & Ree, 1996b). It is likely that an employer selecting employees on the basis of general (cognitive) ability or *g*, will be selecting employees with the ability to acquire new skills from training programmes and from experience on the job fast and well. It is this knowledge of how to perform the job that gives them the ability to perform better on the job. General (cognitive) ability or *g* also impacts on other job aspects over and above the acquisition of job knowledge, but it is postulated that this is the main effect (Schmidt & Hunter, 1998).

Hunter and Hunter (1984) support this statement by maintaining that if people need to complete training in order to be able to do the jobs for which they have been selected, there is no predictor nearly as good as general (cognitive) ability or *g*. This is supported by their findings that the average validity coefficient across jobs for training criteria is 0,54 which is in line with Schmidt and Hunter's (1998) findings of an average validity coefficient of 0,56 for training success criteria. Schmidt and Hunter (1998) report that the combinations that provide the highest validity coefficients for the prediction of training success are *g* plus an integrity test (validity of 0,67), *g* plus a conscientiousness test (validity of 0,65) and *g* plus a work sample test (validity of 0,63).

3.4.3 Moderator variables

A relevant consideration highlighted by Jensen (1986) is the fact that socioeconomic status is a very relevant moderator variable in the relationship between intelligence and occupational level/job status attained (as a measure of job success). He cites the results of Humphreys' (1984) research in the United States, which concluded that if socio-economic status were to be partialled out of the validity coefficient for the prediction of job status and were thus totally reliant on *g*, the advantage of white middle class children over working class children would be reduced by one third. This outcome may prove to be very relevant in the South African situation where large variation in socioeconomic status remains a reality.

3.4.4 General versus specific ability

Recent research in the educational and employment contexts seems to indicate that when the utility of general (cognitive) ability or *g*, i.e. the higher order factor referred to above, is compared with specific cognitive ability (the lower order factors), *g* is the better predictor with very little additional utility being gained from adding specific cognitive ability factors for example, numerical, spatial or verbal abilities (Carretta & Ree, 1997, 2000; McHenry, Hough, Toquam, Hanson & Ashworth, 1990; Olea & Ree, 1994; Ree & Earles, 1992; Ree et al, 1994).

In general, findings indicate that *g* accounts for between 30 – 65% of variance and the largest portion of accounted for by a lower order specific cognitive factor is approximately 8% (Carretta & Ree, 2000). The additional utility of specific cognitive ability factors tend to range between 0,02 and 0,03 seeming to indicate that the multiple aptitude school of thought on intelligence is not empirically supported. Olea & Ree (1994) reported slightly higher incremental validities for specific ability beyond *g*, ranging from 0,075 to 0,139. However, the incremental validity was suggested to be due to specific knowledge about aviation (i.e. aviation controls, instruments and principles) rather than specific cognitive abilities (i.e. verbal, quantitative, spatial and perceptual speed).

This finding is particularly relevant to the South African employment scenario. Bearing in mind that the assessment of specific cognitive ability is on the periphery of Ackerman's (1988) cylinder referred to earlier and is susceptible to the influence of previously acquired skills, the large disparities in the educational system to which different cultural groups were exposed, would definitely impact on the assessment outcomes of previously disadvantaged groups (Taylor, 1994). It is therefore encouraging to know that the assessment of specific ability would add very little additional utility to the selection process and could probably, relatively safely, be omitted, as it is in many assessments utilizing the information processing and learning approaches.

3.5 SUMMARY

In summary the findings pertaining to the current study are:

- There are three schools of thought on intelligence.
- Taylor (1994) formulated an integrated approach encompassing all three schools of thought which underlies the learning potential measure (TRAM 1) used in this study.
- General (cognitive) ability (*g*) is a better predictor of intelligence than any of the specific abilities (s_1, s_2...) with very little additional utility being gained from adding specific cognitive ability factors (Carretta & Ree, 1997, 2000; McHenry, Hough, Toquam, Hanson & Ashworth, 1990; Olea & Ree, 1994; Ree & Earles, 1992; Ree et al, 1994).
- *g* is central to the intelligence constructs in all the approaches to intelligence (Jensen, 1986; Snow et al, 1984; Taylor, 1994).
- *g* is related to learning and learning potential (from the learning / dynamic school of thought) (Jensen, 1986; Snow et al, 1984; Taylor, 1994).
- *g* is related to reaction time (from the information processing school of thought) (Jensen, 1982, Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992).
- Good validity coefficients are reported in the literature for *g* as a predictor of job performance as well as training criteria – the more complex the job, the greater the validity coefficients reported (Gutenberg et al, 1983; Hartigan & Wigdor, 1989; Hunter, 1986; Hunter & Hunter, 1984; Jensen, 1986; Levine et al, 1996; Schmidt & Hunter, 1981; Schmidt et al, 1988; Schmidt & Hunter, 1998).

- Socio-economic status may be a moderator in the intelligence-performance relationship (Humphreys, 1984; Jensen, 1986).

The implications for the current study are that there seems to be good evidence for *g* predicting both job performance and training criteria with little additional utility being gained from specific intelligence factors such as verbal and mathematical ability. The research and literature on this construct also seem to point to learning potential measures (from the learning/dynamic school of thought) and reaction time (from the information processing school of thought) being related to *g*. It would be interesting to see if learning potential also yields similarly high validity in the prediction of job-related criteria. The fact that the majority of the South African population is exposed to less than ideal socio-economic conditions may be a relevant factor to the current study. Due to the limited scope of this study, the extent of this impact will not be assessed or discussed in this study.

In chapter 4, psychomotor ability will be discussed as a potential predictor of job performance.

CHAPTER 4

PSYCHOMOTOR ABILITY AND ITS RELATION TO INTELLIGENCE AS PREDICTOR OF PERFORMANCE

4.1 INTRODUCTION

As with cognitive ability, psychomotor ability has been studied for more than a century (Carretta & Ree, 1997). In general, research findings have been supportive of the validity of psychomotor tests in the prediction of job performance in positions requiring operating or driving skills (Carretta & Ree, 2000; Martinussen, 1996, Schoeman, 1995; Wheeler & Ree, 1997). The findings also seem to point to the fact that as job complexity decreases, the validity of psychomotor skills increases (Hartigan & Wigdor, 1989; Hunter & Hunter, 1984; Levine et al, 1996).

This is one of the main predictors in this study and hence is discussed in more detail in this chapter.

4.2 THEORIES ON THE NATURE OF PSYCHOMOTOR ABILITY

Fleischman (1988) piloted a large proportion of the scientific research in this field. His research centred on the identification of the abilities that psychomotor tests measure, the construct validity of these tests, developing taxonomies of psychomotor abilities and proving that job sample performance tests could safely be replaced by more generic psychomotor tests, with the added advantage of greater standardization.

Fleischman and Quaintance (1984) identified 11 separate psychomotor factors. The argument at this stage of the development of theories on psychomotor ability was that these specific psychomotor factors (eg, eye-hand coordination or reaction time) contributed separately to the validity of a psychomotor battery. In other words, there were no higher order psychomotor factors. Therefore the focus was on the accurate measurement of these separate specific psychomotor factors, since each of these separate factors contributed uniquely to a candidate's overall psychomotor ability measurement (Cronbach as cited in Wheeler & Ree, 1997).

More recent research on psychomotor ability, points to the existence of a higher order general psychomotor factor that underlies all psychomotor tests (Carretta & Ree, 1997; Chaiken, Kyllonen &

Tirre; 2000; Ree & Carretta, 1994). This, contrary to Fleishman and Quaintance's (1984) theory, implies that psychomotor tests measuring various specific factors may be relatively interchangeable.

Ree and Carretta (1994), for example, investigate the relative contribution of the general psychomotor factor (*p*) and the specific psychomotor factors to performance in psychomotor tracking tests and report that 68% of the variance can be ascribed to the general psychomotor factor (*p*), while the lower order factors contribute only 24% of the variance.

4.3 PSYCHOMOTOR ABILITY AS A PREDICTOR OF JOB PERFORMANCE AND TRAINING CRTIRIA

Psychomotor tests were used in the armed forces from as early as the early 1900's. Research in this area started during the First World War, as part of a general drive to improve pilot selection techniques in the United States of America. Early in the war, candidates were selected for flight training on a volunteer basis. When casualty reports started coming in, it was noted that accidents were not always due to equipment error or enemy action and that many casualties were actually due to human error (Griffin & Koonce, 1996). Add to this concern for safety, the fact that pilot training costs were, and still are, astronomical (Duke & Ree, 1996; Griffin & Koonce, 1996; Martinussen, 1996; Ree & Carretta, 1998) and the birth of scientific selection in the aviation sphere is well justified.

The United States of America was not the only country to reach this conclusion. According to Dockeray and Isaacs (as cited in Duke & Ree, 1996), Italy was, in fact, the first country to implement a pilot selection programme, using measures of reaction time, balance, perception of muscular effort, and attention. The French, at this stage, were also investigating reaction time and emotional stability as predictors.

Early studies indicated that psychomotor tests had considerable validity for predicting both Air Force and Navy pass/fail rates (Mashburn as cited in Griffin & Koonce, 1996) and the success or non-success of pilots and bombardiers in training programmes before and during World War II (Melton as cited in Fleischman, 1988).

During World War II, research in the psychomotor field gained momentum (Duke & Ree, 1996, Griffin & Koonce, 1996). Many pilots needed to be trained in a very short period and the training remained costly. Research at that time proved that psychological and, what was then called "apparatus" assessment, had much higher validity than physiological tests, which proved to have no validity over chance (Franzen & McFarland; Viteles as cited in Griffin & Koonce, 1996). Many apparatus or psychomotor tests were developed during this time with encouraging validities for various criteria (Fiske; Melton; Viteles as cited in Duke & Ree, 1996) in the USA, Canada (Signori as cited in Duke & Ree, 1996), the United Kingdom (Parry as cited in Duke & Ree, 1996), and Germany (Duke & Ree, 1996).

Psychomotor assessment remained a part of the United States Air Force pilot-selection programme until 1953 when its use was discontinued, mostly due to the practical difficulties of keeping the apparatus calibrated for many testing sites and equipment failure that started to become quite prevalent as the apparatus aged, with obvious negative effects on the reliability of these tests (Ree & Carretta, 1998).

With the advent of the use of computer technology for the assessment of psychomotor skills, interest in this form of assessment was revived (Duke & Ree, 1996; Fleischman; 1988; Griffin & Koonce, 1996; Ree & Carretta, 1998).

Fleischman (1988) emphasizes the fact that the advent of, what he refers to as, microprocessors and minicomputers made it possible to assess abilities that were not possible to assess adequately using paper and pencil tests, for example, divided attention, ability to concentrate and ability to function under different time pressures. Computers have also made it possible to assess better, various abilities that have, in the past, been assessed using apparatus tests. This is due to the dynamic capabilities of the computer, which lends itself better to the assessment of constructs such as perceptual speed, spatial visualisation and reaction time. Further advantages are the improvement of the reliability of the presentation of the tests as well as the accuracy of measurement and data collection, the elimination of the need to calibrate test apparatus and the reduction in the equipment failure that is typical of apparatus tests (Griffin & Koonce, 1996, Ree & Carretta, 1998).

By far the greatest amount of research in the use of psychomotor ability as predictor in selection exercises within the last three decades, has centred on pilot selection in the military context.

In one of the earlier studies, Long and Varney (as cited in Ree & Carretta, 1998) embark on the very ambitious project of developing a detailed five-hour learning sample of flight tasks known as the Automated Pilot Aptitude Measurement System. This measurement tool can be likened to a combination of a work sample test and a psychomotor test. Although validities were encouraging and face validity was very high, the cost was prohibitive to its implementation in the US Air Force. The Canadian Air Force which follows an assessment centre type of approach for the selection of pilots has, however, implemented the Canadian Automated Pilot Selection System, which is to a great extent based on the original work of Long and Varney.

The Pilot Candidate Selection Method or the PCSM algorithm received a great deal of attention in the research completed in the United States Airforce within the last two decades.

In 1985 the United States Air Force developed an experimental computerised psychomotor test known as the Basic Attributes Test (BAT) (Carretta as cited in Ree & Carretta, 1998). It consists of the following sub-tests: Two Hand Coordination, Complex Coordination, Mental Rotation, Item Recognition, Time Sharing and a facet focusing on attitudes towards risk taking (Griffin & Koonce, 1996). Carretta (1992), reports that when the Air Force Officer Qualifying Test (AFOQT) subtest scores (measuring to a large extent, *g*), the Basic Attributes Test (BAT) scores (measuring psychomotor coordination, perceptual and cognitive processes, and attitudes towards risk) and previous flying experience, are combined in a single regression equation (referred to in the literature as the Pilot Candidate Selection Method or the PCSM algorithm), a multiple R of 0,31 is obtained. Bearing in mind the historically disproportionate ratio of graduates to failures (5:1), the dichotomous nature of the criterion and range restriction (Thorndike & Hagan, 1977), this is probably an underestimation of the battery's validity (Duke and Ree, 1996).

In a cross-validation study, Carretta and Ree (1994) report that when the PCSM algorithm is used, a multiple R of 0,38 is obtained - higher than Carretta's (1992) study referred to above. A supplementary observation is that the PCSM scores do not only predict attrition (i.e. candidates who are not successful in the training programme), but they also predict class rank.

In another cross-validation study, Duke and Ree (1996) use the PCSM algorithm as predictor and the number of flying hours that it takes a candidate to demonstrate competence in flying, as criterion. They report a significant negative correlation between PCSM scores and additional flying hours (-.20 and -.27 for two different aircraft – the first being easier to fly than the second). Although these validity coefficients are in the low to moderate range, they are sufficient to impact on training and consequentially on the costs of training. The researchers proceed to quantify the cost of the additional flying hours that could have been avoided if higher cutoffs were used on the PCSM. (These turn out to be quite significant at $1.1 million in 1996). Duke and Ree (1996) argue that there are sufficient candidates available who apply for pilot training to be able to increase this cutoff and conclude that it hence makes economic sense to do so.

In summary, PCSM scores have been shown to be good predictors of various criteria within the aviation field, namely fighter/non-fighter recommendations (Carretta, 1989), passing/failing (Carretta, 1989, 1992), class rank (Carretta & Ree, 1994) and number of hours taken to complete training (Duke & Ree, 1996).

In studies, using combinations of predictors other than the PCSM algorithm, significant correlations between psychomotor ability as measured by the Porta-Bat and pilot performance have been reported (Cox, 1988; Kantor & Carretta, 1988). Kantor and Carretta (1988) conclude that psychomotor tests significantly predict a pass-fail criterion in flight training. A multiple-regression correlation of 0,22 is reported. When the psychomotor tests are combined with other tests (AFOQT), age and performance in a light aircraft screening programme, a high multiple correlation coefficient of 0,45 is found with the pass-fail criterion. Cox (1988) reports significant correlation of the results of two of the sub-tests (i.e. the Two Hand Coordination Test and the Complex Coordination Test) of the Porta-Bat with a pass or fail criterion.

Wheeler and Ree (1997) investigate the validity of general and specific psychomotor skills for predicting success in aviation training as well as success in flying work sample tests. It is reported that general psychomotor ability is the best predictor of flying criteria with validity coefficients of 0,192 (0,285 once corrected for dichotomization) and 0,278 being reported for training success and performance in the work samples respectively. The only measure that provided incremental utility beyond general psychomotor ability was reaction time. Interestingly, this dimension is often identified as a measure of cognitive ability within the information processing school of thought (Jensen, 1982, 1993; Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992).

Aviation meta-analyses are inclined to report moderate validity coefficients for psychomotor measures (Hunter & Burke, 1994; Martinussen, 1996). In the Hunter and Burke (1994) study, validity coefficients of 0,32 for gross dexterity, 0,10 for perceptual speed and 0,28 for reaction time are reported. Relatively low validity coefficients are reported in the Martinussen (1996) study, with a validity coefficient of 0,20 for psychomotor / information processing being reported. In both studies, it is contended that validity coefficients were probably under-reported, due to the fact that no correction for statistical artifacts such as range restriction or lack of reliability were included (Carretta & Ree, 1997).

In a South African validity and utility study for the selection of train drivers using the Vienna Test System (which is used to assess psychomotor functioning) as predictor and simulator operating performance as criterion, Schoeman (1995) reports an uncorrected validity coefficient of 0,50, which is reported to increase to 0,70 once corrected for shrinkage and 0,87 once corrected for unreliability of the criterion. These coefficients seem spuriously high. Bearing in mind the small sample (N=62), these results need to be interpreted with caution.

Bouwer's (1983) study investigating the differences between heavy duty vehicle drivers who had been divided into strong and weak performing groups, reports that in terms of the psychomotor tests used, significant differences were found between the two groups when using an adjusted version of the GATB dot-test. No significant differences were found between the two groups in terms of their performance on the apparatus tests designed to measure psychomotor skills, more specifically reaction time, two hand coordination, and information processing. The sample size was particularly small (N=58), the design relatively simplistic (with the artificially dichotomous grouping of subjects based on a criterion that is in all probability continuous) and no corrections were attempted for statistical artifacts – an indication that validity coefficients are possibly under-reported and need to be interpreted with caution.

In this section, a concerted effort was made to try to keep the focus on psychomotor ability alone, as a predictor of job performance and training criteria. However, these studies also refer to a general (cognitive) ability measure as additional predictor (Carretta, 1989, 1992; Carretta & Ree, 1994, Duke & Ree, 1996, Hunter & Burke, 1994, Martinussen, 1996). This seems to point to a relationship between general (cognitive) ability and psychomotor ability and hence, that these two variables probably interact to predict performance.

Having focused on general (cognitive) ability or *g* as predictor in chapter 3 and on psychomotor ability as predictor in chapter 4., it makes sense to, at this stage, examine these predictors, firstly, in terms of how they relate to one another and then how they operate together to predict performance and training criteria. First, however, it is important to mention the criterion problem as it relates to validation research in the psychomotor field.

4.4 THE CRITERION PROBLEM IN PSYCHOMOTOR VALIDATION STUDIES

A factor that has always been a problem in validation research in general, and more specifically in research in the psychomotor field, is the criterion and how it should be reliably measured.

Early validation work on various pilot selection batteries including psychomotor assessment was marred by the lack of an acceptable criterion for flight performance, due to inconsistent ratings of instructors, poor record-keeping of pilot performance and a low percentage of pilots who actually failed (Mc Farland, 1953) – the criterion problem at its best (Anastasi, 1988; Cascio, 1991; Thorndike & Hagen, 1991; Walsh, 1989).

Griffin and Koonce (1996) emphasize that the criterion problem in pilot selection in the US military service, is still alive and well. Historically, a dichotomous pass/fail criterion was used in pilot selection studies. This became progressively more inappropriate as improved selection techniques made the selection pool progressively more homogenous (they are generally all good), leading to lower levels of attrition (ie, fewer candidates fail). Due to the decreased variance in the ability of the candidates being assessed, correlations between predictors and criteria by definition cannot be very high (Griffin & Koonce, 1996). It also limits the understanding of the relationships between predictors and criteria and limits statistical power (Cohen 1983, Duke & Ree, 1996, Hunter & Schmidt, 1990).

This led to a move toward different criteria such as check ride or flight grades, and more objective criteria, such as the number of training flight hours required before a candidate is deemed competent to fly solo (Duke & Ree, 1996). Since these criteria are more or less normally distributed and continuous (rather than artificially dichotomous), higher uncorrected correlations have been reported in studies in this field (Carretta & Ree, 1994; Griffin & Koonce, 1996).

4.5 THE RELATIONSHIP BETWEEN GENERAL (COGNITIVE) ABILITY AND PSYCHOMOTOR ABILITY

As suggested in section 4.2, it has been a common view that cognitive and psychomotor ability are distinct from one another (Caroll, 1993; Fleishman & Quaintance, 1984). Carretta and Ree (1997, 2000) contend that this view possibly endures due to the fact that the ways in which these two factors are measured are so different.

In reality though, many recent studies have found a relationship between cognitive and psychomotor ability (Carretta & Ree, 1997; Chaiken et al; 2000; Hunter & Hunter, 1984; McHenry et al, 1990; Rabbitt, Banejeri & Scymnaski, 1989; Ree & Carretta, 1994; Tirre and Raouf, 1998), hence it is concluded that these factors are, in fact, not independent.

Various studies focusing on the direct correlation between psychomotor tests and general (cognitive) ability tests have found significant correlations (Rabbit et al, 1989, Tirre & Raouf, 1998). Rabbit et al (1989) report a very high correlation coefficient of 0,69 between an IQ test as predictor and a psychomotor test called "Space Fortress" as criterion. Tirre and Raouf (1998) obtain a low correlation of 0,20 between a working memory composite and a multi-limb coordination composite with no correction for attenuation of range restriction having been incorporated.

Various factor analytical studies also point to a relationship between psychomotor ability and general (cognitive) ability.

Chaiken et al (2000), for example, create a general psychomotor factor from various psychomotor tests and three cognitive factors, namely, general ability (*g*), temporal processing and processing speed. The correlation between general psychomotor ability and *g* is a very high 0,69 before correction for attenuation and range restriction, a convincing case for the assertion that cognitive ability and psychomotor ability are correlated. (The correlation of the general psychomotor factor with temporal processing was 0,25 and the correlation with the processing speed factor was not significant).

Ree and Carretta's (1994) factor analytic study reports that performance on a psychomotor tracking test yields a general cognitive factor, a general psychomotor factor and specific psychomotor factors. In order to determine the relationship between these factors, they complete a correlation study between four verbal and mathematics tests loading on *g;* and eight psychomotor tests. An average correlation coefficient of 0,34 is reported, once corrected for range restriction.

Carretta and Ree (1997), administered 17 computer-based psychomotor tests designed to measure varying psychomotor factors (arm/hand movement, complex coordination, finger and hand dexterity, kinaesthetic memory, leg reaction time, pursuit tracking and rate control) and the Armed Services Vocational Aptitude Battery (ASVAB) that assesses general cognitive ability (*g*) and three lower order factors namely, verbal/mathematics, speed and technical knowledge to 429 military recruits. They regress each psychomotor score on each of three sets of ASVAB tests. (The first set consists of four verbal and mathematical tests, the second of six speed and technical knowledge tests and the third set consists of all ten ASVAB tests). The correlations (after correction for range restriction) reported are 0,283, 0,412 and 0,440 for the three sets, demonstrating a relatively clear overlap between psychomotor ability and general (cognitive) ability (*g*). This study's results i.e. the correlation of the cognitive and psychomotor scores; the fact that both the paper and pencil and the psychomotor tests load on *g;* the existence of a higher order psychomotor factor; and the existence of several lower order factors largely confirms and extends on their 1994 study. In the 1997 study more psychomotor tests are included, enlarging the psychomotor domain and a technical knowledge test is included, which yields a psychomotor / technical knowledge (PM/TK) factor rather than the pure general psychomotor factor reported in their 1994 study. Furthermore, the results suggest that psychomotor tests may be interchangeable to the degree that they load on *g* and PM/TK respectively.

Having focused, thus far, on the fact that there is a relationship between general mental ability and psychomotor ability, in the following section the focus will be on reviewing research that investigates how these two predictors interact to predict both performance and training criteria.

In general, research seems to point to general (cognitive) ability or *g* being the main predictor of performance and training criteria in jobs requiring psychomotor skills, with psychomotor measures providing incremental validity beyond *g* (Carretta, 1989; Carretta & Ree, 1994, 1997; Gibb & Dolgin, 1989; Hunter & Hunter, 1984; McHenry et al, 1990; Ree & Carretta, 1994).

Incremental validity refers to the ability of one predictor to add to the predictive effectiveness of another predictor (Carretta & Ree, 1996b). In other words, if the validity of *g* is, for example 0,37 and the validity of *g* combined with psychomotor ability is 0,40, then the incremental validity of psychomotor ability is the difference between the two validity coefficients i.e. 0,03.

Two early studies confirmed that psychomotor assessment had incremental validity beyond the United States Air Force (USAF) pencil-and-paper selection battery honing in on *g* (Hunter & Thompson; McGrevy & Valentine as cited in Griffin & Koonce, 1996).

In their meta-analysis focusing on the validity and utility of alternative predictors of job performance, Hunter and Hunter (1984) found validity coefficients of 0,446 for cognitive ability and 0,319 for psychomotor ability. However, when the incremental utility of psychomotor ability beyond *g* is investigated, the increment is statistically significant, but very small. An average increment of 0,056 is reported.

Carretta (1989), also found significant, but small, added variance for psychomotor measures beyond the variance accounted for by the Air Force Officer Qualifying Test (AFOQT). The AFOQT scores are added into a multiple regression model first. When the remaining variables are allowed to enter the regression model in a stepwise fashion, psychomotor ability enters first with a multiple R of 0,44 for the two variables combined.

In a concurrent validity study of nine US Army enlisted entry-level skilled jobs, McHenry et al (1990), found validity coefficients of 0,47 for *g*, 0,47 for spatial ability and 0,37 for psychomotor ability as predictors of general soldiering proficiency – once they had been corrected for shrinkage and range restriction they were reported to be 0,65 for *g*, 0,63 for spatial ability and 0,57 for psychomotor ability, respectively. However, when combined in a multiple regression equation, psychomotor ability yields an increment of only 0,01 and spatial ability an increment of 0,02 beyond *g*. Similarly, small increments in the region of 0,03 and 0,06 have been reported in studies with aviation criteria (Gibb & Dolgin, 1989; Ree & Carretta, 1994).

Carretta and Ree (1994) report incremental validity beyond *g* for the PCSM algorithm referred to earlier. They conclude that the best predictors of training performance remain the AFOQT subtests (R = 0,31) which have been shown to be highly loaded on *g*. When the PCSM algorithm is used (which adds psychomotor ability and experience), a multiple R of 0,38 is reported – hence a small incremental validity of 0,07.

In their 1997 study, Carretta and Ree found that *g* and PM/TK account for similar percentages of the variance found in the joint analysis of cognitive and psychomotor tests, with *g* contributing 31.6% and PM/TK contributing 31.07%. No specific factor (eg, finger dexterity, multi-limb control etc.) accounts for more than 7% of the variance i.e. the two higher order factors contributed more to the common variance (62.66%) than all the lower-order factors combined. It is also reported that the psychomotor tests all contribute to both the general cognitive ability factor (*g*) and the PM/TK factor. These findings suggest that psychomotor tests will not add much incremental validity beyond *g*. These findings are generally consistent with the findings of their 1994 study.

A relevant issue pertaining to the relationship between general cognitive ability and psychomotor ability is job complexity.

4.5.1 The influence of job complexity

There seems to evidence in the research for the trend that as information processing demands of the job increases, the validity of psychomotor measures decreases (Carretta & Ree, 1997, Hartigan & Wigdor, 1989; Hunter & Hunter, 1984, Levine et al, 1996).

In a study by Hunter (as cited in Levine et al, 1996), the various tests in the General Aptitude Test Battery (GATB) are classified in the composite categories cognitive, perceptual and psychomotor and GATB scores are analyzed for 12000 jobs. The findings of this study point to the validity of the cognitive tests of the GATB across jobs, but more specifically point to increases in the validity of the predictor, as the complexity of the job increases (as referred to in chapter 3). The findings also support the validity of psychomotor tests for jobs with lower levels of cognitive complexity.

Levine et al (1996) report similar results with the highest validity of psychomotor measures being reported for those jobs requiring the least amount of information processing (an overall validity coefficient of 0,23 versus a validity coefficient for assembly workers of 0,39). The implication is that as the jobs become less complex or cognitively demanding, that psychomotor tests become more relevant. In fact, the validity and incremental validity of tests contributing to their PM/TK factor was demonstrated for the 20% of United States workers performing low complexity jobs by Hunter & Hunter (1984).

The corollary of the assertion that the validity of psychomotor ability increases as job complexity decreases and decreases as job complexity increases, i.e. the contention that the validity of *g* decreases as job complexity decreases and increases as job complexity increases, is well supported in the literature and was discussed in section 3.4.1 (Gutenberg et al, 1983; Hunter as cited in Levine et al, 1996; Hunter 1986; Hunter & Hunter, 1984; Jensen, 1986; Levine et al, 1996).

4.5.2 Why do psychomotor and intelligence measures correlate?

The literature points to three possible explanations for the correlation between psychomotor and intelligence measures:

- Performance on both measures requires the ability to reason. Ree and Carretta (1994) suggest that psychomotor and intelligence factors correlate due to the fact that performing both types of tests require a certain amount of reasoning. In their 1996 article, they extend on this suggestion by contending that the measurement of *g* is unavoidable in all measures of ability due to the fact that when responding to test material, regardless of whether the test requires a psychomotor response, specialized knowledge or verbal skills, reasoning is unavoidable and causes *g* to be measured (Carretta & Ree, 1996b). Jensen (1980) goes as far as to suggest that most tests would be reduced to practical uselessness once *g* has been partialled out (see chapter 3).

- Performance on both measures requires a certain degree of learning to take place. The correlation between *g* and psychomotor ability may also be a reflection of the fact that a certain degree of learning is required to perform well in psychomotor tests (Carretta & Ree, 1996b). Chaiken et al (2000) suggest that the correlation is due to that subjects differ in terms of working memory capacity (*g*) and that it impacts on learning in complex and novel tasks such as those that are often involved in psychomotor tests.

- Measures such as information processing speed, working memory capacity, and reaction time, which often underlie good performance in many psychomotor tests, have also been identified as measures of cognitive ability (Jensen, 1982; 1993; Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992), hence, it should not be surprising that correlations between the two performance on the two measures are found. As far as the abovementioned reaction time component is concerned, research has shown that tests assessing simple reaction time, requiring one fast response to one signal, have shown low positive correlation with *g*, whereas choice reaction time tasks that require the ability to choose from a variety of possible responses and make an appropriate response (eg. pressing a lighted button from a choice of eight buttons) have shown moderate correlations with *g* (Jensen, 1993).

- Reaction time may indeed be a measure of general (cognitive) ability as postulated by the information processing school of thought on intelligence (Fogarty & Stankoff, Hemmelgarn & Kehle, Vernon as cited in Jensen 1986; Jensen, 1982, 1993) and hence a correlation between general (cognitive) ability or *g* and psychomotor ability is to be expected.

4.6 SUMMARY

In this chapter, psychomotor ability was discussed as a predictor of job performance and training criteria. Research findings suggest that

- Psychomotor ability should predict job performance in lower level entry positions requiring operating skills (Bouwer, 1985; Carretta, 1989, 1992; Carretta & Ree, 1994, Duke & Ree, 1996, Hartigan & Wigdor, 1989; Hunter & Burke, 1994; Hunter & Hunter, 1984; Levine et al, 1996; Martinussen, 1996; Schoeman, 1995; Wheeler & Ree, 1997).
- Performance on general cognitive ability and psychomotor ability measures should be positively correlated (Carretta & Ree, 1997; Chaiken et al; 2000; Hunter & Hunter, 1984; McHenry et al, 1990; Rabbitt, Banejeri & Scymnaski, 1989; Ree & Carretta, 1994; Tirre and Raouf, 1998); possibly due to measures such as information processing speed, working memory capacity, and reaction time, which typically underlie good performance in many psychomotor tests, also having been identified as measures of cognitive ability (Jensen, 1982, 1986, 1993; Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992)
- Psychomotor ability should provide significant but small incremental validity beyond general (cognitive) ability (Carretta, 1989; Carretta & Ree, 1994, 1997; Gibb & Dolgin, 1989; Hunter & Hunter, 1984; McHenry et al, 1990; Ree & Carretta, 1994).

The implications for the current study are that there is sound evidence for psychomotor ability predicting both job performance and training criteria for occupations requiring operating skills. Although the research and literature seem to indicate that psychomotor ability is positively correlated with general (cognitive) ability or *g*, there seems to be evidence of psychomotor ability providing significant but small incremental validity beyond general (cognitive) ability, more particularly so in less complex occupations. It would be interesting to see if these findings are confirmed in the current research.

The methodological approach that has been adopted in this research will be outlined in chapter 5.

CHAPTER 5

METHODOLOGY

5.1 INTRODUCTION

This chapter deals with the research design and methodology used in this study.

5.2 RESEARCH DESIGN

The research design was a non-experimental, field setting, concurrent validity design to evaluate the criterion-related validity of a selection battery. The predictor and criterion measures were administered to the load and haul operators at approximately the same time.

Figure 5.1: Variables used in the research design

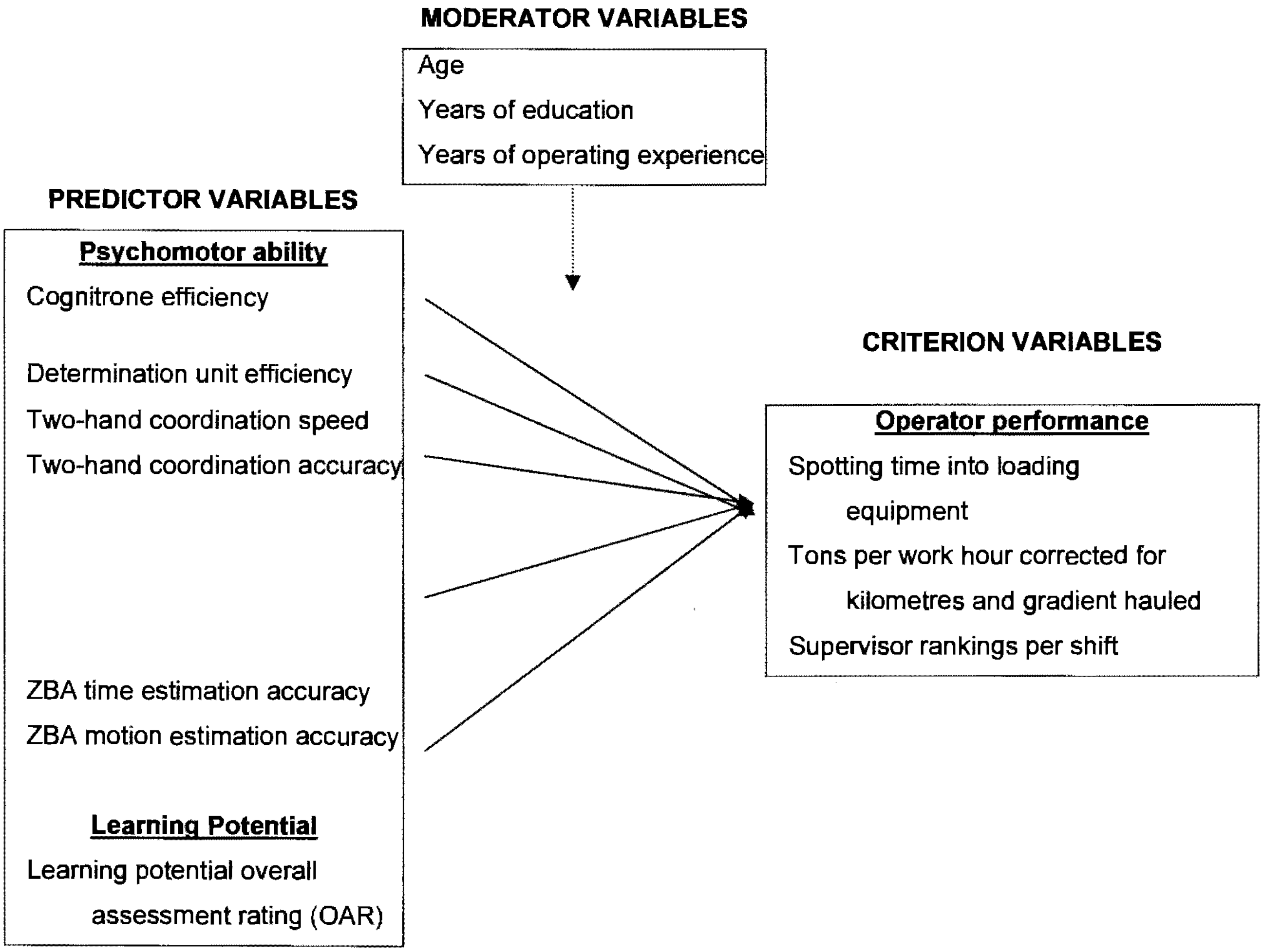


Figure 5.1 illustrates the predictor and criterion variables that were used, as well as potential moderator variables, the effects of which were evaluated and, where significant, controlled for statistically to avoid spuriousness.

5.3 POPULATION AND SAMPLE

The population incorporated 128 haul truck operators at an open-pit mine. In order to maximise the statistical power of the design, no sampling took place, that is the whole population for which predictor and criterion data were available, was included in the study. The demographics of the population are outlined in table 5.1.

Table 5.1: Demographics of the population

Race		Sex		Average age	Average number of years' education	Average number of years' operating experience
Black	White	Male	Female			
128	0	125	3	41.555	8.228	15.127

5.4 RESEARCH METHOD

In terms of the predictor variables, the entire population was assessed on the TRAM 1, (see 5.5.1.2) in groups as per their availability. They were also assessed on the various subtests of the Vienna Test System, (see section 5.5.1.1).

The criterion measures were obtained simultaneously. The Spotting time and Corrected tons hauled variables were obtained from the computerised dispatch system and averaged per operator over a period of three months. The Supervisor ranking variable was obtained for each operator from his or her supervisor (see section 5.5.2.2).

5.5 MEASURING INSTRUMENTS

5.5.1 Predictor variables

Predictors were selected according to the Society for Industrial Psychology's (1992) guidelines for the selection of predictors (see chapter 2), namely

- Predictors should be reliable and valid.
- Predictors should be chosen on the basis that there is a logical, empirical or theoretical reason for them to be included.
- Predictors should be selected based on scientific knowledge rather than expedience.
- Predictors should be as objective as possible.

The following sub-tests of the Vienna Test System were included in the battery: the Cognitrone, Determination unit, Two-hand coordination and *Zeit bewegung abschätzung* (ZBA) sub-tests.

5.5.1.1 The Vienna Test System

Cognitrone

This subtest of the Vienna Test System was designed to assess candidates' ability to concentrate and to adjust their work tempo to different stimuli patterns. The test was included because of its logical conceptual link with haul truck operator performance. The haul truck operator is required to demonstrate sustained concentration throughout the shift, taking into consideration the demands which the continually changing operating environment places on the haul truck operator.

The test is based on Reulecke's concentration theory, which postulates that concentration is made up of three variables, namely energy (concentration consumes energy), function (different actions require different levels of concentration) and precision (the quality of task completion). Individuals involved in tasks requiring concentration must continually regulate the energy, function and precision of their actions. This can be exhausting and cannot be maintained on a continuous basis (Reulecke as cited in Schuhfried, 2000a).

In this test, the candidates are required to indicate whether any of four figures presented on a computer screen is similar to the figure in the test question. The problems are relatively simple and the candidates are required to complete as many questions as possible as accurately as possible. The test duration is 7 minutes.

The test yields various options in terms of differentiated results. In this study "sum correct reactions" was utilised. It gives an indication of performance quality and, to some extent, also provides data on processing speed (Schuhfried, 2000a). In this study, this measure was referred to as "Cognitrone efficiency".

The test has a split-half reliability of 0,95 (Schuhfried, 2000a). Criterion-related validity studies found significant correlations between test results and safety criteria, such as accident frequency and driver errors (Bukasa, Wenninger & Brandstätter as cited in Schuhfried, 2000a; Cale as cited in Schuhfried, 2000a). However, no criterion-related studies are available with the Cognitrone as predictor and operator or driver performance as criterion. A correlation of 0,482 was reported with the Determination unit subtest of the Vienna Test System, discussed in the following sub-section (Wagner as cited in Schuhfried, 2000a). This could be due to both tests tapping similar needs for sustained concentration, efficient information processing and quick reaction time. Negative impacts for age were reported in two studies (Oehlschlagel & Moosbrugger, Wagner as cited in Schuhfried, 2000a).

Determination Unit

This subtest of the Vienna Test System was designed to assess candidates' reaction speed, reactive stress tolerance and ability to demonstrate sustained multiple-choice reactions to rapidly changing stimuli (Schuhfried, 1996). Like the Cognitrone, this test was included because of its conceptual links with haul truck operator performance requirements. This subtest concentrates on the operators' appropriate and fast responses in rapidly changing environments that may involve stress. This is relevant in any situation where there is moving machinery, even more so in the mining environment, where operators of machinery need to cooperate closely with very little margin for error, to achieve production objectives.

As partial cognitive performances, the Determination unit requires the discrimination of colours and acoustic signals, memorisation of the relevant characteristics of stimulus configurations and response buttons and memorisation and application of assignment rules. This test challenges candidates to make continuous sustained and varied reactions to stimuli which are changing rapidly.

The test requires individuals to react to different coloured visual stimuli as well as acoustic stimuli that require either finger or foot responses. The test starts off slowly, gains speed to a very fast response requirement (approximating high stress situations, eg, accident or near-accident situations) and then slows down marginally (approximating the period just after the accident/near-accident). Bearing in mind that the reactions required are not simple, a certain amount of overlap between general cognitive ability (*g*) and the performance in this test should be expected (Jensen, 1993).

The measure used in this research is the "overall results correct" result, which reflects the total number of appropriate timely and delayed responses for the entire test (encompassing the slow, medium and fast phases of the test). This measure will be referred to in this study as "Determination unit efficiency". The "overall results on time" result was omitted, since it registers all responses given within the stimulus presentation time, whether correct or not – it is felt that the "overall results correct" measure is less open to contamination by candidates who respond fast and inappropriately.

The Determination Unit demonstrates internal consistency of 0,99 (Schuhfried, 1996). Various criterion-related validity studies found significant correlations between test results on the Determination unit and driving performance criteria. Correlations have been reported for the Determination Unit results with driving behaviour during a test drive as well as results of a driving test (Klebelsberg & Kallina; Karner & Neuwirth as cited in Schuhfried, 2000). Encouraging correlations between test results on the Determination unit and driving safety criteria have also been reported in terms of frequency of accidents and driver errors (Cale; Wenninger and Brandstatter as cited in Schuhfried, 2000).

Two-hand coordination test

This subtest of the Vienna Test System was designed to assess hand-eye and hand-hand coordination. It was included as a predictor in this study because of the two-hand coordination requirements of haul truck operating activities.

In this test candidates are required to move a cursor on a given track with the aid of two joysticks, one which can only move forward and backward and one that can only move right and left. Hence candidates must use both hands in a coordinated way to move the cursor through the track within acceptable accuracy limits. The track consists of three sections varying in the demands made on the left and right hands, namely an inverted L, V and inverted C shape. Candidates are required to complete four runs.

The results yielded are "total mean duration" (speed dimension) and "total percentage error duration" (accuracy dimension). These variables are referred to as "Two-hand coordination speed" and "Two-hand coordination accuracy", respectively in this study.

Internal consistency (Cronbach's Alpha) of the standardised variables lies between 0,847 and 0,968 (Schuhfried, 2000c). No criterion-related validity studies could be located in the literature review.

Zeit Bewegung Abschätzung (ZBA)

This subtest of the Vienna Test System was designed to assess candidates' ability to correctly estimate and anticipate motion and distance pertaining to moving objects. It was included in this study because of its perceived logical relevance to the "spotting in" activities (see chapter 1), since the better haul truck operators can estimate both distance and direction, the easier it will be for them to "spot into" the loading equipment correctly and quickly on their first attempt, resulting in better performance on the "spotting in" criteria, (see section 5.5.2.1).

In this test a slow-moving dot moves across the screen. At a stage it disappears and candidates are required to indicate both where and when it will hit a line. The test is very similar to the LAMP tests developed in World War II (Gibson as cited in Ree & Carretta, 1998), where an aeroplane appears on the left side of the screen, travels across the screen and then disappears behind a cloud. Candidates are required to estimate when the aeroplane will reappear on the other side of the cloud on the right hand side of the cloud (Ree & Carretta, 1998).

For the purposes of this study only linear progressions were included. Sine wave progressions were omitted, since they relate to operating activities more complex than the operating activities required for the operation of haul trucks.

The results used in this study were "median of deviation time during a linear progression", which is measured in seconds and "median of direction deviation during a linear progression", which is measured in pixels. In the current study these variables were referred to as "ZBA time estimation" and "ZBA motion estimation", respectively.

Internal consistency is 0,92 for the ZBA time estimation measure and 0,69 for the ZBA motion estimation measure (Schuhfried, 2000b).

Validity studies on the previous version of this test (the Distance Estimation Test) indicate that drivers who overestimate distance (i.e. who stop too late) are more problematic than drivers who underestimate distance (Schuhfried, 2000b). This relates to the prediction of safety criteria. No validity studies relating the test to operator or driver performance criteria were reported for the previous version of the test. There are no validity studies available on the current version of the test (ie, the ZBA). No studies on the relative effects of age or any other potential moderator variable have been reported.

5.5.1.2 The TRAM 1Learning Potential Test Battery

General (cognitive) ability or *g* has been found to be a valid indicator of job performance when compared with other potential predictors (Gutenberg et al, 1983; Hartigan & Wigdor, 1989; Hunter, 1986; Hunter & Hunter, 1984; Jensen, 1986; Levine et al, 1996; Schmidt & Hunter, 1981; Schmidt et al, 1988; Schmidt & Hunter, 1998). Furthermore, culture-fairness in the assessment of such cognitive potential is a particularly important issue in the South African context (Taylor, 1994). Therefore, the TRAM 1 from the learning potential school of thought was selected as a predictor in this study (Taylor, 1999).

The TRAM 1 is a learning potential assessment instrument designed for candidates who fall in the illiterate and semi-literate ranges or, put differently, in the educational range of candidates having no formal schooling up to grade 10 (Taylor, 1999). It was included in this study as a culture-free measure of learning potential, which should also give a relatively good indication of fluid intelligence (*gf*) and general cognitive ability or *g* (Taylor, 1994).

The test requires the candidates to translate symbols into other symbols, using a dictionary. The symbols are pictorial or quasi-geometric. The symbols are translated using some underlying rule (eg, opposites - sun, moon; or the symbols being used together - teacup, teapot). In Phase A1 the candidates first complete the translation process by themselves. They are then given a lesson to explain the underlying rules, followed by the completion of Phase A2. They are then given a new test book and a new dictionary to assess the transfer of skills and finally complete a memory test (Taylor, 1999).

Scores are provided on the following dimensions: Automatisation, Transfer, Memory and Understanding, Speed and Accuracy. A composite score (Overall assessment rating) on the testee's overall performance is also generated (Taylor, 1999). Only the overall assessment rating will be used as predictor in this study and will be referred to as "Learning potential OAR".

The various dimensions yield reliability coefficients ranging from 0,62 to 0,95 (Taylor, 1999)

In terms of validity, Taylor (1999) found that the composite score on the TRAM 1 correlated significantly ($r = 0,59$) with academic performance in an ABET course and with academic performance in N1 studies ($r = 0,51$).

5.5.2 Criterion variables

Cascio (1987) contends that "any predictor measure will be no better than the criterion used to establish its validity". In an attempt to balance the shortcomings of both objective (Cascio, 1991; Deary, 2001) and subjective criteria (Anastasi, 1988; Cascio, 1991; Guion, 1987; Thorndike et al, 1991), both objective and subjective criterion measures were used in this study with the focus on the pertinent factors to determine suitability of a criterion as outlined in chapter 2, namely relevance, freedom from bias, availability, (Thorndike et al) and freedom from contamination (Anastasi, 1988; Cascio, 1991; Society for Industrial Psychology, 1992).

No attempt was made to determine the reliability of the various criteria, due to difficulties in obtaining the data required to determine such reliabilities. Although the Society for Industrial Psychology (1992) suggests that it is desirable (but not essential) for criterion measures to be highly reliable, this would have lowered the overall validity that could be obtained (Carretta & Ree, 2000; Cascio, 1991; Schmidt et al, 1976) in the current study.

5.5.2.1 Objective criterion variables

Scheepers (1973) pointed out that an acceptable driving criterion could only be achieved by registering the movements of the vehicle on a continuous basis in much the same way as a flight recorder registers the movements of an aircraft in flight. At the operation in question, every truck is constantly monitored by a Geographical Positioning System (GPS) and the computerised dispatch centre has continuous data on every truck/operator combination in terms of location, tons loaded, status (eg, service/breakdown etc) and other variables relevant to the monitoring of every individual truck-operator combination's performance in the field. This data was used for the collection of the objective criterion data.

Spotting time into loading equipment

This criterion variable was included because the operating skills required for operators to drive from point A to point B are relatively simple. The more complex operating skills are their ability to

- manoeuvre and place the truck in an optimal position for the shovel or front end loader to be able to load, without having to swing further than necessary or overextend the "arm" of the shovel. This process is referred to as "spotting in".
- reverse into the ideal position to tip into the crusher or to "spot into" the safety berm in order to tip at the waste or ore dumps.

It was argued that the more difficult operators found these two skills, the longer it would take them to load or tip; and thus the less productive that particular truck/operator combination would be.

The dispatch system provides data on both these variables per truck/operator combination on a continuous basis. The "spotting in" time criterion is calculated by determining the time between the "full action of the previous truck" and the "first bucket action of the new truck". This criterion is calculated in "overtrucked" situations (ie, when there is at least one truck waiting to be loaded).

Since the measuring equipment at the various tipping/dumping sites lack accuracy, only the time taken to spot in under the bucket was used as criterion measure in this study; in other words, the spotting times into the dumping or tipping sites were omitted. Furthermore, to remove the impact of the different loading equipment (ie, shovel vs front end loaders), only "spotting in" times for the shovel were included. This variable was referred to as "Spotting time" in this study. "Spotting time" was also averaged over a three-month period.

Tons per work hour corrected for kilometres hauled and gradient

Every operator drives different routes to the various pieces of loading equipment, hence they travel a different number of kilometres in order to transport their load (either ore or waste). The routes that they travel also differ substantially in terms of road gradients. It was thus not possible to use tons hauled per hour as an uncorrected indicator of productivity. Some adjustment was required for both kilometres hauled and gradient of the road that is driven.

The following calculation was performed to correct the data extracted from the computerised Dispatch System:

The tons hauled were divided by the hours worked (WH). WH refers to haul truck operating time exclusive of time spent in the queue waiting for loading equipment availability (i.e. WH = operating time – queue time). This was then divided by the kilometres travelled yielding a "Tons corrected for distance hauled" (TCDH) measure. Road gradient was incorporated by multiplying the linear distance travelled by a correction factor calculated by the dispatch system from the road network data that is captured on the system to yield an "Effective flat haul rate" (EFH). Finally, TCDH was corrected to a "Tons per work hour corrected for kilometres hauled and gradient" (TPKH) by multiplying TCDH with EFH (i.e. TPKH = TCDH * EFH). This variable (i.e. TPKH) was referred to in the current study as "Corrected tons hauled". As with the spotting times, these results were also averaged over a three-month period.

5.5.2.2 Subjective criterion variables – supervisor ranking

Supervisor rankings were included as an additional criterion variable. These were obtained in a relatively rudimentary fashion with the aim of corroborating the findings of the objective criteria, bearing in mind the shortcomings of objective criteria (Cascio, 1991; Deary, 2001) stated in chapter 2. The supervisor of every shift was asked to rank the operators on his shift, using the paired comparisons method (Cascio, 1991). The brief was to determine which of every pair of operators, the supervisor would select if the loading conditions were particularly difficult (eg, due to congestion in the loading area or wet road conditions) with number 1 being his best operator, number 2 his second best operator and so forth.

No attempt was made to compare operators amongst shifts because of the unreliability that could potentially be caused by different supervisors not being familiar with the performance of all the operators on all the shifts. Great care was taken to ensure that this criterion was not contaminated by supervisors having access to either predictor or objective criterion scores (Cascio, 1991; Anastasi, 1988; Society for Industrial Psychology, 1992).

5.5.3 Moderator variables

It is common practice in applied psychological research to investigate the effects of biographical variables such as age, sex and education moderating the relationship between the predictor and criterion variables (Cascio, 1991). Anastaasi (1988) stresses the need to include only those variables for which there is evidence of moderating effects. Accordingly, in this study, age, education and experience were investigated as potential moderators in the predictor–criterion relationships:

5.5.3.1 Age

In their meta-analyses, Hunter and Hunter (1984) and Schmidt and Hunter (1998), found validity for age as a predictor of job performance. However, Oehlschlagel and Moosbrugger (as cited in Schuhfried, 1999) and Wagner (as cited in Schuhfried, 1999) found that as age increases, performance in psychomotor tests decrease. It was thus anticipated that age might be a moderating variable; hence it was included as an additional variable.

5.5.3.2 Years of education

Hunter and Hunter (1994), Martinussen (1996) and Schmidt and Hunter (1998) found that correlations of years of education with job performance and training outcomes were relatively small (in the 0,10 to 0,20 range).

It was not certain whether these validity coefficients are relevant in the South African context, where the quality of the schooling system has been of questionable standard and access to the schooling system, for the greater part of the population, has been problematic. Little research is available on this topic. Since the schooling system to which the current South African working population was exposed, did not enforce schooling for all learners, it was anticipated that there would be significantly variability in the South African context compared to the international situation. It was also anticipated that the number of years of schooling that candidates claim to have completed, might give a very rough indication of their current literacy and numeracy level. This, in turn, could impact on the ease with which they acquire the knowledge and skills to perform the job. This was especially relevant due to the considerable variation among the candidates in this study, with education levels varying from illiterate to post-grade 12.

Hence, the researcher was of the opinion that the South African educational situation for the current workforce was sufficiently different from the international educational scenario to warrant the investigation years of education as a potential moderator of job performance of haul truck operators. In the current study, this variable will be referred to as "Years of education".

5.5.3.3 *Years of operating experience*

Martinussen (1996), Schmidt and Hunter (1998), Schmidt et al (1988) and Shinar (1978) indicate that experience may be a moderating variable in a validity study such as this one, hence it was included as a potential moderator variable. For the purposes of this research, years of operating experience were deemed to be experience in operating any mobile machinery including cars, trucks, forklifts or any other mobile machinery, and the variable was referred to as "Years of operating experience".

5.6 DATA ANALYSIS

5.6.1 Descriptive univariate statistics

The means and standard deviations of all the predictor variables, criterion variables and additional variables were calculated. The arithmetic mean is the most common measure of the central tendency of a set of scores (Mc Call, 1990; Howitt & Cramer, 1997). The standard deviation, however, indicates the average amount by which scores differ from the mean (Howitt & Cramer, 1997). The variance of a set of scores is the mean of the squared deviations of the scores from their mean (Rosenthal & Rosnow, 1991). The standard deviation(s) is defined as the positive square root of the variance (McCall, 1990). The size of the standard deviation is an indicator of how much variability there is in the scores for a specific variable (Howitt & Cramer, 1997). In addition, the skewness and kurtosis values of each distribution were calculated and reported and a statistical test of normality was performed. These statistics are briefly discussed below.

The values of skewness and kurtosis should be zero in a normal distribution (Field, 2000). Positive values of the skewness indicate a pile-up of scores on the left (or lower end) of the distribution, while negative values of skewness indicate a pile-up of scores on the right (or higher end) of the distribution (Huysamen, 1987). Positive values of kurtosis indicate a pointed or leptokurtic distribution and negative values of kurtosis indicates a flat or platykurtic distribution (Huysamen, 1987). Skewness and kurtosis values can be converted to z-scores as follows (Field, 2000):

> Zskewness = (S – 0) / SEskewness, where S is the skewness value and SEskewness is the standard error of the skewness value.

Zkurtosis = Square root of |(K – 0)/ SEkurtosis)|, where K is the kurtosis value and SEkurtosis is the standard error of the kurtosis value.

The SPSS programme does not compute these z-values and the researcher had to calculate them. As a rule of thumb, z-values larger than 2 in absolute value, are considered indicative of deviation from normality SPSS, 2000).

Furthermore, the so-called Kolmogorov-Smirnov test of the normality of a distribution was performed .If the p-value of this test statistic is smaller than 0,05, the distribution deviates in some way from normality. The idea is that when this is the case, that the significant skewness and kurtosis values, or even a histogram of the distribution may give more information about the shape of the distribution (Field, 2000).

An important issue to consider was whether the form of the various distributions had implications for the validity of the correlational procedures used in the present study. According to Hays (1963) there is no necessity to make any assumptions at all about the form of the distribution, the variability of Y scores within X columns, or the true level of measurement represented by the scores, in order to employ regression and correlational indices to describe a given set of data. Yet, Hays (1963) emphasizes that assumptions about distributions can become necessary when the concern is with inferring the true linear relations in the population from sample data. Bearing in mind that one of the assumptions underlying the use of multivariate techniques, such as regression, originally planned in the current study (see section 5.6.3 below), is linearity between the independent variables and the dependent variables (Kerlinger & Pedhuzar, 1973; Lewis-Beck, 1980), the researcher initially thought that this concern might be pertinent to the study.

Hays (1963) states further that the form of the distributions of X and Y can affect the range of correlations possible in a set of data if the form of the distributions of X and Y are dissimilar (such as one being positively skewed and the other being negatively skewed, for example).Thus, in order to better understand the nature of the variables in the current study, skewness, kurtosis and normality of distribution statistics were reported.

5.6.2 Correlations and partial correlations

The Pearson Moment correlation coefficient (Hays, 1963) was calculated to determine the degree of relationship amongst the predictor, criterion and additional variables. The statistical significance of all the correlations were determined for different levels of significance.

The following sets of correlations were calculated:

- correlations amongst criteria
- correlations and partial correlations between predictors and criteria
- correlations of moderator variables to predictors
- correlations of moderator variables to criteria
- correlations amongst moderator variables
- correlations amongst predictors

A correlation coefficient varies in value from -1 (a perfect negative value) to + 1 (a perfect positive correlation). Values close to 0 indicated no linear correlation. Correlations are never perfect so that +1 and -1 are never achieved. Suppose a positive correlation of 0,5 between variable X and Y is found. This means that the higher a person's score on X, the higher that same person's score will be on Y. Or conversely, the lower a person's scores on X, the lower that person's score on Y.

If the correlation is negative, however, say -0,5, then the higher a person's score on X, the lower that same person's same score is likely to be on Y. Or, conversely, the lower a person's score on X, the higher that person's score on Y (Huysamen, 1987). When the correlation matrix is large, it sometimes becomes difficult to make sense of the relations between variables and a data-reduction technique, such as factor analysis, may be helpful.

Partial correlations were also calculated between predictor and criteria, controlling for the influence of moderator variables. Partial correlations refer to the correlation between two variables in which the effects of other variables are held constant (Field, 2000).

5.6.3 Stepwise Regression Analysis

In this study, the researcher planned for each of the three job-success variables to be regressed on the predictor and potential moderator variables in an attempt to ascertain the extent to which each of the predictors explained each particular job-success criterion. The moderator variables were also to be included as this would make the control of these variables possible.

The stepwise regression procedure progresses in a stepwise manner. It begins by selecting as a first step the predictor variable which explains the job-success criterion best. This is the predictor with the largest simple correlation with the criterion. In step 2, the stepwise regression procedure correlates the partial correlation for each of the remaining variables, controlling for the predictor already in the equation and then enters the variable with the largest partial correlation as the second variable in the equation. In step 3 the procedure is repeated. At each step, the procedure checks whether previously entered variables may have become less useful and the least useful predictor is removed (Field, 2000).

In the present study, however, the correlation and partial correlations (controlling for biographic variables) between predictor and job-success variables revealed very few significant correlations, hence the stepwise regression analysis proved to be a trivial exercise, as it was already clear from the correlations that at most a single predictor variable would be entered into the equation. Consequently, the regression analysis results are not reported in this study.

5.6.4 Factor analysis

Factor analysis is a statistical technique that is excellently suited to the investigation of the underlying structure of a set of variables (Kerlinger, 1986). Factor analysis is especially useful when the purpose is to uncover dimensions in a set of variables. The rationale is that those variables which refer to the same dimension or share the same dimension, inter-correlate with one another because of this sharing of the same dimension or because they are influenced by the same source. When a group of variables are indicators of a particular dimension (or factor), these variables correlate with this dimension which, in turn, causes the items themselves to be inter-correlated.

Kerlinger (1986, p.569) points out that: "Factor analysis serves the cause of scientific parsimony. It reduces the multiplicity of tests or measures that belong together - which ones virtually measure the same thing, in other words, and how much they do so. It thus reduces the number of variables with which the scientist must cope. It also helps the scientist locate and identify unities or fundamental properties underlying tests and measures."

In the present study the factor analysis programme of the statistical software package SPSS (SPSS, 2000) was used to investigate the underlying structure of the predictor variables. The steps followed in the factor analysis of these predictor variables were as follows:

- computation of a matrix of correlations between the predictor variables
- tests of sampling adequacy
- subjection of the correlation matrix to a Principle Factor Analysis
- decision on the number of factors (dimensions) to be extracted
- oblique rotation of the factor solution to a more interpretable solution utilizing a mathematically calculated criterion - in the present study the criteria used were the direct Oblimin procedure and the Promax rotation (Field, 2000).

Each of these steps is discussed briefly below.

5.6.4.1 The computation of a matrix of correlations amongst the predictor variables

In order to achieve two of the main aims of this study, namely to uncover the relationship between general (cognitive) ability or *g* and psychomotor ability and to determine how general cognitive ability and psychomotor ability interact to predict performance, which underlie H4 and H5, respectively, it was necessary to determine the relationships between the predictors. Correlations between the predictor variables were computed. In an attempt to make more sense of these correlations, it was necessary to employ some data-reduction technique. The researcher decided to use factor analysis.

5.6.4.2 Tests of sampling adequacy

Before researchers actually commence performing a factor analysis on the correlations between variables, they need to test for so-called sampling adequacy. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy assesses whether the partial correlations among variables are sufficiently small. Field (2000) points out that it is usually required that the KMO value (which can vary between 0 and 1) be larger than 0,5 as a large value indicates that patterns of correlations are relatively compact so that factor analysis should yield distinct and reliable factors.

Bartlett's test of sphericity assesses whether the correlation matrix is an identity matrix, which would indicate that the factor model is inappropriate (Field, 2000). The p-value associated with Bartlett's test of sphericity should be small (eg, smaller than 0,05) as this would mean that the correlation matrix is not an identity matrix.

5.6.4.3 Subjecting the correlation matrix to a Principle Factor Analysis

Various methods of factor extraction are available. In the present study, the Principal axis method of factor analysis was used. Field (2000) points out that Principal axis factoring is the preferred method of factor analysis when factor analysis is used in an exploratory manner as is the case in the present study.

5.6.4.4 Deciding on the number of factors (dimensions) to be extracted.

The researcher must decide in advance the number of factors to extract. The following guidelines have been developed to assist researchers in making this decision (Stevens, 1992):

- The so-called Kaiser-Guttman rule specifies that the number of factors to extract is determined by the number of factors with *eigen* values of a size larger than 1,0.
- Use of the "scree"-test to decide the number of factors to extract. For this purpose, the *eigen* values associated with possible underlying factors, are plotted against the factor numbers and Cattell's so-called "scree" test is performed which involves studying the slope of the plotted *eigen* values. The *eigen* value of a factor reflects the amount of variance that the particular factor explains in terms of the data being studied. The larger the *eigen* value of a factor, relative to size of the *eigen* values of the other factors, the more variance that factor explains relative to what the other factors explain. Cattell (1979) contends that one should extract factors that account for the majority of the variability in the original data. An inspection of the *eigen* values usually reveals that the initial drop in the *eigen* values of the first one or two consecutive factors (factors 1 and 2 for instance) is large but grows less and less as more factors are considered. At a particular stage, the drop becomes small and constant so that the shape of the graph now looks like a straight line with a gradual downward slope. This straight-line segment is referred to as a "scree" and there can be more than one. According to Catell, one should note the number of the factor at which the first "scree" begins. If for example, the "scree" starts at factor two, then there is only one factor that underlies the variance.

- In the event that the researcher expects a particular number of factors, such as, say, two factors, the researcher may specify this number of factors to be extracted.
- The researcher often needs to cater for more than one possibility. The researcher may for example extract a single factor, a two-factor and a three factor solution and then judge which solution makes most theoretical sense.

In the present study the Kaiser-Guttman rule indicated a three-factor solution underlying the correlations between the predictors whereas the "scree test" indicated a single factor only. The researcher had expected a two factor solution, namely general cognitive ability and psychomotor ability.

It was hence decided to obtain a single, a two- and a three-factor solution and interpret all of them in order to see which solution made most sense.

5.6.4.5 Rotation of factor solutions

The unrotated factor solution is not interpretable. In the present study it was decided to rotate the solution obliquely (which means that the factors are allowed to correlate with one other). Bearing in mind that no solution is unique, it was decided to use two methods of oblique rotation to verify findings. The researcher was of the opinion that this would allow for a better understanding of the various variables and their inter-relationships.

The two methods used were the direct Oblimin method, which is the method suggested by Field (2000) in cases where there is a preference for oblique rotation, and Promax rotation. In the case of the direct Oblimin method, the delta parameter was set at 0 which ensures that the correlations allowed between the factors are not large. This is also the default setting of the SPSS program. In the case of the Promax rotation the parameter Kappa was set at 4 (Cureton & Mulaik, 1975; Field, 2000), which is also the default value of the programme and allows factors to be moderately inter-correlated (SPSS 2000).

Both these rotation methods yield several factor solution matrices, of which the so-called factor pattern solution matrix is the more important (Cattell, 1979) and is reported in the present study. The values in these factor pattern solution matrices are called factor loadings and indicate the regression of the items on the factors. As a rule of thumb, factor loadings larger than 0,30 in absolute value (Field, 2000) will be considered significant loadings.

5.7 LEVEL OF STATISTICAL SIGNIFICANCE

Most computer programs report a so-called p-value with each statistical test of a null hypothesis. This p-value gives the probability of the result under the null hypothesis. Should the researcher wish to reject the null hypothesis, he or she requires a small p-value. Normally this is the case with statistics such as t- and F-values and correlation coefficients. If this is not the case and one wishes to retain the null hypothesis, then of course, a large p-value is required. In the present study this was, for example, the case when the Kolmogorov-Smirnov test of the normality of a distribution was performed (Field, 2000). The null hypotheses with these tests are that the distribution is normal in shape and the researcher usually would prefer to retain the null hypothesis. Therefore a large p-value is needed.

Normally the researcher sets a so-called level of significance in advance of any statistical tests being calculated and then compares the resulting p-values against this level of significance. If the p-value is smaller then the level of significance, the null hypothesis is rejected. If not, the null hypothesis is not rejected.

Conventionally, the levels 0,05 and 0,01 are used by most researchers as levels of significance for statistical tests performed. In choosing a level of significance for the present research, the following points of view were taken into account:

The Society for Industrial Psychology (1992) holds that a type 2 error is often underemphasized and that Type 1 and Type 2 errors should receive equal attention. The argument, hence is for less stringent significance levels due to researchers in the social sciences being equally concerned with missing a significant result or making a type-II error, as about falsely concluding a significant result (committing a type I error). Hays (1963) and Winer (1971) state that when both types of errors (type I and type II) are equally important, significance levels such as 0,20 (and possibly even 0,30) are more appropriate than the conventionally used 0,05 and 0,01 levels.

At the same time, the argument for more stringent significance levels states that as the total number of statistical tests to be performed on the same sample data increases, the probability of a type I error increases (Hays,1963). One approach to counter this accumulating effect is to set the level of significance smaller for the individual statistical test so as to compensate for the overall type I error effect.

Given the above considerations, it was decided to maintain the conventional 0,05 as the level of significance for each individual statistical test in this study.

5.8 STATISTICAL COMPUTER PACKAGE

All statistical analyses in the present study were computed using the SPSS (Statistical Package for the Social Sciences) statistical package for Windows version 10.1. SPSS Inc. can be contacted at 444 North Michigan Avenue, Chicago, Illinois 60611 (USA), (also see the website http://www.spss.com).

5.9 HYPOTHESES

The methodological approach as outlined in this chapter was deemed adequate for the investigation of the following hypotheses:

H1: Learning potential is a valid predictor of haul truck operator performance in an open-pit mine.

H2: Psychomotor ability is a valid predictor of haul truck operator performance in an open-pit mine.

H3: There is a positive correlation between learning potential and psychomotor ability measures.

H4: The combined use of the TRAM 1 and Vienna Test System measures applied in this research yields a general (cognitive) ability factor (g) and a general psychomotor factor.

H5: Psychomotor ability provides significant incremental validity beyond learning potential in the prediction of haul truck operator performance in an open-pit mine.

It should be noted that the effects of the potential moderator variables, namely age, years of education and years of operating experience, were evaluated for all of the above hypotheses and, where significant, were controlled for statistically to avoid spuriousness.

5.10 SUMMARY

In this chapter the research design and methodology employed to investigate the hypotheses stated in section 5.8 were outlined and discussed.

With these hypotheses in mind, the results obtained through the pursual of the above methodology are presented in chapter 6.

CHAPTER 6

RESULTS, CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

6.1 INTRODUCTION

In this chapter, the findings of this study are discussed. Firstly, the univariate statistics for all the predictors, anticipated moderators and criteria are reported and their relevance discussed, followed by the relevant results for each of the hypotheses. After that, the researcher discusses the conclusion reached and the limitations of this study, before making recommendations for future research.

6.2 POPULATION AND SAMPLE

Due to various practical considerations, all the data in terms of predictor, criterion and moderator variables were not available for the whole population. Table 6.1 below indicates the frequency of cases per predictor, criterion and moderator variable.

Table 6.1: Frequency of cases per predictor, criterion and moderator

Predictors	Cases	Criteria	Cases	Moderators	Cases
Learning potential overall assessment rating	121	Spotting time into loading equipment	110	Age	128
Cognitrone efficiency	127	Tons corrected for distance hauled	94	Years of education	127
Determination unit efficiency	127	Supervisor rankings	110	Years of operating experience	118
Two-hand coordination speed	127				
Two-hand coordination accuracy	127				
ZBA time anticipation accuracy	127				
ZBA motion anticipation accuracy	127				

Bearing in mind the reduced frequencies highlighted above, sample size in certain of the above instances may limit the extent to which adequate statistical power can be achieved to provide a meaningful test of the hypotheses in the present study especially bearing in mind that the current research involves a relatively complex multivariate design in which criterion unreliability and range restriction may be present (Cascio, 1991; Schmidt et al, 1976).

6.3 UNIVARIATE STATISTICS

Univariate statistics are reflected in table 6.2.

According to table 6.2, the distributions of Years of education, Cognitrone efficiency and Determination unit efficiency are negatively skewed (ie, that they are distributed asymmetrically and skewed to the left). This results in the curve piling up at the upper end of the horizontal axis (where scores indicate good efficiencies) with the curve trailing off to the lower end. Although they deviate from normality in terms of demonstrating absolute z-values that are higher than 2, these effects are not severe.

The Spotting time distribution is positively skewed; that is the scores pile up at the lower end indicating short spotting times (which is indicative of good performance) – again the skewness is not severe.

For the ZBA motion estimation accuracy, ZBA time estimation accuracy and Two-hand coordination accuracy, the skewness is severe and positive (which, bearing in mind that low scores are indicative of good performance in all these subtests, is indicative of a large proportion of the sample demonstrating good performance on these subtests). This may also be indicative of a problem in the test design, which does not allow for greater variation in performance amongst candidates. This could thus suggest that there is room for improvement in these particular sub-test designs.

Table 6.2: Univariate statistics

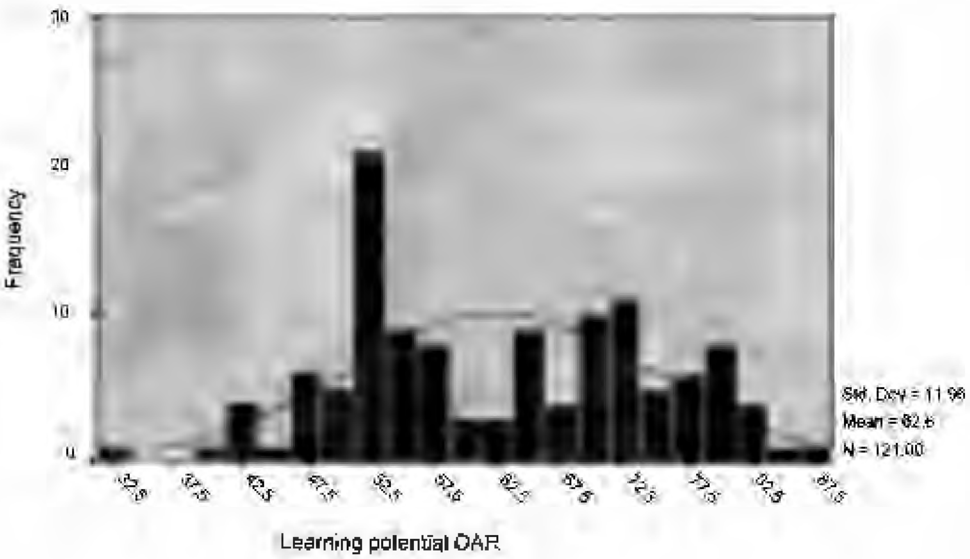
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness			Kurtosis			Kolmogorov-Smirnov(a)		
						Statistic	Std. Error	z-value	Df	Sig.	z-value	Statistic	Df	Sig.
Age	128	25.000	60.000	41.555	8.123	0.204	0.214	0.953	-0.574	0.425	1.162	.068	81	.200(*)
Years of education	127	0.000	13.000	8.228	2.789	-0.513	0.215	-2.386	0.209	0.427	0.700	.107	81	.023
Years of operating experience	118	1.000	35.000	15.127	8.074	0.315	0.223	1.413	-0.375	0.442	0.921	.069	81	.200(*)
Learning potential OAR	121	31.960	86.720	62.637	11.965	0.017	0.220	0.077	-0.957	0.437	1.480	.124	81	.004
Cognitrone efficiency - sum correct reactions	127	83.000	696.000	426.630	117.041	-0.431	0.215	-2.005	0.050	0.427	0.342	.091	81	.093
Determination unit efficiency - overall results correct	127	167.000	519.000	411.961	72.161	-0.839	0.215	-3.902	0.585	0.427	1.170	.076	81	.200(*)
Two-hand coordination speed - seconds per run	127	23.470	123.630	62.326	22.586	0.580	0.215	2.698	-0.339	0.427	0.891	.085	81	.200(*)
Two-hand coordination accuracy - total % error duration	127	0.000	8.780	1.067	1.512	2.712	0.215	12.614	8.679	0.427	4.508	.251	81	.000
ZBA time estimation accuracy - deviation in seconds	127	0.190	4.010	1.096	0.804	1.607	0.215	7.474	2.478	0.427	2.409	.182	81	.000
ZBA motion estimation accuracy - deviation in pixels	127	5.000	155.000	28.126	20.148	3.041	0.215	14.144	14.286	0.427	5.784	.198	81	.000
Spotting time in seconds	110	62.000	108.000	79.073	7.579	0.865	0.230	3.761	1.374	0.457	1.734	.102	81	.036
Corrected tons hauled	94	2042.114	2922.394	2469.941	161.762	0.226	0.249	0.908	0.844	0.493	1.308	.068	81	.200(*)
Supervisor rankings	110	1.000	30.000	14.736	8.206	0.022	0.230	0.096	-1.188	0.457	1.612	.085	81	.200(*)

Predictably, the three variables indicating severe positive skewness, namely ZBA motion estimation accuracy, ZBA time estimation accuracy and Two-hand coordination accuracy, are also severely leptokurtic in terms of kurtosis, with the majority of the candidates obtaining good scores.

These kurtosis observations may be a reflection of the presence of range restriction. It can be argued that since this is a concurrent validity study, and hence that all the subjects are experienced operators, the scores on the above-mentioned variables are inclined to group together on the end of the scale that reflects good performance for each particular measure (Anastasi, 1988; Cascio, 1991).

As expected, all the variables that indicate skewness or kurtosis also demonstrate Kolmogorov-Smirnov test of normality statistics indicating deviations from normality. The exception is Determination Unit efficiency variable, where a skewness z-score of -3,902 is reported. This is above the absolute value threshold of 2 indicating significant skewness – yet the Kolmogorov-Smirnov statistic indicates a normal distribution curve. This finding is difficult to explain.

Figure 6.1: Histogram of the Learning Potential OAR measure



Furthermore the Learning potential variable indicates a significant deviation from normality despite demonstrating no significant skewness or kurtosis. Examination of the histogram emphasizes that there seems to be a bipolar tendency, (see figure 6.1 above). This could possibly be explained by the nature of the TRAM 1 test as a measure of learning potential, where "catching on" (Jensen, 1986, p. 315) and automatisation (Sternberg, 1984) or "improvement with practice" (Jensen, 1986, p. 315) play a significant role. The two peaks of the distribution may very well be indicative of those candidates who "caught on" to the learning tasks and automatised or improved with practice, and those who did not.

6.4 INTERPRETATION ACCORDING TO HYPOTHESES

6.4.1 Hypotheses 1 and 2

H 1 states that learning potential is a valid predictor of haul truck operator performance and H2 states that psychomotor ability is a valid predictor of haul truck operator performance. The following aspects are relevant to these hypotheses: the correlations amongst the various job performance criteria, the correlations between the predictors and criteria, the correlations amongst the proposed moderators and the partial correlations amongst the predictors and criteria (controlling for such moderators).

6.4.1.1 Correlations amongst criteria

Table 6.3 indicates the inter-correlations amongst the job performance criteria utilised in this study. This has significance to H1 and H2 in that it serves as an indicator of the reliability and validity of the various job performance criteria.

There was a significant and strong positive correlation ($r = 0{,}407$) between Spotting time (an objective criterion) and Supervisor rankings (the subjective criterion). This may, to a certain degree, be indicative of the reliability and validity of these two criteria. The argument is that the general spotting times of each operator into the various pieces of loading equipment is actually observable – (more so than the tons hauled per operator, which is inclined to get lost in the realities of different operators driving different haul distances and road gradients; and production being reported per day per shift – rather than per shift per operator). This correlation is all the more significant in that it occurs despite the trend for correlations between objective criterion measures and subjective ratings to typically be low (Cascio, 1991).

The fact that these two criteria correlate could be a relatively good indication that they measure the same dimension relatively reliably. A certain degree of overlap between these two criteria should have been anticipated due to the brief that the supervisors were given when ranking their candidates, using the paired comparisons method (Cascio, 1991). They were asked to determine which of every pair of operators, the supervisor would select if the loading conditions were particularly difficult (eg, due to congestion in the loading area or wet road conditions). The supervisor ranking criterion by definition, then, should be expected to load, to some degree, on the same dimension as the spotting time criterion.

The reliability and validity of the Corrected tons hauled criterion, however, is not supported. There is some indication that the road network data on the Dispatch System was not entirely updated at the time of the study, and that the flat haul rate correction factor that was implemented using this road network data as foundation was thus not totally correct. This criterion, then, was possibly not entirely reliable. It was not possible to determine the exact extent of this unreliability and the consequential effect size on the validity of the criterion. Nonetheless, the fact that there is no significant correlation between the two criteria that do correlate and the Corrected tons hauled criterion could indicate that the Corrected tons hauled criterion was not as reliable as the researcher had hoped. Although the Society for Industrial Psychology (1992) states that it is desirable (but not essential) for criterion measures to be highly reliable, low reliability of the criterion measure places a ceiling on the validity coefficients that are attainable. This may lead to a type 2 error, namely in this case, missing a significant validity coefficient that was present (Carretta & Ree, 2000, Cascio, 1991; Schmidt et al, 1976).

Table 6.3: Correlations amongst criteria

		Spotting time	Corrected tons hauled	Supervisor rankings
Spotting time	Pearson Correlation	1	-.027	.407(**)
	Sig. (2-tailed)	.	.793	.000
	N	110	94	99
Corrected tons hauled	Pearson Correlation	-.027	1	-.016
	Sig. (2-tailed)	.793	.	.883
	N	94	94	85
Supervisor rankings	Pearson Correlation	.407(**)	-.016	1
	Sig. (2-tailed)	.000	.883	.
	N	99	85	110

** Correlation is significant at the 0.01 level (2-tailed).

6.4.1.2 Correlations between predictors and criteria

Table 6.4 indicates the correlations of the predictors selected for this study with the criteria. A one-tailed test of significance was used for the correlations in this table, since the correlations would logically be in a specific direction.

Table 6.4: Correlations of predictors to criteria

		Spotting time	Corrected tons hauled	Supervisor rankings
Learning potential OAR	**Pearson Correlation**	-.132	.144	-.242(**)
	Sig. (1-tailed)	.092	.089	.007
	N	104	90	104
Cognitrone efficiency - sum correct reactions	**Pearson Correlation**	-.200(*)	-.019	-.146
	Sig. (1-tailed)	.018	.427	.064
	N	110	94	110
Determination unit efficiency - overall results correct	**Pearson Correlation**	-.077	-.069	-.029
	Sig. (1-tailed)	.212	.256	.383
	N	110	94	110
Two-hand coordination speed - seconds per run	**Pearson Correlation**	.126	-.035	.049
	Sig. (1-tailed)	.096	.367	.304
	N	110	94	110
Two-hand coordination accuracy - total % error duration	**Pearson Correlation**	-.044	.021	-.052
	Sig. (1-tailed)	.325	.422	.295
	N	110	94	110
ZBA Time estimation accuracy - deviation in seconds	**Pearson Correlation**	.020	.091	.124
	Sig. (1-tailed)	.418	.193	.099
	N	110	94	110
ZBA Motion estimation accuracy - deviation in pixels	**Pearson Correlation**	.028	.000	.029
	Sig. (1-tailed)	.385	.499	.382
	N	110	94	110
** Correlation is significant at the 0.01 level (1-tailed).				
* Correlation is significant at the 0.05 level (1-tailed).				
When N size is smaller than 128, it is because of incomplete pair-wise data.				

There were disappointingly few significant correlations although there were some significant correlations. This was contrary to Bouwer's (1983) findings of no significant correlations for psychomotor measures with job performance for heavy duty truck drivers in a South African study.

Only one significant correlation was reported for the Spotting time criterion, namely a negative correlation (r = -0,20) with Cognitrone efficiency. This means that operators who achieved many correct responses on the Cognitrone, were likely to require fewer seconds to spot in – that is, their spotting performance is better. This seems to provide partial support for H2.

Only one significant correlation was reported for the Supervisor ranking criterion, namely a negative correlation (r = -0,242) with Learning potential. As a lower score on the Supervisor ranking criterion indicates better performance of the haul truck operator (i.e. the number 1 candidate is the best performer on every shift), this means that candidates who performed better on the learning potential measure, were ranked as better operators by their supervisors. This provided evidence in support of H1.

No significant correlations were reported for the Corrected tons hauled criterion. This may be due to the unreliability of the criterion referred to in section 6.4.1.1.

The fact that high correlations of Learning potential with performance were not found should probably not have been surprising, since the haul truck operator position could probably be classified as a lower complexity job. The literature highlights that general (cognitive) ability or *g* is inclined to correlate better with job performance in more complex jobs (Levine et al, 1996; Gutenberg et al, 1983, Hunter, 1986; Jensen, 1986). Nonetheless, various studies have indicated that although the validity of cognitive ability varies across jobs it never approaches zero (Hartigan & Wigdor, 1989; Hunter & Hunter, 1984; Hunter, 1986; Schmidt & Hunter, 1981; Schmidt et al, 1988) – a contention that seems to be partially supported in terms of the Supervisor ranking criterion in the current study.

6.4.1.3 Moderator variable relationships

In order to understand and control for the potential effects of the anticipated moderator variables, namely Age, Years of education and Years of operating experience, on the relationships between the predictors and the job performance criteria, the researcher had to examine the correlations of these variables to the predictors, the criteria and how these anticipated moderator variables were related to one another.

- *Correlations of moderator variables to predictors*

As can be seen from table 6.5 below, there were various significant correlations between the moderator variables and the respective predictors. This would indicate that the moderator variables did, in fact, impact on the performance of the various candidates on the predictors of this study. The moderators, could very well impact on the predictor-criterion relationship in this way.

Table 6.5: Correlations of moderator variables to predictors

Correlations

	Age			Years of education			Years of operating experience		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Learning potential OAR	-.342(**)	0.000	121	.392(**)	0.000	120	-.283(**)	0.002	117
Cognitrone efficiency - sum correct reactions	-0.160	0.072	127	0.090	0.314	126	0.002	0.986	117
Determination unit efficiency - overall results correct	-.337(**)	0.000	127	.328(**)	0.000	126	-.279(**)	0.002	117
Two-hand coordination speed – seconds per run	0.025	0.782	127	-0.122	0.174	126	-.187(*)	0.043	117
Two-hand coordination accuracy - total % error duration	.289(**)	0.001	127	-0.029	0.746	126	.306(**)	0.001	117
ZBA Time estimation accuracy - deviation in seconds	-0.025	0.782	127	-0.068	0.452	126	-0.032	0.735	117
ZBA Motion estimation accuracy – deviation in pixels	0.122	0.171	127	-0.014	0.879	126	0.131	0.159	117
** Correlation is significant at the 0.01 level (2-tailed).									
* Correlation is significant at the 0.05 level (2-tailed).									

More particularly, the Learning potential measure correlated negatively with Age (r = -0,342), positively with Years of education (r = 0,392) and negatively with Years of operating experience (r = -0,283). The fact that the correlation with Years of operating experience was negative, was probably related to the negative correlation of the measure with age, since the more experienced operators were also likely to be older.

In exactly the same way, the Determination unit efficiency measure correlated negatively with Age (r = -0,337), positively with Years of education (r = 0,328) and negatively with Years of operating experience (r = -0,279). In sections 6.4.2 (Table 6.9) for the correlation matrix between the predictors and 6.4.3 (Tables 6.11, 6.12 and 6.13) for the factor analysis findings, both of which indicated that there was a significant overlap in terms of the dimension assessed by the Learning potential and measured by the Determination unit efficiency.

Furthermore, the results indicated a significant positive correlation between the Two-hand coordination accuracy as measured in terms of percentage error duration and Age (r = 0,289) and between the Two-hand coordination accuracy measure and Years of operating experience (r = 0,306). In other words, the older the operators were and the more years' experience they had as operators (a dimension which is also related to age), the more likely it was that they would make more errors the Two-hand coordination test.

The final significant correlation related to Two-hand coordination speed measure. More particularly, there was be a significant negative correlation between the Two-hand coordination speed measure and Years of operating experience (r = 0,187) (ie, candidates who were more experienced, were likely to perform slower on the measure than those with fewer years of experience).

This study did not support the negative correlation between age and performance on the Cognitrone reported by Oehlschlagel and Moosbrugger, and Wagner (as cited in Schuhfried, 2000a).

- *Correlations of moderator variables to criteria*

There were no significant correlations between the moderator variables and the criteria of this study. Table 6.6 below, reflects that the proposed moderators did not moderate in terms of the performance criteria.

Table 6.6: Correlations of moderator variables to criteria

Correlations									
	Age			Years of education			Years of operating experience		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Spotting time	-0.019	0.847	110	-0.015	0.875	110	-0.068	0.497	102
Corrected tons hauled	-0.128	0.218	94	0.185	0.075	94	-0.055	0.611	89
Supervisor rankings	0.005	0.961	110	-0.061	0.527	109	0.036	0.72	104
** Correlation is significant at the 0.01 level (2-tailed).									
* Correlation is significant at the 0.05 level (2-tailed).									

- *Correlations amongst moderator variables*

There were a few significant correlations amongst the moderator variables, namely Age, Years of education and Years of operating experience (see table 6.7 below).

Table 6.7: Correlations amongst moderators

		Age	Years of education	Years of operating experience
Age	Pearson Correlation	1		
	Sig. (2-tailed)	.		
	N	128		
Years of education	Pearson Correlation	-.499(**)	1	
	Sig. (2-tailed)	.000	.	
	N	127	127	
Years of operating experience	Pearson Correlation	.663(**)	-.351(**)	1
	Sig. (2-tailed)	.000	.000	.
	N	118	117	118
** Correlation is significant at the 0.01 level (2-tailed).				

More specifically, Age was negatively correlated with Years of education (r = -0,499), (ie, the older the operators, the less years of education they were likely to have completed). Age was also positively correlated with Years of operating experience (r = 0,663), namely the older the operators, the more years of operating experience they were likely to have had. Because the Years of operating experience was correlated positively with Age and Age was negatively correlated with Years of education completed, it was not surprising that Years of operating experience was correlated negatively with Years of education completed (r = -0,351).

In order to avoid spuriousness, the effects of the various significant correlations discussed in section 6.4.1.3 were controlled for statistically.

6.4.1.4 Partial correlations between predictors and criteria

Table 6.8: Partial correlations of predictors with criteria

		Spotting time	Corrected tons hauled	Supervisor rankings
Learning potential OAR	**Pearson Correlation**	-.1801	.0526	-.2567(*)
	Sig. (1-tailed)	.057	.324	.012
	N	76	76	76
Cognitrone efficiency - sum correct reactions	**Pearson Correlation**	-.2109(*)	-.0428	-.1763
	Sig. (1-tailed)	.032	.355	.061
	N	76	76	76
Determination unit efficiency - overall results correct	**Pearson Correlation**	-.0836	-.1122	.0272
	Sig. (1-tailed)	.233	.164	.407
	N	76	76	76
Two-hand coordination speed - seconds per run	**Pearson Correlation**	.0298	-.0366	.0102
	Sig. (1-tailed)	.398	.375	.465
	N	76	76	76
Two-hand coordination accuracy - total % error duration	**Pearson Correlation**	-.0291	.1171	-.0431
	Sig. (1-tailed)	.400	.154	.354
	N	76	76	76
ZBA Time estimation accuracy - deviation in seconds	**Pearson Correlation**	-.0992	.1129	.0336
	Sig. (1-tailed)	.194	.162	.385
	N	76	76	76
ZBA Motion estimation accuracy - deviation in pixels	**Pearson Correlation**	.0655	.0186	.0075
	Sig. (1-tailed)	.284	.436	.474
	N	76	76	76
** Correlation is significant at the 0.01 level (1-tailed).				
* Correlation is significant at the 0.05 level (1-tailed). When N size is smaller than 128, it is because of incomplete pair-wise data				

Table 6.8 indicates the partial correlations of the predictors with the criteria, controlling for Age, Years of education and Years of operating experience.

The effect of the various correlations of the moderator variables with one another and the various predictors was a negligible reduction in the reported correlations. Hence, the partial correlations reflected in table 6.8 were slightly higher for the two significant correlations (i.e. $r = -0{,}2109$ and $r = -0{,}2567$, respectively) than those reflected in table 6.3 (i.e. $r = -0{,}200$ and $r = -0{,}242$ respectively) in which the moderator variables were not taken into account.

Furthermore it should be noted that, although not quite significant at the 0,05 level, the partial correlations indicated a near-significant correlation (0,057) of this criterion with Learning Potential. The relationship shows promise for further research and could support H1.

In summary, H1 was partially verified in the sense that there was a significant correlation between the Learning potential measure and Supervisor ranking criterion and a near-significant ($p = 0{,}057$) partial correlation between the Learning potential measure and the Spotting time criterion. No significant correlations were reported for the Corrected tons hauled criterion. This may be due to the unreliability of the criterion referred to in section 6.4.1.1.

H2 was confirmed only for the Cognitrone efficiency predictor. The extent to which this measure loads on a psychomotor factor as opposed to a general (cognitive) ability factor of *g* still needs to be confirmed (see section 6.4.3).

6.4.2 Hypothesis 3

Hypothesis 3 relates to the relationship between the predictors. More specifically, the hypothesis is that there is a positive correlation between learning potential and psychomotor ability.

6.4.2.1 Correlations amongst predictors

Table 6.9 below reflects the correlation matrix for the predictors used in the current study. As can be seen, there were several significant correlations amongst the predictor data.

Table 6.9: Correlations amongst predictors

		Learning potential OAR	Cognitrone efficiency - sum correct reactions	Determination unit efficiency - overall results correct	Two-hand coordination speed - seconds per run	Two-hand coordination accuracy - total % error duration	ZBA time estimation accuracy - deviation in seconds	ZBA motion estimation accuracy - deviation in pixels
Learning potential OAR	Pearson Correlation	1						
	Sig. (2-tailed)	.						
	N	121						
Cognitrone efficiency - sum correct reactions	Pearson Correlation	.285(**)	1					
	Sig. (2-tailed)	.002	.					
	N	120	127					
Determination unit efficiency - overall results correct	Pearson Correlation	.425(**)	.285(**)	1				
	Sig. (2-tailed)	.000	.001	.				
	N	120	127	127				
Two-hand coordination speed - seconds per run	Pearson Correlation	-.096	-.166	.038	1			
	Sig. (2-tailed)	.298	.062	.669	.			
	N	120	127	127	127			
Two-hand coordination accuracy - total % error duration	Pearson Correlation	-.236(**)	-.093	-.297(**)	-.228(**)	1		
	Sig. (2-tailed)	.009	.301	.001	.010	.		
	N	120	127	127	127	127		
ZBA time estimation accuracy - deviation in seconds	Pearson Correlation	-.076	-.105	-.174(*)	.000	-.097	1	
	Sig. (2-tailed)	.412	.239	.050	.996	.277	.	
	N	120	127	127	127	127	127	
ZBA motion estimation accuracy - deviation in pixels	Pearson Correlation	-.151	-.075	-.141	-.010	.150	.080	1
	Sig. (2-tailed)	.099	.403	.114	.910	.092	.374	.
	N	120	127	127	127	127	127	127

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Learning potential appeared to correlate with three of the predictors designed to assess psychomotor ability, namely Cognitrone efficiency, Determination Unit efficiency, and Two-hand coordination accuracy. For all of these predictors, the learning potential score was correlated with psychomotor predictor scores indicating good performance. Learning potential was correlated positively with Cognitrone efficiency (r = 0,285) measured in terms of the sum of correct reactions; positively with Determination Unit efficiency (r = 0,425) measured in terms of overall results correct; and negatively with Two-hand coordination accuracy (r = -0,236) reflecting a smaller percentage of error time. This supported the H3 hypotheses, namely that learning potential and psychomotor ability were positively correlated. This finding also corroborates findings indicating that *g* and psychomotor ability are positively correlated (Carretta & Ree, 1997; Chaiken et al; 2000; Hunter & Hunter, 1984; McHenry et al, 1990; Rabbitt, Banejeri & Scymnaski, 1989; Ree & Carretta, 1994; Tirre and Raouf, 1998). Moreover, this finding adds to the literature in terms of the fact that the measure of general cognitive ability used in this study was a learning potential measure.

These positive correlations of learning potential with psychomotor measures were possibly due to both the learning potential measure and the psychomotor measures loading on *g*. This supposition is supported in the research literature (Carretta & Ree, 1997; Chaiken et al, 2000; Ree & Carretta, 1994, Vernon as cited in Walsh & Betz, 1990). Jensen, (1982, 1986, 1993), Kranzler and Jensen, (1991), Kyllonen and Christal, (1990) and Miller and Vernon, (1992) maintain that this is due to measures such as information processing speed, working memory capacity, and reaction time, which typically underlie good performance in many psychomotor tests, also being measures of cognitive ability. This possible loading of both the learning potential measure and the psychomotor measures on *g*, will have to be confirmed for this study via factor analysis (see section 6.4.3).

When the correlations amongst psychomotor predictors were investigated, the following significant correlations were found. The Determination Unit correlated negatively with both the Two-hand coordination accuracy (r = -0,297) and ZBA time estimation accuracy (r = -0,174) measures. Lower scores on both the Two-hand coordination accuracy and ZBA time estimation accuracy measures, indicated good psychomotor ability, this translated into good performance on the Determination Unit measure correlating with good performance on the Two-hand coordination accuracy and ZBA time estimation accuracy predictor measures. In the same way, the Cognitrone efficiency was positively correlated with Determination Unit efficiency (r = 0,285). This supports Wagner's (as cited in Schuhfried, 2000a), findings of a correlation of 0,482 between the Determination unit and the Cognitrone.

All the abovementioned inter-correlations could indicate a higher order psychomotor factor (Carretta & Ree, 1997; Chaiken et al, 2000; Ree & Carretta, 1994, Vernon as cited in Walsh & Betz, 1990). This had to be confirmed via factor analysis. (See section 6.4.3).

A further correlation was the predictably negative correlation between Two-hand coordination speed and Two-hand coordination accuracy (r = -0,228). In other words, the faster candidates move the cursor through the various sections of the track, the more errors they are likely to make.

6.4.3 Hypotheses 4

According to H4, the combined use of the TRAM 1 and Vienna Test System measures applied in this study, yields a general (cognitive) ability or *g* factor and a general psychomotor factor (Carretta & Ree, 1997; Chaiken et al, 2000; Ree & Carretta, 1994). As noted in section 6.4.2, the investigation of the correlations amongst the predictors pointed to the possibility of these higher order factors operating in the current study.

6.4.3.1 Tests of sampling adequacy

To assist with the interpretation of the correlations in tables 6.4, 6.8 and 6.9, factor analysis was performed on the inter-correlations, since it is excellently suited to uncovering the underlying structure of a set of variables (Kerlinger, 1986).

As a first step, the correlation matrix was tested for sampling adequacy and the null hypothesis of an identity matrix. See section 5.6.4.2 in chapter 5 for a discussion of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The results of these measures are given in Table 6.10 below.

Table 6.10: Results of KMO and Bartlett's tests of sampling adequacy for factor analysis

Kaiser-Meyer-Olkin measure of sampling adequacy		.626
Bartlett's test of sphericity	**Approx. Chi-Square**	73.684
	Df	21
	Sig.	.000

The KMO value is above 0,5 which is acceptable. The Bartlett's test of sphericity is highly significant which is what is required for a factor analysis to make sense. It was thus decided to proceed with the factor analysis of the correlation matrix.

6.4.3.2 Principal factor analysis

Factor analysis was performed on the data, using the Principal Axis method of factor analysis, since it is the preferred method when factor analysis is used on an exploratory basis, as in this study (Field, 2000).

6.4.3.3 Deciding on the number of factors to be extracted

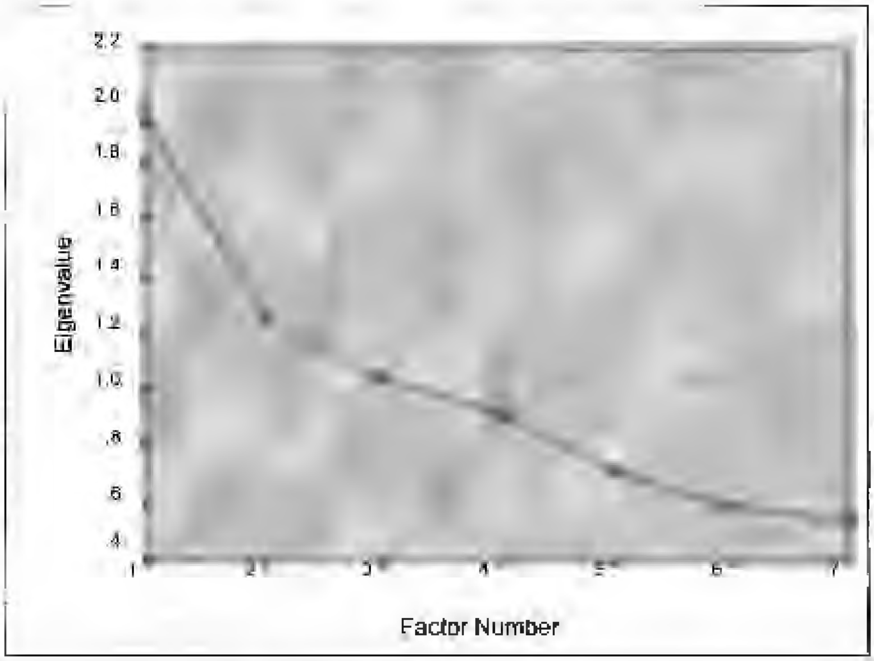
Various strategies are available to the researcher in deciding on how many factors to extract, (see chapter 5, section 5.6.4.4). The first strategy referred to is the Kaiser-Guttman criterion, which specifies that the number of factors to extract is determined by the number of factors with *eigen* values of size larger than 1,0 (Stevens, 1992).

The second strategy referred to is the use of the "scree" test to decide the number of factors to extract.

In order to implement both of these strategies, the *eigen* values of the various factors were plotted against the number of factors that can be extracted. The *eigen* value of a factor indicates its importance in terms of the overall variability in the data for which the factor accounts. A plot of the *eigen* values against the number of the factors can thus serve as an important guide to the number of factors that need to be extracted (see section 5.6.4.4 in chapter 5). Such a plot is given in figure 6.2 below. From this line-graph, the following observations can be made:

- There are three factors with *eigen* values larger than 1,0 – hence the Kaiser-Guttman criterion suggests that three factors should be extracted.
- The line appears more or less straight from factor 2 onwards. The so-called "scree" (Cattell, 1979) therefore, starts at factor 2. According to Cattell's (1979) so-called "scree" test, this would indicate that only one factor should be extracted.

Figure 6.2: Plot of *eigen* values of possible factors against the number of the factor (the so-called "scree" plot)



To complicate matters further, the researcher, anticipated two factors, namely *g* and psychomotor ability (Carretta & Ree, 1997; Ree & Carretta, 1994; Vernon as cited in Walsh & Betz, 1990). Hence, it was decided to extract a single-factor, a two-factor and a three-factor solution and to study these factor solutions in order to make sense of the inter-relatedness of the predictors.

6.4.3.4 One-factor solution

As a first step, a principal axis factor analysis was performed on the correlation matrix and a single factor extracted. The factor loadings are given in table 6.11.

From the data below, it would appear that four of the seven predictors load higher that 0,30 in absolute value on a single factor. The remaining three predictors have small loadings.

Table 6.11: Principal axis single-factor solution of correlations between predictors

	Single Factor
Determination unit efficiency - overall results correct	.718
Learning potential OAR	.614
Two-hand coordination accuracy - total % error duration	-.410
Cognitrone efficiency - sum correct reactions	.382
ZBA Motion estimation accuracy - deviation in pixels	-.242
ZBA Time estimation accuracy - deviation in seconds	-.106
Two-hand coordination speed - seconds per run	.032
Extraction method: Principal Axis factoring.	

Both cognitive (or intelligence or learning potential) tests and psychomotor tests are inclined to load on *g*, and this may very well constitute a general (cognitive) ability or *g* factor (Carretta & Ree, 1997; Chaiken et al; 2000; Hunter & Hunter, 1984; McHenry et al, 1990; Rabbitt, Banejeri & Scymnaski, 1989; Ree & Carretta, 1994; Tirre and Raouf, 1998). This is due to measures such as information processing speed, working memory capacity, and reaction time, which typically underlie good performance in many psychomotor tests, also being measures of cognitive ability (Jensen, 1982, 1986, 1993; Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992). The very high loading of the Determination unit efficiency variable on the postulated *g* factor, was due to the fact that the responses required by the measure were not simple (ie, there are many varied responses to various stimuli) (Jensen, 1993).

The evidence presented for the one-factor solution supported H4 in terms of both the learning potential measure (TRAM 1) and the psychomotor measures (Vienna Test System subtests) loading on a general cognitive ability or *g* factor.

6.4.3.5 Two-factor solution

Two-factor solutions were also obtained by Principal Axis factoring. In an attempt to achieve interpretable solutions, the solutions were rotated. Two rotation methods were used in an attempt to verify findings and to gain a better understanding of the various variables and their inter-correlations. The two methods of rotation used were:

- Promax rotation using the default setting of the SPSS program (Kappa = 4)
- Oblimin rotation using the default setting of the SPSS program (Delta = 0)

The pattern matrices are reported in Table 6.12 below. The factor loadings in these matrices are standardized regression coefficients and not correlations as would have been the case in orthogonal rotated solutions.

According to Table 6.12 below the Promax rotation gave a slightly different result from that obtained through the Oblimin rotation.

Table 6.12: Principal axis two-factor solution of correlations between predictors

	Promax rotated solution		***Oblique rotated solution***	
	Factor 1	**Factor 2**	**Factor 1**	**Factor 2**
Learning potential OAR	.639	.046	.658	.006
Cognitrone efficiency – sum correct reactions	.528	-.133	.473	.157
Determination unit efficiency - overall results correct	.498	.312	.628	-.242
ZBA Motion estimation accuracy - deviation in pixels	-.168	-.121	-.218	.095
ZBA Time estimation accuracy - deviation in seconds	-.148	.059	-.124	-.064
Two-hand coordination accuracy - total % error duration	-.081	-.586	-.323	.517
Two-hand coordination speed - seconds per run	-.354	.542	-.130	-.509
Extraction Method: Principal Axis Factoring.				

Both solutions suggest the following:

- Factor 1 consists of the Learning potential OAR, Cognitrone efficiency and Determination unit efficiency.
- Factor 2 consists of the Two-hand coordination tests.
- The ZBA tests are poorly represented by the two factors.

The two solutions differ in that the Promax solution places Two-hand coordination speed with factor 1 while the Oblimin solution does not. The reverse situation holds for Two-hand coordination accuracy as the Oblimin rotation place this test with factor 1 but the Promax rotation does not. A further difference is that the Promax solution places Determination unit efficiency with factor 2 while the Oblimin solution does not.

The examination of the nature of the three tests included in factor 1 in the two-factor solution (regardless of the factor analysis rotation used) suggests that factor 1 is again indicative of a general (cognitive) ability or *g* factor, which supported H4 in terms of the existence of a general (cognitive) ability or *g* factor.

If the nature of the three (Promax) or two (Oblimin) tests that load on factor 2, namely the Two-hand coordination accuracy and Two-hand coordination speed (for both methods of rotation) and the Determination unit efficiency (for the Promax solution) is examined, these tests, comparatively speaking, make more demands on motor coordination.

In the case of the Two-hand coordination tests, fine motor coordination is required to move a cursor on a given track with the aid of two joysticks, one which can only move forward and backward and one that can only move right and left. Hence the candidate must use both hands in a coordinated way to move the cursor through the track within acceptable accuracy limits. The track consists of three sections varying in the demands made on the left and right hands (ie, an inverted L, V and inverted C shape). In the case of the Determination unit, gross motor coordination is required to make quick hand or foot responses to different coloured visual stimuli as well as acoustic stimuli. The test starts off slowly, gains speed to a very fast response requirement and then slows down marginally.

Accordingly, the Two-hand coordination accuracy, the Two-hand coordination speed and the Determination unit efficiency variables may load on a general psychomotor factor (Carretta & Ree, 1997; Chaiken, 2000; Ree & Carretta, 1994; Vernon as cited in Walsh & Betz, 1990).

In summary, the two-factor solution discussed above supported H4, namely that the combined use of the TRAM 1 and Vienna Test System measures applied in this study, yields a general (cognitive) ability or *g* factor and a general psychomotor factor (Carretta & Ree, 1997; Ree & Carretta, 1994; Vernon as cited in Walsh & Betz, 1990).

The fact that the psychomotor measures seem to load more on the general (cognitive) factor (*g*) than on the postulated psychomotor factor may be evidence for:

- Reaction time, indeed, being a measure of general (cognitive) ability as postulated by the information processing school of thought on intelligence (Fogarty & Stankoff, 1982; Hemmelgarn & Kehle, 1984; Jensen, 1982, 1993; Vernon as cited in Jensen 1986) and furthermore that measures such as information processing speed, working memory capacity, and reaction time, which typically underlie good performance in many psychomotor tests are also measures of cognitive ability (Jensen, 1982, 1986, 1993; Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992).
- The contention that the measurement of *g* is unavoidable in all measures of ability due to the fact that when responding to test material, regardless of whether the test requires a psychomotor response, specialized knowledge or verbal skills, reasoning is unavoidable and causes *g* to be measured (Carretta & Ree, 1996a and b). In fact, according to Jensen (1980), most tests would be reduced to practical uselessness once *g* has been partialled out.
- The redundancy of psychomotor tests, since there is not much incremental utility gain from their use, especially in the case of entry level jobs. Hunter and Hunter (1984) maintain that, when used in isolation, there is no other predictor with as much validity as general (cognitive) ability or *g*.

6.4.3.6 Three-factor solution

Three-factor solutions were also obtained by Principal Axis factoring and the solutions rotated in order to achieve interpretable solutions using the following two rotation methods:

- Promax rotation using the default setting of the SPSS program (Kappa = 4)
- Oblimin rotation using the default setting of the SPSS program (Delta = 0)

The pattern matrices (containing standardized regression coefficients) are reported in table 6.13 below.

Table 6.13: Principal axis three-factor solution of correlations between predictors

	Promax rotated solution			Oblique rotated solution		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Determination unit efficiency – overall results correct	.670	-.102	-.101	.666	-.099	-.042
Learning potential OAR	.669	.169	-.026	.643	.154	.004
Cognitrone efficiency - sum correct reactions	.462	.231	-.067	.419	.223	-.056
Two-hand coordination accuracy - total % error duration	-.393	.297	-.253	-.482	.327	-.302
ZBA Motion estimation accuracy - deviation in pixels	-.237	.055	.064	-.233	.050	.041
Two-hand coordination speed - seconds per run	-.103	-.705	-.138	-.034	-.662	-.072
ZBA Time estimation accuracy - deviation in seconds	-.167	.148	.577	-.077	.072	.536
Extraction Method: Principal Axis factoring.						

From Table 6.13, it is again clear that in both rotations the predictors Learning potential OAR, Cognitrone efficiency and Determination unit efficiency, load strongly on factor 1, thereby supporting H4 in terms of the existence of a general (cognitive) ability factor.

The position of the Two-hand coordination test variables in terms of factor 1 is now somewhat clearer. The predictor, Two-hand coordination accuracy, loads negatively (above 0,30 in absolute value) on factor 1 in both the Promax and the Oblimin rotated solutions, indicating that few errors (ie, good performance) on the Two-hand coordination accuracy correlated significantly with good performance on the three measures that appear to be the basis of factor 1.

In terms of factor 2, the position in terms of the Determination unit test, is also somewhat clearer in that neither of the rotations indicate a significant loading on factor 2, which was hypothesized to be the psychomotor factor in the two-factor solution discussed in section 6.4.3.5 above. Only the Two-hand coordination test variables are hypothesized to load on this factor in the three-factor solution (the Two-hand coordination accuracy variable has a 0,297 standardised regression coefficient with this factor in the Promax solution, which is just short of the 0,30 criteria set for inclusion, but is included in the Oblimin solution with a 0,327 standardised regression coefficient). Bearing in mind that there are now only two variables loading on this factor, there probably is not enough evidence to make any interpretations about their loading on a single factor. Being cautious in the interpretation of the data, this would probably not constitute sufficient evidence for the existence of a general psychomotor factor in the current study.

In terms of factor 3, Table 6.13, again indicates that the Promax rotation gave a slightly different result to that obtained through an Oblimin rotation. Both solutions suggest that the ZBA time estimation variable loads on this factor. The rotations differ in that the Oblimin rotation includes the Two-hand coordination accuracy variable, whereas the Promax rotation does not.

Only two variables load on this factor in the Oblimin rotation and only one variable loads on this factor in the Promax rotation, hence no interpretations could be made about the variables that load on this factor. However, the ZBA time estimation variable involves the accurate prediction of the speed at which an object is moving and deviations are measured in milliseconds. The Two-hand coordination accuracy variable, on the other hand, involves the assessment of the accuracy of small, precise movements. It can conceptually be argued that it makes sense for these two variables to load on the same factor, which could conceptually be linked to something analogous to a psychomotor precision factor. This may be worthy of further investigation if future studies.

To summarise the findings relating to H4:

All the factor analysis solutions support the hypothesis that the combined use of the TRAM 1 and Vienna Test System measures applied in this study, yields a general (cognitive) ability or *g* factor.

Although the two-factor solution indicates that the combined use of the TRAM 1 and Vienna Test System measures applied in this research yields a general psychomotor factor, the three-factor solution does not.

Hence the current research provides partial support for H4.

The fact that the psychomotor measures seem to load more on the general (cognitive) factor (*g*) than on the postulated psychomotor factor may be evidence that:

- Reaction time is, indeed, a measure of general (cognitive) ability as postulated by the information processing school of thought on intelligence (Fogarty & Stankoff, 1982; Hemmelgarn & Kehle, 1984; Jensen, 1982, 1993; Vernon as cited in Jensen 1986) and furthermore that measures such as information processing speed, working memory capacity, and reaction time, which typically underlie good performance in many psychomotor tests are also measures of cognitive ability

(Jensen, 1982, 1986, 1993; Kranzler & Jensen, 1991; Kyllonen & Christal, 1990; Miller & Vernon, 1992).

- The measurement of *g* is unavoidable in all measures of ability due to the fact that when responding to test material, regardless of whether the test requires a psychomotor response, specialized knowledge or verbal skills, reasoning is unavoidable and causes *g* to be measured (Carretta & Ree, 1996 a and b). In fact, according to Jensen (1980) most tests would be reduced to practical uselessness once *g* has been partialled out.
- Psychomotor tests may be redundant, since there is not much incremental utility gain from their use, especially in the case of entry level jobs where it is argued that, when used in isolation, there is no other predictor with as much validity as general (cognitive) ability or *g* (Hunter & Hunter, 1984).

6.4.4 Hypothesis 5

According to H5, psychomotor ability provides significant, but small, incremental validity beyond learning potential in the prediction of haul truck operator performance in an open-pit mine (Carretta, 1989; Carretta & Ree, 1994, 1997; Gibb & Dolgin, 1989; Hunter & Hunter, 1984; McHenry et al, 1990; Ree & Carretta, 1994).

It was planned to do stepwise multiple regression analysis in order to provide support for this hypothesis. Each of the three job-success variables was to have been regressed on the predictor and moderator variables in an attempt to ascertain the extent to which each of the predictors explained each particular job-success criterion. The moderator variables were also to have been included as this would have made the control of these variables possible.

Due to the fact that, in the present study, the correlation and partial correlations (controlling for the moderator variables) between predictor and job-success variables reveal very few significant correlations, the stepwise regression analysis proved to be a trivial exercise, as it was already clear from the correlations that, at most, a single predictor variable would be entered into the equation. Hence the regression analysis results are not reported in this study.

6.5 CONCLUSION

From the hypotheses of this study, the following was concluded relating to H1 and H2:

- Only one significant correlation is reported for the Spotting time criterion, namely a negative correlation with Cognitrone efficiency (r = -0,200 which improves to r = -0,2109 when the moderator variables are controlled for statistically). This means that those operators, who achieve many correct responses on the Cognitrone, are likely to require fewer seconds to spot in – i.e. their spotting performance is better. This supports H2 and the validity of psychomotor performance in lower level entry positions requiring operating skills (Carretta, 1989, 1992; Carretta & Ree, 1994, Duke & Ree, 1996, Hartigan & Wigdor, 1989; Hunter & Burke, 1994; Hunter & Hunter, 1984; Levine et al, 1996; Martinussen, 1996; Schoeman, 1995; Wheeler & Ree, 1997). There was however, no support for H2 for any of the other predictor-criterion relationships.
- Only one significant correlation is reported for the Supervisor ranking criterion, namely a negative correlation with Learning potential (r = -0,242 which improves to r = -0,2567 when the moderator variables are controlled for statistically). Due to the fact that a lower score on the Supervisor ranking criterion indicates better performance of the haul truck operator (i.e. the number 1 candidate is the best performer in every shift), this means that candidates who performed better on the learning potential measure, were ranked as better operators by their supervisors. This provides evidence in support of H1 and the validity of general (cognitive) ability to predict job performance and training criteria (Gutenberg et al, 1983; Hartigan & Wigdor, 1989; Hunter, 1986; Hunter & Hunter, 1984; Jensen, 1986; Levine et al, 1996; Schmidt & Hunter, 1981; Schmidt et al, 1988; Schmidt & Hunter, 1998).
- No significant correlations were reported for the Corrected tons hauled criterion. This may be due to the unreliability of the criterion referred to in section 6.4.1.1.
- There is a significant and strong positive correlation between Spotting time and Supervisor rankings (r = 0,407), which provides encouraging support for the reliability and validity of two of the criteria relating to H1 and H2. However, there are indications of criterion unreliability in terms of the criterion, Corrected tons hauled. However, the extent of the unreliability cannot be determined.
- The partial correlations of the predictors with the criteria, controlling for age, years of education and years of operating experience indicate that there is a minimal moderating effect of the variables identified as possible moderators in this study, leading to a very slight reduction in the

correlations between the predictors and criteria. Hence, the partial correlations are slightly higher for the two significant correlations (i.e. the Learning potential – Supervisor ranking correlation improves from r = -0,200 to r = -0,2109 and the Cognitrone efficiency – Spotting time correlation improves from r = -0,242 to r = -0,2567). Furthermore, although not quite significant at the 0,05 level, the partial correlations indicate a near-significant correlation (0,057) of this criterion with Learning Potential. The relationship shows promise for further research and may be indicative of possible further evidence in support of H1.

In terms of H3 there is a positive correlation between learning potential and psychomotor ability, the following conclusions were reached:

- Performance on learning potential measures seems to correlate with performance on psychomotor measures, indicating that *g* and psychomotor ability are positively correlated (Carretta & Ree, 1997; Chaiken et al; 2000; Hunter & Hunter, 1984; McHenry et al, 1990; Rabbitt, Banejeri & Scymnaski, 1989; Ree & Carretta, 1994; Tirre and Raouf, 1998) and adds to the literature in terms of the fact that the measure of general cognitive ability used in this study, is a learning potential assessment instrument.
- The inter-correlations found amongst the various psychomotor measures may be indicative of a higher order psychomotor factor (Carretta & Ree, 1997; Chaiken et al, 2000; Ree & Carretta, 1994, Vernon as cited in Walsh & Betz, 1990). This is partially confirmed by the two-factor rotation solution referred to in section 6.4.3.5.

According to H4, the combined use of the TRAM 1 and Vienna Test System measures applied in this study, yields a general (cognitive) ability factor (g) and a general psychomotor factor (Carretta & Ree, 1997; Chaiken et al, 2000; Ree & Carretta, 1994). The following conclusions pertaining to H4, were reached:

- All the factor analysis solutions seem to support the hypotheses that the combined use of the TRAM 1 and Vienna Test System measures applied in this study, yields a general (cognitive) ability factor (*g*).
- Although the two-factor solution seems to provide evidence that the combined use of the TRAM 1 and Vienna Test System measures applied in this study, yields a general psychomotor factor, the three-factor solution does not.

According to H5, psychomotor ability provides significant incremental validity beyond learning potential in the prediction of haul truck operator performance in an open-pit mine. Due to the fact that, in the present study, the correlation and partial correlations (controlling for biographic variables) between predictor and job-success variables revealed very few significant correlations, the stepwise regression analysis proved to be a trivial exercise, as it was already clear from the correlations that, at most, a single predictor variable would be entered into the equation. Hence, no results pertaining to this hypothesis were reported in this study.

6.6 LIMITATIONS

6.6.1 Design limitations

No correction for range restriction was attempted (Lawley, 1943; Thorndike as cited in Carretta & Ree, 2000). It can be argued that because this was a concurrent validity study, and all the subjects were hence experienced operators, limited variability in the performance on the psychomotor measure could be anticipated. The fact that there was severe kurtosis present in some of the psychomotor predictors, namely the ZBA motion estimation accuracy, ZBA time estimation accuracy and Two-hand coordination accuracy measures, supported the notion of the operation of range restriction in this study.

More variability would have been possible if it had been a predictive design, where candidates had been taken in from outside the company and trained from scratch (Anastasi, 1988). The presence of range restriction in this study could have led to the validity estimates reported being much lower than they were in reality (Anastasi, 1988; Carretta & Ree, 2000; Martinussen, 1996; Ree et al, 1994). In other words, it could have led to a type 2 error operating in this study, that is that significant validity coefficients could have been missed, despite their being present in reality.

An important limitation of this study was that no attempt was made to control for motivational aspects of operator performance. This is a typical limitation of concurrent validity designs (Anastasi, 1988; Cascio, 1991; Jennings as cited in Cascio, 1991), which may have had a significant impact on the results. In terms of the objective criteria, namely Spotting time and Corrected tons hauled, operator performance was assessed over a three-month period based on performance records captured on a routine basis by the computerised dispatch system. Operators were hence not aware of their performance being assessed. It is thus conceivable that the predictor variables relate to driving ability, whilst the criterion variables utilised in the study relate to driving behaviour. Driving behaviour can be affected by a myriad of motivational factors that may have impacted on the correlations found in the study (Cascio, 1991, Deary, 2001, Shinar, 1978; Thorndike et al, 1991). This aspect would also have had the effect of depressing the validity coefficients reported in this study thereby making the possibility of a type 2 error in this study a distinct possibility.

6.6.2 Limitations pertaining to the criteria

As in many validation studies, more specifically in the psychomotor field (Griffin & Koonce; 1996), the criterion problem (Anastasi, 1988; Cascio, 1991; Thorndike & Hagen, 1991; Walsh, 1989) was a relevant factor in the current study.

Due to the fact that the road network data and the consequent correction factor that was built in to correct for road gradient (i.e. the "effective flat haul rate") were suspect, criterion unreliability is suspected for the criterion, Corrected tons hauled. No attempt was made to determine the extent of such unreliability, bearing in mind that there was insufficient data available to attempt to estimate its effect. Hence, the statistical correction of the validity coefficients could also not be attempted – a procedure referred to as correction for attenuation (Carretta & Ree, 2000, 2001; Society for Industrial Psychology, 1992). Although the Society for Industrial Psychology (1992) states that it is desirable (but not essential) for criterion measures to be highly reliable, low reliability of the criterion measure, places a ceiling on the validity coefficients that are attainable. This may also lead to a type 2 error, namely in this case, missing a significant validity coefficient that is, in fact, present (Carretta & Ree, 2000, Cascio, 1991; Schmidt et al, 1976).

A further limitation relating to the criteria was that the subjective criterion, namely the Supervisor ranking criterion was obtained in a relatively rudimentary fashion using the paired comparisons ranking method (Cascio, 1991), the focus being more on attempting to corroborate the findings of the objective criteria, bearing in mind the shortcomings of objective criteria highlighted in chapter 2 (Cascio, 1991, Deary, 2001, Shinar, 1978; Thorndike et al, 1991). Although the strong correlation between good performance on the Spotting time criterion and the Supervisor ranking criterion seems to be indicative of good reliability and validity of the Supervisor ranking criterion, it is conceivable that even better reliability and validity as well as a better understanding of the performance levels of the various operators could have been obtained, had a more robust performance measure such as behaviourally anchored rating scales been used (Cascio, 1991; Latham & Saari, 1984). The fact that no attempt was made to rank operators across shifts was a further limitation of the Supervisor ranking criterion.

6.7 RECOMMENDATIONS

6.7.1 Correcting design limitations

The researcher recommends that an attempt be made to corroborate and extend on the findings of this study, by using a predictive design in cross-validation (Cascio, 1991). This would minimize the range restriction, which is typical of concurrent designs and is hypothesized to be active in the current study, mostly due to the levels of experience that the candidates have, impacting on their scores on the predictors and hence confounding the relation between predictor and criterion.

In order to attempt to limit the effects of motivational levels on the objective performance criteria measures (Anastasi, 1988; Cascio, 1991; Jennings, as cited in Cascio, 1991; Shinar, 1978), whether in a concurrent or predictive design, it is recommended that all subjects be informed that there will be systematic monitoring of each individual's operating performance over a three-month period. Although this will not totally cancel out motivational effects, it may have the effect of limiting their impact.

If better predictive or concurrent validity coefficients were to be found in these studies, it would be possible to better explore H5 in terms of the incremental validity of psychomotor ability beyond learning potential (Carretta, 1989; Carretta & Ree, 1994, 1997; Gibb & Dolgin, 1989; Hunter & Hunter, 1984; McHenry et al, 1990; Ree & Carretta, 1994).

The existence of the general (cognitive) ability or *g* factor reported in this study as well as the existence of any other factors beyond *g* (e.g. a possible psychomotor factor or psychomotor precision factor) can also be confirmed.

6.7.2 Correcting limitations pertaining to the criteria

In order to minimize the effect of the criterion problem (Anastasi, 1988; Cascio, 1991; Thorndike & Hagen, 1991; Walsh, 1989), the reliability and validity of both the objective and the subjective criteria need attention.

In terms of the objective criteria, an obvious necessity, is the professional technical auditing of the computerized dispatch system before any cross-validation studies should be undertaken in order to ensure that the objective criteria are reliable.

In terms of the subjective criteria, it is recommended that a behaviourally anchored rating scale (Cascio, 1991; Latham & Saari, 1984) be developed to ensure more detailed, quantifiable, reliable and valid supervisor assessments per operator.

6.7.3 Exploring safety as a criterion

In chapter 1, it was emphasized that there were three criteria against which the effectiveness of selection procedures could be measured, namely productivity, safety and length of time taken to complete the training programme. Due to the limited scope of this study, the only criterion dealt with here was productivity. No attempt was made to link the predictors to either safety or training criteria, both of which hold promise for further research.

Focusing on the safety criterion, safety is a crucial consideration in the mining industry, and arguably enjoys even more focus than productivity. The Mines Health and Safety Act (1996) very much places the onus on the employer to ensure that safety risks are minimized. The consequences of not doing so are onerous. It hence makes sense that the mining industry should be interested in identifying potential operators, who display the least risk from a safety perspective.

It can be argued that the Vienna Test System subtests conceptually relate better to safety than to productivity measures.

- The Cognitrone (Schuhfried, 2000a) yields data in terms of the candidate's ability to concentrate and to adjust his or her work tempo to different stimuli patterns. Hence, all the Vienna Test System subtests have a strong conceptual link to safety.
- The Determination unit (Schuhfried, 1996) specifically focuses on the operator's capacity to make appropriate and fast responses in rapidly changing environments that may involve stress. The test starts off slowly, gains speed to a very fast response requirement (approximating high stress situations e.g. accident or near-accident situations) and then slows down marginally (approximating the period just after the accident/near-accident).

- The Two-hand coordination subtest (Schuhfried, 2000c) is specifically focused on the candidate's hand-eye and hand-hand coordination, which is conceptually related to safety in terms of small movements that need to be made during the spotting in process in tight loading conditions.
- The Distance estimation time and motion measures (Schuhfried, 2000b), for example, attempt to identify those candidates who are least likely to underestimate distance and hence stop too late or cut in front of moving machinery when it is not safe to do so.
- The Signal detection subtest (Schuhfried, 2000d), the results of which were not used in this validation exercise, yields data on the candidate's ability to sustain concentration levels in monotonous conditions.

Despite its importance, there are very few good validity studies relating to safety criteria. This may be due to the difficulty of obtaining safety criteria that are reliable (accidents are generally speaking, relatively infrequent events and near-misses are seldom reliably reported).

In order to add to the literature in this crucial field, it is recommended that the safety variables pertinent to Haul truck operator performance should be measured using a simulator. The predictor data from the current research can then be correlated with simulator performance focusing on safety variables, such as the number of times during the simulation exercise that the operator stopped too late; underestimated the speed of approaching vehicles; or displayed risk behaviour (e.g. driving too fast, overtaking on an incline etc.).

The simulator can also be used to train and assess operators in terms of operating in conditions that cannot be practiced or assessed "live", such as accident situations (break failures, tyre bursts, near-misses, slippery road conditions, operating in tight conditions) and so on. This would make a significant contribution to both the research literature and practical safety in the mining industry.

6.8 SUMMARY

This study yielded encouraging results for the hypotheses that learning potential is a valid predictor of operator performance in an open-pit mine, with a significant correlation being reported for Learning potential with the Supervisor ranking criterion and a near-significant partial correlation being reported for Learning potential with the Spotting time criterion.

The hypothesis that psychomotor ability is a valid predictor of operator performance in an open-pit mine, was only partially supported with only one significant correlation being reported, namely the Cognitrone efficiency predictor with the Spotting time criterion. No support was found for the correlation between any of the psychomotor measures predicting any of the other performance criteria.

The current results also provide support for the well-researched contention that there is a positive correlation between general (cognitive) ability or *g* and psychomotor ability and adds to the research literature in terms of the fact that the measure used to assess general (cognitive) ability, is a learning potential measure.

The hypothesis that the combined use of the TRAM 1 and Vienna Test System measures utilised in this study yields a general (cognitive) factor (*g*) and a general psychomotor factor, was partially supported. Factor analysis provided relatively consistent evidence for a general (cognitive) ability factor (*g*) underlying both the TRAM 1 and Vienna Test System measures. The evidence for the existence of a general psychomotor factor was less convincing.

Bearing in mind that very few significant correlations were reported between the predictors and the criteria utilised in this study, the hypothesis relating to psychomotor ability providing significant, but small, incremental validity beyond learning potential, could not be investigated.

Overall, evidence could only be found for the partial validity of the learning potential and psychomotor ability measures utilised in this study. Thus the process cannot be seen to be totally compliant with section 8 of the Employment Equity Act, (55 of 1998), (although it is conceivable that it is probably more compliant than the unstructured interview measure that was in operation before its implementation).

There are two possible implications of these findings. The first implication is that it would make sense to complete further research to correct the limitations in this study which may have led to a type 2 error operating in the research and also to pursue the investigation of utilizing a safety measure (possibly using a simulator) as criterion in future studies. In this way it is hoped that the economic utility of using the TRAM 1 and Vienna Test System as selection measures may be verified more consistently.

If this is not the case, it points to the second possible implication of the findings of this study, namely the need for further research into more valid predictors of operator performance. In this way compliance with the Employment Equity Act will be ensured and the mining industry will potentially harness the economic gains that will be obtainable if more valid selection measures were to be employed.

REFERENCES

Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288-318.

Aiken, L.R. (1997). *Psychological Testing and Assessment.* (9th ed.). Boston: Allyn & Bacon.

Anastasi, A. (1988). *Psychological Testing.* (2nd ed.). New York: MacMillan.

Anastasi, A. & Urbina, S. (1997). *Psychological Testing.* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Anglo American Platinum Corporation Limited. (2000). *Anglo Platinum Annual Report 2000.* Johannesburg: Bastion.

Austin, J.T. & Hanisch, K.A. (1990). Occupational attainment as a function of abilities and interests: a longitudinal analysis using project TALENT data. *Journal of Applied Psychology, 75*, 77-86.

Avolio, B.J. & Waldman, D.A. (1990). An examination of age and cognitive test performance across job complexity and occupational types. *Journal of Applied Psychology, 75,* 43-50.

Baddeley, A.D. (1986). *Working memory.* Oxford: Clarendon Press.

Binning, J.F. & Barret, G.V. (1989) Validity of personnel decisions: a conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology. 74* (3), 478 – 494.

Bouwer, D.T. (1984). The role of psychological and psychometrical factors as predictors of job success of heavy vehicle drivers. Unpublished M.A dissertation. Johannesburg: Rand Afrikaans University.

Bransford, J.P., Delclos, V.R., Vye, N.J., Burns, M.S. & Hasselbring, T.S. (1987). State of the art in future directions. In C.S. Lidz (Ed.), *Dynamic assessments* (pp. 479-496). New York: Guilford.

Budoff, M. (1987). The validity of learning potential assessment. *Human Psychopharmacology Clinical and Experimental,* 13 (2), 64-69.

Campion, M.A., Pursell, E.D. & Brown, B.K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview, *Personnel Psychology, 41,* 25-42.

Carretta, T.R (1989).USAF pilot selection and classification. *Aviation Space Environment, 60*, 46-49.

Carretta, T.R. (1992). Recent developments in US Air Force pilot candidate selection and classification. *Aviation, Space, and Environmental Medicine, 63*, 1112-1114.

Carretta, T.R. & Ree, M.J. (1994). Pilot-candidate selection method: sources of validity. *International Journal of Aviation Psychology, 4,* 103-117.

Carretta, T.R. & Ree, M.J. (1995). Air Force Officer Qualifying Test Validity for predicting pilot training performance. *Journal of Business and Psychology, 9,* 379-88.

Carretta, T.R. & Ree, M.J. (1996a). Factor structure of the Air Force Officer Qualifying Test: Analysis and Comparison. *Military Psychology, 8*, 29-42.

Carretta, T.R & Ree, M.J. (1996b). US Airforce Pilot Selection Tests: What is measured and what is predictive? *Aviation, Space and Environmental Medicine, 67* (3), 279 – 283.

Carretta, T.R. & Ree, M.J. (1997). Expanding the nexus of cognitive and psychomotor abilities. *International Journal of Selection and Assessment, 5* (3), 149-158.

Carretta, T.R. & Ree, M.J. (2000). General and specific cognitive and psychomotor abilities in personnel selection: The prediction of training and job performance. *International Journal of Selection and Assessment, 8* (4), 227-236

Carretta, T.R. & Ree, M.J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9* (4), 325-335.

Carrol, J.B. (1993). *Human cognitive abilities: a survey of factor analytic studies.* New York: Cambridge University Press.

Cascio, W.F. (1991). *Applied Psychology in Personnel Management* (3rd ed.). Virginia: Reston.

Cattell, R.B. (1971). *Abilities: their structure, growth and action.* Boston: Houghton Mifflin.

Cattell R B. (1979). *The scientific use of factor analysis in behavioural and life sciences.* New York: Plemum.

Chaiken, S.R., Kyllonen, P.C. & Tirre, W.C. (2000). Organisation and components of psychomotor ability. *Cognitive Psychology, 40,* 198-226.

Chamber of Mines. (2002). *South Africa's Mineral Wealth.* Retrieved July 2, 2002 from the World Wide Web: http://www.bullion.org.za/bulza/educatn/career.htm

Chamber of Mines. (2001). *Industry overview 1999.* Retrieved October 2, 2001 from the World Wide Web: http://www.bullion.org.za/bulza/Platinum.htm

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurements, 7,* 249-253.

Constitution of South Africa Act 108 of 1996. *Government Gazette,* No. 17678. Cape Town: Government Printer.

Cox, R.H. (1988). Utilization of psychomotor screening for USAF pilot candidates: Enhancing predictive validity. *Aviation Space and Environmental Medicine, 59* (7), 640-645.

Cureton, E.E. & Stanley A.M. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika, (40)* 2, 449-460.

Deary, I.J. (2001). Individual differences in cognition: British contributions over a century. *British Journal of Psychology, 91* (1), 217-238.

De Jager, J.J. & Van der Walt, H.S. (1993). *The reliability and validity of the stress tolerance test of the Vienna Test System for operators of earth moving equipment at a diamond mine.* Pretoria: Human Science Research Council.

De Jager, J.J. & Van der Walt, H.S. (1997). *The validity of the Vienna Test System for operators of moving equipment at a coal mine.* Pretoria: Human Science Research Council.

De Jager, J.J. (1997). *The external validity of the Vienna Test System by predicting the effectiveness and productiveness of truck drivers transporting fluid cargo.* Pretoria: Human Science Research Council.

Drucker, P.F. (1992). *Managing for the Future.* New York: HRD Press.

Duke, A.P. & Ree, M.J. (1996). Better candidates fly fewer training hours: another time testing pays off. *International Journal of Selection and Assessment, 4* (3), 115-121.

Dunnette, M.D. (1966). *Personnel selection and placement.* Belmont, CA: Wordsworth.

Employment Equity Act 55 of 1998. *Government Gazette,* No. 19370. Cape Town: Government Printer.

Field, A. (2000). *Discovering Statistics using SPSS for Windows.* London: Sage.

Fleischman, E.A. (1988). Some new frontiers in personnel selection research. *Personnel Psychology, 41,* 679 – 701.

Fleischman, E.A. & Quaintance, M.K. (1984). *Taxonomies of human performance: the description of human tasks.* Orlando, FL: Academic Press.
Gibb, G.D. & Dolgin, D.L. (1989). Predicting military flight training success by a compensatory tracking task. *Military Psychology, 1*, 235-240.
Giliomee, H. (1989). Sharing the company: Co-determination in the 1990's. *Indicator South Africa, 6* (4), 63-67.
Griffin, G.R. & Koonce, J.M. (1996). Review of psychomotor skills in pilot selection research of the U.S. military services. *International Journal of Aviation Psychology, 6* (2), 125-147.
Guion, R.G. (1974). Open a new window: validities and values in psychological measurement. *American Psychologist, 29,* 287 – 296.
Guion, R.M. (1987). Changing views for personnel selection research. *Personnel Psychology, 40,* 199-213.
Gustafsson, J.E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research,* 28, 407-34.
Gutenberg, R.L., Arvey, R.D., Osburn, H.G. & Jeanneret, P.R. (1983). Moderating effects of decision-making/information processing job dimensions on test validities. *Journal of Applied Psychology, 8*, 602-608.
Hakamies, B.L. & Henriksson, P. (1999). Cohort effects in older drivers' accident type distribution: Are older drivers as old as they used to be? *Transportation Research Part F: Traffic Psychology and Behaviour, 2* (3), 131-138.
Hartigan, J.A. & Wigdor, A.K. (1989). *Fairness in employment testing: validity generalization, minority issues and the General Aptitude Test Battery.* Washington, DC: National Academy.
Hays, W.L. (1963). *Statistics.* New York: Holt, Rinehart & Winston.
Howitt, D. & Cramer, D. (1997). *An introduction to Statistics for Psychology" A Complete Guide for Students.* London: Prentice Hall.
Humphreys, G.L. (1984). General intelligence. In C.R. Reynolds & T.R. Brown (Eds), *Perspectives on Bias in Mental Testing*. New York: Plenum.
Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge and job performance. *Journal of Occupational Behaviour, 29,* 340-362.
Hunter, D.R. & Burke, E.F. (1994). Predicting aircraft pilot training success: A meta-analysis of published research. *The International Journal of Aviation Psychology, 4,* 297-313.
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. (1990). *Methods of Meta-analysis.* Newburypark, CA: Sage.
Hunter, J.E & Schmidt, F.L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, public policy and law, 2,* 447-472.
Huysamen, G.K. (1987). *Introductory Statistics and Research Design for the Behavioural Sciences Volume 1.*: Cape Town: H&R Academica.
Jensen, A.R. (1980). *Bias in Mental Testing.* NY: The Free Press.
Jensen, A.R. (1982). Reaction time and psychometric *g.* In H.J. Eysenck (ed.), *A Model for Intelligence.* New York: Springer.
Jensen, A.R. (1986). *g*: Artifact or reality? *Journal of Vocational Behaviour, 29*, 301 – 331.
Jensen, A.R. (1993). Why is reaction time correlated with psychometric *g*? *Current Directions in Psychological Science, 2*, 53 – 56.

Kaplan, R.M. & Saccuzzo, D.P. (1997). *Psychological Testing: Principles, Applications and Issues.* (4th ed.). California: Brooks/Cole.
Kantor, J.E. & Carretta, T.R. (1998). Aircrew selection systems. *Aviation Space and Environmental Medicine, 59,* A32-A38.
Kerlinger, F.N. & Pedhuzar, E.J. (1973). *Multiple regression in behavioural research.* New York: Holt, Rinehart & Winston.
Kerlinger, F.N. (1986). *Foundations of behavioural research* (3rd ed.). Orlando, Florida: Holt, Rheinhart & Winston.
Kim J. & Mueller C.W. (1978). *Statistical methods and practical issues*. London: Sage.
Kim, J. & Mueller, C.W. (1978). *Introduction to factor analysis: What it is and how to do it.* London: Sage.
Klineberg, O. (1980). Historic perspective: cross-cultural psychology before 1960. In H.C. Trianis & W.W. Lamber (Eds), *Handbook of Cross-cultural Psychology: Volume 1 Perspectives.* Boston: Allyn & Bacon.
Kranzler, J.H. & Jensen, A.R. (1991). The nature of psychometric *g*: unitary process or a number of independent processes? *Intelligence, 15,* 397–422.
Kyllonen, P.C. & Christal, R.E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389 – 433
Labour Relations Act 66 of 1995. *Government Gazette*, Vol 366, no. 16861. Cape Town: Government Printer.
Larson, G.E. & Saccuzzo, G.E. (1989). Cognitive correlates of general intelligence: toward a process theory of *g. Intelligence,* 13, 5-31.
Latham, G.P. & Saari, L.M. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology, 69,* 569-573.
Latham, G.P. & Wexley, K.N. (1981). Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley.
Laughton, P. (1990). The dynamic assessment of intelligence: a review of three approaches. *School Psychology Review, 19,* 459-470.
Levine, E.L., Spector, P.E., Menon, S., Narayanan, L. & Cannon-Bowers, J.A. (1996). Validity generalization for cognitive, psychomotor, and perceptual tests for craft jobs in the utility industry. *Human Performance, 9* (1), 1-22.
Lewis-Beck, M.S. (1980). *Applied regression: an introduction.* Beverley Hills: Sage.
Lopes, A., Roodt, G & Mauer, R. (2001). The predictive validity of the Apil-B in a financial institution. *Journal of Industrial Psychology, 27* (1), 61-69.
Luthans, F. (1995). *Organizational Behavior*. New York: McGraw-Hill.
Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology, 6* (1), 1-20.
Mauer, K.F. (2000a). *An appraisal of psychological test and other assessment devices used in the various divisions of AMPLATS.* Pretoria: Author.
Mauer, K.F. (2000b). *Psychological test use in South Africa*. Retrieved July 25, 2002 from the world wide web: http://www.pai.org.za/board.html.
Mc Call, R.B. (1990). *Fundamental Statistics for Behavioural Sciences.* (5th ed.). Fort Worth: Harcourt Brace Javanovich.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. & Mauer, S.D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599-616.

Mc Henry, J.J., Hough, L.M., Toquam, J.L., Hanson, M.A. & Ashworth, S. (1990). Project A validation results: The relationship between predictor and criterion domains. *Personnel Psychology, 45,* 335-353.

Meyer, W.F., Moore, C. & Viljoen, H.G. (1989). *Personality theories from Freud to Frankl.* Johannesburg: Lexicon.

Miller, L.T. & Vernon, P.A. (1992). The general factor in short term memory, intelligence and reaction time. *Intelligence, 16,* 5 – 29.

Mines Health and Safety Act 29of 1996. *Government Gazette*, No. 17242. Cape Town: Government Printer.

Mouton, J, (1997). Positivism. In J. Snyman (Ed), *Conceptions of Social Enquiry* (2nd ed.), pp. 1-35. Pretoria: Human Science Research Council.

Murphy, K.R. & Davidshofer, C.O. (1994). *Psychological testing: Principals and Applications.* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Nell, V. (1991). Standardising the WAIS-III and the WMS-III for South Africa: legislative, psychometric and policy issues. *South African Journal of Psychology, 29* (3), 128-138.

Olea, M.M. and Ree, M.J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology,* 79, 845-851.

Oosthuizen, F.J. (1976). The development of a psychological selection battery for the selection of vehicle drivers (National Service level) in the South African Defence Force. Unpublished M.A dissertation. Bloemfontein: University of the Orange Free State.

Rabbit, P., Banerji, N. and Szymanski, A. (1989). Space fortess as an IQ test? Predictions of learning and of practiced performance in a complex interactive video-game. *Acta Psychology, 71,* 243-57.

Ree, M.J. & Carretta, T.R. (1994). The correlation of general cognitive ability and psychomotor tracking tests. *International Journal of Selection and Assessment, 2,* 209-216.

Ree, M.J. & Carretta, T.R. (1996a). Central role of g in military pilot selection. *International Journal of Aviation Psychology, 6* (2), 111-123.

Ree, M.J. & Carretta, T.R. (1998). Computerised testing in the United States Air Force. *International Journal of Selection and Assessment, 6* (2), 82-106.

Ree, M.J. & Carretta, T.R., Earles, J.A.& Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology, 79,* 298-301.

Ree, M.J. & Earles, J.A. (1992). Intelligence is the best predictor of job performance. *Current directions in psychological science, 1,* 86-89.

Reed, T.E. & Jensen, A.R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16,* 259 – 272

Resnick, L.B. & Neches, R. (1984). Factors affecting individual differences in learning ability: In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence, (Vol 2).* Hillsdale(NJ): Erlbaum.

Robertson, I.T., Gratton L & Rout, I.J. (1990). The validity of situational interviews for administrative jobs. *Journal of Organisational Behaviour, 11,* 69-76.

Rosenthal, R. & Rosnow, R.L. (1991). *Essentials of Behavioural Research: Methods and Data Analysis.* (2nd ed.). New York: McGraw-Hill.
Rust, J. & Gulumbok, S. (1999). *Modern psychometrics: the science of psychological assessment.* (2nd ed.). London: Routledge.
Salgado, J.F. (2000). Personnel selection at the beginning of the new millennium. *International Journal of Selection and Assessment, 8* (4), 191-193.
Scheepers, J.M. (1973). Identifying the accident-prone driver. *Robot, 69,* 15-21.
Schermerhorn, J.R., Hunt, J.G. & Osborn, R.N. (2000). *Organizational Behavior.* New York: Wiley.
Schmidt, F.L. & Hunter, J.E. (1981). Employment testing: Old theories and new research findings. *American Psychologist, 36,* 1128-1137.
Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin, 124* (2), 262 – 274.
Schmidt, F.L., Hunter, J.E., McKenzie, R.C. & Muldrow, T.W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology, 64*, 609 – 626.
Schmidt, F.L., Hunter, J.E., Outerbridge, A.N. & Goff, S. (1988). The joint relation of experience and ability with job performance: A test of three hypotheses. *Journal of Applied Psychology, 73,* 46-57.
Schmidt, F.L., Hunter, J.E. & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: a red herring. *Journal of applied Psychology, 66* (2), 166-185.
Schmidt, F.L., Hunter, J.E. & Urry, V.W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61,* 473-485.
Schoeman, C.J.P. (1995). Die utiliteit van die VTS as keuringsinstrument vir treindrywers. Unpublished MCom dissertation. University of Pretoria.
Schuhfried, G. (1996). *Vienna Determination Test. Release 25.00.* Moedling: Author.
Schuhfried, G. (2000a). *Cognitrone. Release 26.01.* Moedling: Author.
Schuhfried, G. (2000b). *Time-Movement Anticipation. Release 22.00.* Moedling: Author.
Schuhfried, G. (2000c). *Two-Hand Coordination. Release 22.03.* Moedling: Author.
Schuhfried, G. (2000d). *Signal Detection. Release 21.00.* Moedling: Author.
Shinar, D. (1978). *Psychology on the road – the human factor in traffic safety.* New York: John Wiley.
Shobris, J.G. (1996). The anatomy of intelligence. *Genetic, social & general psychology Monograms, 122* (2), 133-158.
Snow, R.E., Kyllonen, P.C. & Marshalek, B. (1984). The topography of ability and learning correlations. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 2.* Hillsdale, NJ: Erlbaum.
Society for Industrial Psychology. (1992). *Guidelines for the validation and use of personnel selection procedures.* Pretoria: Society for Industrial Psychology.
Statistical package for the Social Sciences (SPSS) for Windows version 10.1. (2001). [Computer software]. Chicago: SPSS.
Sternberg, R.J. (1984). Towards a triarchic theory of human intelligence. *Behavioural and Brain Sciences, 7,* 269-287.
Stevens, J. (1992). *Applied Statistics for the Social Sciences.* (2nd ed.). Hillsdale, NJ: Erlbaum.

Suzuki, L.J., Meller, P.J. & Ponterotto, J.G. (1996). *Handbook of multicultural assessment.* San Francisco: Jossey-Bass.
Taylor, T.R. (1992). New perspectives on fairness in selection. Unpublished report.
Taylor, T.R. (1994). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology,* 24, 184-193.
Taylor, T.R. (1999). Administrator's manual for TRAM-1 Battery: Assessment of Learning Potential for Semi-literates. Johannesburg: Aprolab.
Thorndike, R.L. & Hagan, G.P. (1977). *Measurement and Evaluation in Psychology and Education.* (4th ed.). New York: Wiley.
Thorndike, R.M., Cunningham, G.K., Thorndike, R.L. & Hagen, E.P. (1991). *Measurement and evaluation in psychology and education.* (5th ed.). New York: Macmillan.
Tirre, W.C. & Raouf, K.K. (1998). Structural models of cognitive perceptual-motor abilities. *Personal and Individual Differences, 24,* 604-614.
Vernon, P.A. (1986). Relationships between speed-of-processing, personality, and intelligence. In C. Bagley & G.K. Verma (Eds), *Personality, cognition and values.* London: Macmillan.
Vernon, P.A. (1990). An overview of chronometric measures of intelligence. *School Psychology Review, 19,* 399-410.
Walsh, W.B. (1989). *Tests and Measurement.* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
Walsh, W.B. & Betz, N.E. (1990). *Tests and assessment.* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
Wheeler, H.L. (1993). The fairness of an engineering selection battery in the mining industry. Unpublished M.Comm dissertation. Pretoria: University of South Africa.
Wheeler, L.J. & Ree M.J. (1997). The role of general and specific psychomotor tracking in validity. *International Journal of Selection and Assessment, 5* (2), 128-136.
Winer, B.J. (1971). *Statistical principles in experimental design.* (2nd ed.). Tokyo: McGraw-Hill.
Woodruffe, C. (1990). *Assessment Centres.* London: Institute of Personnel Management.