CHAPTER 2

VALIDATION OF SELECTION INSTRUMENTS

This chapter deals with employee selection. The main focus will be on the use of psychometric testing in this process, and more specifically, on the validation of the selection instruments used. The use of instruments for the assessment of learning potential and abilities, as well as situational judgment tests and assessment centres in the selection decision-making process are discussed. Then follows a discussion of the development of a criterion measure and the chapter concludes with a summary of the main discussions.

2.1 SELECTION

The performance of employees is a major determinant of how successful an organisation is in reaching its strategic goals. According to Gatewood, Feild and Barrick (2008) this performance is made up of two factors, namely, the ability of the employee and the effort that the employee puts forth. These two factors can be influenced by the organisation. Ability is a function of two organisational practices, namely, selection and training. Effort is a function of the organisation's practices for motivating employees, such as compensation, job design and communication. Motivation practices assume that employees have the ability to perform the job and are intended to mobilise employees to use these abilities in a concerted and continuous manner (Gatewood et al., 2008). The selection process is therefore an organisation's first opportunity to ensure that persons with the required skills enter the organisation or different positions as it is based on the assumption that the procedure can predict one or other important and relevant behavioural requirement or job performance aspect (Edenborough, 2005; PAI, 2005).

Thus, being one of only two ways of ensuring that employees have the abilities to perform their work, selection becomes a top priority for organisations to ensure their competitiveness and ultimate survival (Cascio, 2006; Gatewood et al., 2008; Mirvis & Hall, 1994; Ployhart, 2006; Ulrich et al., 1999; Van der Merwe, 2002).

Most of the discussions in this chapter will centre on Selection. This will be followed by a look at the development of a criterion measure.

2.1.1 Selection definition and process

Roe (2005, p. 74) defines a selection system as 'a configuration of instruments, procedures and people created with the purpose of selecting candidates for certain positions, in such a way that they can be expected to optimally fulfil pre-defined expectations'. This process leads to the choice of one or more candidates over others for one or more jobs or roles. The purpose of any procedure used in selection is to find out something about the candidate or candidates that will be of relevance to their functioning in the job or role concerned. The task in selection is therefore to match knowledge, skills, abilities and other personal characteristics to requirements of the work (Borman, Hanson & Hedge, 1997; Edenborough, 2005).

Gatewood et al. (2008, p. 3) describes selection as 'the process of collecting and evaluating information about an individual in order to extend an offer of employment'. Such employment could be a first position for a new employee or a different position for a current employee.

Finding the right candidate through objective selection, holds widespread gains and addresses the future interests of both the organisation and the individual (Gatewood et al., 2008; SHL, 1995; Smith & Smith, 2005).   Improving the person-job match through objective selection often leads to higher employee

18

commitment and satisfaction and can reduce staff turnover (Edenborough, 2005; SHL, 1995). According to Smith and Smith (2005), good selection can increase productivity by ten percent of labour costs. Waste of associated costs with regards to advertising, time, administration, interviewing, salaries and training is avoided (Edenborough, 2005; SHL, 1995) as fewer unsuitable people are appointed and go through the nightmare of failure and dismissal (Smith & Smith, 2005). The net benefit of structured and objective selection in terms of work output may equate to between ten and twenty percent above the level obtained by random selection. Apart from the benefit to the organisation, objective selection is also fair to the employee in terms of person-job matching and personal development (Edenborough, 2005; SHL, 1995). The benefits spread to colleagues, customers and even society at large. Colleagues gain because they are spared the burden of coping with the mistakes of an incompetent co-worker, while increased productivity produces a sounder economy and tax base (Hunter & Hunter, 1987; Smith & Smith, 2005).

A typical selection process is depicted in Figure 1. This process details the steps commonly followed in arriving at a selection decision.
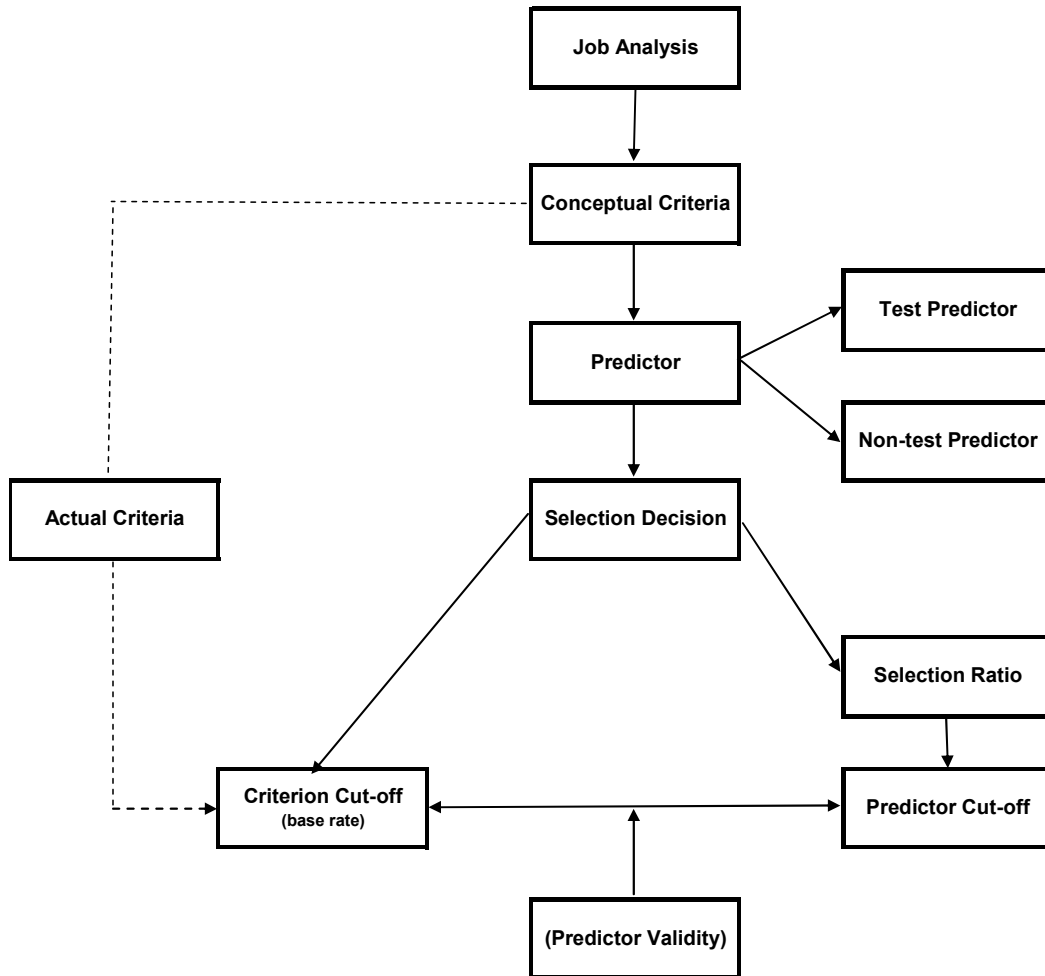
Figure 1. Typical selection process
(Unisa: Department of Industrial and Organisational Psychology,
2008, p. 12)

The starting point in this process is to identify the knowledge, skills and abilities (KSAs) required to effectively do the job to be filled. This can only be achieved by conducting a comprehensive job analysis (Borman et al, 1997; Gatewood et al, 2008; SHL, 1995; Smith & Smith, 2005).

Through analysing the job, a competency framework is determined against which to assess the individual (SHL, 1995). By means of test- and non-test predictors,

information is collected from individuals to determine how much of the necessary knowledge, skills and abilities required to perform the job, they possess. A predictor can be defined as an aid to decision-making applied in the context of selection or other personnel decisions and include for example, standardised ability tests, personality inventories, biographical data forms, situational tests, assessment centre evaluations and ratings based on interviews (PAI, 2005). Gatewood et al. (2008) refer to two principles for choosing the predictors to be used. These are the measurement of the knowledge, skills and abilities identified as needed for the job, and the differentiation among individuals.

The collection of information by means of the predictors enables a comparison of individuals against the criteria determined through the job analysis, as well as against each other, on which a selection decision can be made. The percentage of employees who will be identified as successful following the use of a predictor depends on three factors, namely predictor validity, selection ratio and base rate (Aiken, 2003; Gatewood et al., 2008; Unisa, 2008). Prediction is different from a decision, as it involves estimating a criterion (future job performance), and a decision involves choosing between accepting or rejecting the individual for the position (Born & Scholarios, 2005).

Prediction is achieved by processing information from instruments according to a set procedure (Roe, 2005). Predictor validity refers to the extent to which the predictor (test or non-test) or battery of predictors is related to the criterion of interest (Levy, 2006). Predictor validity will be discussed in more detail in Section 2.3.

The accuracy of the prediction is a crucial factor for the quality of the decision, but the selection ratio is another influence that may have a limiting effect on the soundness of decisions. The selection ratio is the number of persons hired to the number of applicants available, that is, the number of credible applicants or the

proportion of applicants to be selected (Aiken, 2003; Born & Scholarios, 2005; Gatewood et al., 2008; Smith & Smith, 2005).

Another influencing factor is base-rate or criterion cut-off, which refers to the percentage of employees that are successful on the job without the use of the test (Gatewood et al., 2008). This represents the proportion of applicants who would be expected to perform satisfactorily on the job even if no selection instrument or procedure were employed (Aiken, 2003; Born & Scholarios, 2005). If the cut-off score of a test is set very low, there will be many false positives (incorrect acceptances), while there will be many incorrect rejections (false negatives) if the cut-off score is set too high. Furthermore, the lower the selection ratio, the higher the cut-off score, and vice versa.

The percentage of successful applicants varies inversely with the selection ratio, but it varies directly with the validity of the test or other selection device (Aiken, 2003).

The essential principle in the evaluation of any selection procedure is that evidence must be accumulated to support an inference or assumption of job relatedness, as required by the Employment Equity Act (No. 55 of 1998, Section 8) (PAI, 2005). Four important aspects of this process will be discussed in more detail in the following paragraphs. These are job analysis, predictors, criterion measures and validation.

2.1.2 Job analysis

One of the major purposes of job analysis is the development of a selection procedure (Landis, Fogli & Goldberg, 1998; PAI, 2005; Schneider & Konz, 1989).

The purpose of a selection procedure is the identification of the best individuals for a job. The starting point of the process should therefore be the gathering of information about the specific job. To this end, job analysis has been a useful tool. Not only does it provide a base for the other steps in the selection process, but it can also form a part of the justification, and perhaps legal defence, of the process (Landis et al.; 1998; Schneider & Konz, 1898; Smith & Smith, 2005).

Job analysis refers to defining a job in terms of its component tasks or duties and the behaviours, knowledge and skills required to perform it (Levy, 2006; Voskuijl, 2005). The job analysis process yields two major outcomes, namely job descriptions and job specifications. The job description details the task requirements, procedures, methods and performance standards of the job (Voskuijl, 2005; Levy, 2006). Gatewood et al. (2008) also refer specifically to the results (products or services), equipment and material used, and environment (working conditions, hazards and work schedules) that characterise the job. The job specification details the people requirements such as skills, knowledge, abilities, behaviours and other personal characteristics called for by the job (Levy, 2006; Voskuijl, 2005).

Through this process, a competency framework is determined against which to assess individuals (SHL, 1995). Defining the job in terms of duties and responsibilities or in terms of outputs, performance and accountabilities are helpful in determining the appropriate use of tests in selection (Edenborough, 2005). It is important to build an understanding of the organisation's needs so that sound hypotheses can be formulated about the relationships between predictors and criteria (PAI, 2005).

A number of procedures for analysing the job exist. The PAI (2005) emphasises an understanding of the organisation's requirements or objectives when selecting an appropriate method for conducting a job analysis. Job analysis methods used include interviews with incumbents, direct observation of job incumbents, diaries

kept by incumbents, studying of documentation such as instruction materials and maintenance records, critical incidents, task analysis and the repertory grid (Gatewood et al., 2008; Smith & Smith, 2005; Voskuijl, 2005).

The world of work is changing. Lawler, as cited in Borman et al. (1997), believes that selection in competency-based organisations will require identifying persons who fit the 'learning environment' and are likely to be capable of developing the skills necessary to do the organisation's work. Selecting for adaptability, interactional skills, a willingness to learn and a repertoire of multiple skills are predicted to be of increasing importance in future organisations (Borman et al., 1997). Selecting for motivational and dispositional characteristics such as versatility will allow the job to be shaped by individuals themselves (Born & Scholarios, 2005).

Schneider and Konz (1989) and Landis et al. (1998) discuss strategic job analysis where the emphasis is on the specification of tasks to be performed and the knowledge, skills and abilities required for effective performance for a job as it is predicted to exist in the future. These changes can be the result of, for example, mergers and acquisitions, downsizing, automation, and transformations such as a move from a quantity to a quality strategy or a change in emphasis from production to service. They suggest an additional step to traditional job analysis methods where information is gathered by means of a workshop with subject matter experts about the kinds of issues in the job, the company, and/or the larger environment that may affect the job in the future. The tasks and knowledge, skills and abilities (KSAs) are then revised in the light of the expected future changes by rating the target job in the future on (a) the importance and time spent on each task or task cluster, and (b) the importance of the KSAs, the difficulty to learn the KSAs and when the KSAs will be learned. A comparison of present and future ratings allows an assessment of the extent to which changes in the internal and external environment of an organisation yield significant task and KSA changes for the job of interest.

Job analysis is an important component of the selection process as it provides the basis for all the other steps. To select or develop predictors, the relevant criteria that are important for job success need to be identified, which will assist in developing or modifying predictors to be valid indicators of the criteria (Levy, 2006).

2.2 SELECTION INSTRUMENTS / PREDICTORS

In the selection process various tools are used to aid decision-making. These methods include, for example, the curriculum vitae, interviews, reference checking, assessment centres and various psychometric tests (Van der Merwe, 2002) and the aim is to gather information about individuals to enable a comparison against pre-determined criteria.

Experience has shown that psychometric tests are generally more reliable and valid than the other techniques (Van der Walt as cited in Van der Merwe, 2002). Research has shown that well constructed tests can be good predictors of job performance (SHL, 1995) and they are commonly used as part of the chain of activity, but their positioning in the chain may vary substantially (Edenborough, 2005).

A distinction is made between non-test and test predictors. Interviews and assessment centres, for example, are regarded as non-test predictors, while tests of, for example ability, personality and learning potential are regarded as test predictors.

As mentioned in Section 2.1.1, the accuracy of the prediction is a crucial factor for the quality of the selection decision and it is therefore important that valid predictors of the criterion of interest are included. In the discussion of the various

predictors, the results of meta-analytical studies are included. Meta-analysis is a collection of methods used to aggregate results across studies quantitatively to draw more accurate conclusions about inconsistent findings in a particular area of research (Gaugler, Douglas, Rosenthal, Thornton & Bentson, 1987; Hunter & Hunter, 1984).

2.2.1 Psychometric assessment

Psychometric tests have proved to be valid and fair and studies have shown that these tests are about fifteen times more effective than screening interviews. These tools can make a major contribution given the high costs of staff turnover and the heightened importance of identifying key staff from previously disadvantaged groups (Van der Merwe, 2002).

Psychological assessment refers to the application of systematic processes to gather information about individuals or groups, describing it with the aid of numerical scales or fixed categories and in so doing, improving understanding of or reaching decisions about the people concerned, either currently or in a predictive sense (De Beer 2000c; Edenborough, 2005; Foxcroft & Roodt, 2001; Levy, 2006).

Aiken (2003, p. 187) defines psychometric testing as 'carefully chosen, systematic and standardised procedures for evoking a sample of responses from a candidate, which are evaluated in a quantifiable, fair and consistent way'.

In selection, psychometric tests are used as predictors to forecast performance on the job or some other work-related criterion. Information is given by a psychometric test through providing those taking the test with the opportunity to respond to a series of items or events that relate directly or indirectly to a particular area of behaviour (Edenborough, 2005). Organisations will often

employ a set of predictors or battery of tests as a way of predicting criterion performance (Levy, 2005; Van der Merwe, 2002). This multiple test approach would likely increase the proportion of the criterion variance that would be accounted for.

Psychometric assessments play an important role in significantly improving the selection process for both new entrants and internal promotions, as well as in staff development practices (Van der Merwe, 2002).

Assessment centres, ability assessment, the measurement of learning potential and situational judgment tests will be discussed in the following paragraphs.

2.2.1.1 Assessment Centres

An assessment centre can be described as the standardised evaluation of individuals on a set of predictors by multiple raters or assessors (Gatewood et al., 2008; Levy, 2006). It attempts to predict performance criteria by focusing on multiple exercises that simulate activities and situations found on the job, and the interpretation being according to a competency model developed by a formal job analysis (Edenborough, 2005; Foxcroft & Roodt, 2001).

Typical exercises included in an assessment centre are the in-basket, leaderless group discussion, case analysis, fact-finding and analysis, and presentations (Gatewood, 2008; Jaffee, 1971; Smith & Smith, 2005). The average assessment centre uses approximately five exercises to measure ten competencies, with the six most common being interpersonal skills/social sensitivity, communication, motivation, persuasion/influence, organisation/planning and problem solving (Ployhart, 2006).

The multiple techniques used in the assessment centre increase the overall validity and predictive validity thereof. The weaknesses of any particular method can be compensated for by the strengths of another, while participants can compensate for a lack of skill in certain areas by showing their strengths in another (Smith & Smith, 2005).

According to Muchinsky, Kriek and Schreuder (2005) the assessment centre was assessed against validity (the ability of the predictor to forecast criterion performance accurately), fairness (the ability of the predictor to render unbiased predictions of job success across applicants in various subgroups), applicability (application across the full range of job and applicant types) and cost of implementing the method. Assessment centres ranked moderate on applicability, high on costs, and high on both validity and fairness.

A meta-analytical study conducted by Gaugler et al. (1987) reported a mean validity coefficient of 0,37 for assessment centres. In a study by Cohen, Moses and Byham as cited in Hunter and Hunter (1984) a median correlation of 0,63 was found predicting potential or promotion, but only 0,33 predicting supervisor ratings of performance. This is interpreted as assessment centres acting as policy-capturing devices which are sensitive to the personal mannerisms used by top management in promotion. To the extent that these mannerisms are unrelated to actual performance, the high correlation between assessment centres and promotion represents a shared error in the stereotype of a good manager (Hunter & Hunter, 1984).

The use of assessment centres provides a means of evaluating an individual's capabilities as they relate to crucial supervisory skills. It is an attempt to create situations that resemble those in which a supervisor frequently finds himself, that is, dealing with written material (as in the in-basket exercise) or dealing with other people (as in the leaderless group discussion) (Jaffee, 1971).

2.2.1.2 Ability assessment

There is a common belief that mental functioning or intelligence is important for most jobs and tests of cognitive ability are among the most frequently used predictors in selection contexts. Levy (2006) distinguishes between general and specific cognitive ability tests.

General cognitive ability tests measure the individual's general capacity to learn, whereas tests of specific cognitive ability attempt to predict the likelihood that an individual will do well in a particular job given his specific abilities. Measures of specific cognitive functions include those measuring verbal functions, perception and motor performance (Foxcroft & Roodt, 2001).

Meta-analytical studies as cited in Gatewood et al. (2008) show the validity of measures of specific cognitive abilities to be around 0,40 and 0,50. Schmidt and Hunter (1998) show the validity of mental ability tests obtained through a meta-analysis to be 0,51. They also showed that there is a significant gain in predictive validity when ability tests are used in combination with other selection instruments.

Ployhart (2006) discusses various strategies used to reduce subgroup differences in ability test performance. Such include supplementing cognitive with relevant non-cognitive predictors, or using situational judgment tests or assessment centres or ensuring that the assessments have minimal reading requirements.

Tests of cognitive ability are relevant to most jobs, since the rate at which people process information is determined by intelligence. In turn, the processing of information controls the rate at which people learn the job and their ability to cope with new and novel aspects of the job (Aiken, 2003).

The results of studies discussed by Jaffee (1971) show that there is a moderate relationship between intelligence or cognitive ability and success as a supervisor, but in most cases, it is apparently not strong enough to justify the selection of supervisors solely by this means.

2.2.1.3 Assessment of learning potential

Cognitive ability tests have shown a difference in mean scores for different racial and ethnic groups (Hunter & Hunter, 1987). Claassen (1997) has indicated that these changes in intelligence scores are usually linked to educational opportunity, language proficiency, and general socio-economic level. Standard tests of cognitive ability generally measure the products of prior learning and hence rely heavily on the assumption that testees have had comparable opportunities to acquire the skills and abilities being measured.

Given the multicultural nature of the South African society and the fact that people from disadvantaged backgrounds have not had the opportunity to develop their cognitive potential fully, a need was verbalised for alternative predictors or determinants of performance other than ability to be used in conjunction and to add to the prediction supplied by such tests to increase validity and decrease adverse impact (De Beer, 2005).

De Beer (2005, p. 719) describes learning potential as having to do with overall cognitive capacity, including both present and projected future performance. This learning potential is based on Vygotsky's zone of proximal development and refers to the difference between the individual's actual level of performance (that is, present performance level without help) and the potential level of performance (that is, performance level after some form of help or learning has been provided) (De Beer, 2005; Van Eeden, De Beer & Coetzee, 2001).

This type of testing is referred to as dynamic testing, built on the idea that the individual's ability to profit from the guided feedback conveys a sense of the difference between his latent capacity and his observed capacity (Sternberg, 1997). Sternberg and Grigorenko (2001) define it as testing plus an instructional intervention. They describe two common formats of dynamic testing, which are the inclusion of an instruction between a pre-test and a post-test, and an instruction in response to the individuals solution to each test item. Unlike static testing, dynamic testing emphasises the quantification of the psychological processes involved in learning and change.

A modern method of identifying learning potential is through computerised adaptive assessment. In this testing the difficulty level of items is calculated beforehand and the items are then interactively selected during the assessment to match the estimated ability level of the test taker (De Beer, 2000c; De Beer, 2005; Foxcroft & Roodt, 2001).

Taylor as cited in Van Eeden et al. (2001) emphasises the importance of identifying those individuals who have potential for development, even though their abilities are currently limited by past disadvantages. Such individuals should be given the opportunity to develop specific skills through training programmes.

In response to the need to construct an instrument for the measurement of learning potential in the domain of general non-verbal figural reasoning ability, De Beer (2000c) developed the Learning Potential Computerised Adaptive Test (LPCAT) in South Africa. This test allows the assessment of fluid cognitive ability in a more equitable and culture-fair manner.

2.2.1.4 Situational Judgment tests

Situational judgment tests are designed to measure judgment in work settings. In those tests classified as situational, a scenario is described and the respondent must identify an appropriate response from a list of alternatives. Other measures do not present a situation, but rather require respondents to indicate their level of agreement with statements concerning the appropriateness of various work-related behaviours (Gatewood et al., 2008; McDaniel, Morgeson, Finnegan & Campion, 2001). These tests are therefore samples of likely job performance, rather than signs of possible job performance as the test items are directly developed or sampled from the criterion behaviours that the test is designed to predict (Chan & Schmitt, 2002).

Because job-relevant situations could occur in both task (technical) and contextual (motivational or interpersonal) domains, the ability to make effective judgments or responses to situations is likely to affect both task and contextual performance dimensions, in addition to overall job performance (Chan & Schmitt, 2002).

A meta-analytical study by McDaniel et al. (2001) reported the estimated population validity of situational judgment tests to be 0,34. Their study revealed the mean validity from studies with predictive designs (0,18) to be lower than the mean validity from studies with concurrent designs (0,35).

Clevenger, Pereira, Wiechmann, Schmidt and Harvey (Gatewood et al., 2008) have shown that situational judgment tests provide incremental validity even when a selection battery that includes measures of mental ability, consciousness, job experience and job knowledge is used. These tests also have minimal capacity to moderate gender and ethnic differences (Ployhart, 2006).

2.3   VALIDATION OF ASSESSMENT INSTRUMENTS

The importance of selection validity is emphasised by the Employment Equity Act (No. 55 of 1998, Section 8) that governs the use of assessments in South Africa. It states that 'psychometric testing and other similar assessments of an employee are prohibited unless the test or assessment being used has been scientifically shown to be *valid* and reliable, can be applied fairly to all employees; and is not biased against any employee or group'.

Apart from legislative requirements, high selection validity also translates into considerable financial savings for organisations (Hunter & Hunter, 1984). It is therefore imperative that organisations embark on validation studies that can offer them an insight into whether or not people who perform well during the assessment process tend to be successful on the job.

Validity can be defined as the extent to which a test measures what it was designed to measure (Aiken, 2003; Foxcroft & Roodt, 2001). Van der Merwe (2002) states the importance of tests being validated for different organisations in which they are used. This also implies that in each organisation, for each specific job, the validity of the test battery being used must be ensured.

The validity of a measure is directly proportional to its reliability, that is, the consistency with which it measures whatever it measures (Anastasi & Urbina, 1997; Foxcroft & Roodt, 2001).

The methods by which the validity of a test may be determined include
(1) analysing the content of the test (content validity),
(2) computing the correlation between scores on the test and those on the criterion of interest (criterion-related validity), and
(3) investigating the particular psychological characteristics or constructs measured by the test (construct validity) (Aiken, 2003).

2.3.1 Types of validation studies

There are three types of validity or validation procedures, namely criterion-related validity, content validity and construct validity (Foxcroft & Roodt, 2001).

2.3.1.1 Criterion-related validity

Personnel selection procedures are used to predict future performance or other work behaviour. Assessment results should thus be interpreted in terms of expected job performance (the criterion) and not in terms of the construct being assessed (PAI, 2005).

Procedures in which test scores are compared with ratings, classifications or other measures of performance are called criterion-related validation (Aiken, 2003). This quantitative procedure involves the calculation of a correlation coefficient between a predictor or predictors and a criterion (Foxcroft & Roodt, 2001; Levy, 2006). Evidence for criterion-related validity is demonstrated when a test is shown to be effective in estimating an examinee's performance on some outcome measure or criterion, that is, when a relationship exists between the results of a selection procedure (predictor) and one or more measures of work-related behaviour or work outcomes (Gregory, 2000; PAI, 2005).

A distinction is made between two approaches to criterion-related validity, namely concurrent validity and predictive validity. This classification is based on the presence or absence of a time lapse between the collection of the predictor and criterion data (PAI, 2005). There could also be a difference in the employment status of the sample used in the two designs (Barrett et al., 1981; PAI, 2005). In situations involving concurrent validation, incumbents, rather than applicants, are used as participants, and predictors and criteria are measured at the same time (Levy, 2006).

Concurrent validity involves the accuracy with which a measure can identify or diagnose the current behaviour or status regarding specific skills or characteristics of an individual (Foxcroft & Roodt, 2001; Gregory, 2000). Thus, whenever a criterion measure is available at the time of testing, the concurrent validity of the test can be determined (Aiken, 2003; Levy, 2006; PAI, 2005). As per Figure 2 that follows, the criterion measure (for example work performance) is obtained at approximately the same time as the predictor data (for example, psychometric test results) (Gregory, 2000). The process determines the accuracy with which these instruments can identify the current behaviour (work performance) of an individual (Anastasi & Urbina, 1997; Foxcroft & Roodt, 2001).

```
┌─────────────────────────────────┐
│          Job analysis           │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Determining relevant KSAs and  │
│  other characteristics to perform│
│      the job successfully.      │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Choose or develop experimental │
│     predictors of these KSAs.   │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    Select criteria of job success.│
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│ Administer predictors to current │
│  employees and collect criterion │
│              data.              │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Analyse predictor and criterion │
│        data relationships.      │
└─────────────────────────────────┘
```
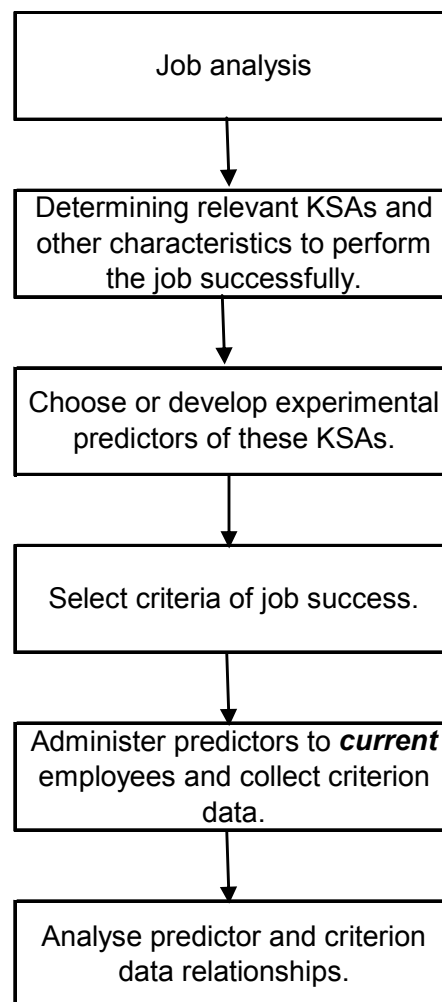
Figure 2. Major steps undertaken in conducting concurrent validity studies (Gatewood et al., 2008, p. 170)

Predictive validity refers to the accuracy with which a measure can predict the future behaviour or status of an individual (Aiken, 2003; Anastasi & Urbina, 1997; Foxcroft & Roodt, 2001; Gregory, 2000). Thus, when scores on the criterion do not become available until some time after the test has been administered, the focus is on predictive validity (Aiken, 2003; Levy, 2006; PAI, 2005). In such studies, as depicted in Figure 3, the scores on the predictor are obtained first where after there is an interval before the criterion scores are obtained (Smith & Smith, 2005) and the accuracy with which the instruments can predict future behaviour (for example, work performance) is determined (Anastasi & Urbina, 1997; Foxcroft & Roodt, 2001).
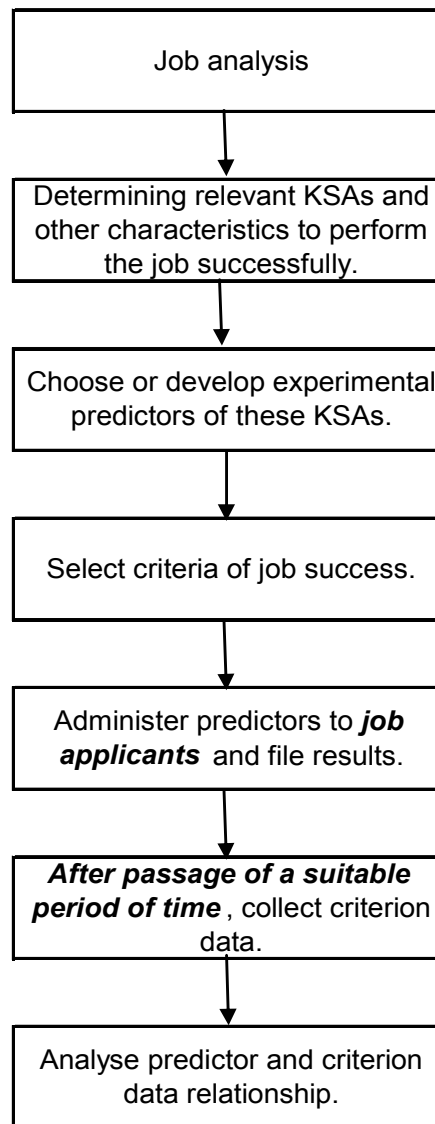
```
┌─────────────────────────────┐
│        Job analysis         │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│ Determining relevant KSAs and│
│ other characteristics to perform│
│     the job successfully.   │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│ Choose or develop experimental│
│    predictors of these KSAs. │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│  Select criteria of job success.│
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│  Administer predictors to *job*│
│  *applicants* and file results.│
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│ *After passage of a suitable*│
│ *period of time*, collect criterion│
│           data.             │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│  Analyse predictor and criterion│
│      data relationship.     │
└─────────────────────────────┘
```

Figure 3. Major steps undertaken in conductive predictive validity studies (Gatewood et al., 2008, p. 170)

Predictive validity has often been regarded as a scientifically superior strategy for the validation of selection tests (Aiken, 2003; Barret et al., 1982). Criticism of concurrent validity studies include missing persons, restriction of range, motivational and demographic differences between present employees and job

applicants, and confounding by job experience (Barrett et al., 1981; Levy, 2006). Because of these factors, correlation results need to be interpreted with caution.

To establish predictive validity requires scope for effective, controlled research over a period of time, large numbers of the research sample and some stability of roles and performance measures (Edenborough, 2005; Smith & Smith, 2005). In practice, such conditions are often difficult to obtain and makes the validation effort costly. Therefore, organisations often undertake validation in the here and now by means of concurrent validation.

2.3.1.2 Content validity

Content validity is concerned with whether the content of a measure elicits a range of responses that are representative of the entire behaviour domain that the test is supposed to sample (Aiken, 2003; Foxcroft & Roodt, 2001; Gregory, 2000). It is a non-statistical type of validity where test items are evaluated by a panel of experts during the test construction phase (Foxcroft & Roodt, 2001; Gregory, 2000).

Content validity is a useful concept when a great deal is known about the variable the researcher wishes to measure (Gregory, 2000). It is most often used with achievement tests for which there is usually no specified external criterion (Foxcroft & Roodt, 2001).

2.3.1.3 Construct validity

The construct validity of a test or instrument refers to the extent to which it measures the theoretical construct or trait it was intended to measure (Aiken, 2003; Foxcroft & Roodt, 2001; Levy, 2006; Smith & Smith, 2005). A test

designed to measure a construct must estimate the existence of an inferred, underlying characteristic based on a sample of behaviour (Gregory, 2000).

In this research a criterion-related validity study was conducted. A concurrent validity approach was followed. In terms of this approach, test scores were correlated with a measure of job performance for a sample of current job incumbents.

For a criterion-related validity study it should be possible to develop a relevant, reliable and uncontaminated criterion measure.

2.4 DEVELOPMENT OF A CRITERION MEASURE

A criterion is any outcome measure against which a test is validated (Gregory, 2000). Because important organisational decisions will be made directly on the basis of criteria, organisations need to ensure that these measures are reliable, appropriate and free of contamination from the test itself (Gregory, 2000; Levy, 2006). To the extent that problems in a criterion measure exist, the effectiveness or validity of the predictor is attenuated. It may lead to the erroneous conclusion that there is no significant relationship between the predictor and the criterion. Jenkins already argued in 1947 (Pursell, Dossett & Latham, 1980) that validation is not a simple technical problem and the criterion may provide researchers with as great a technical challenge as the procedures incident to the assembly of good predictors.

Literature lists a few characteristics of a good criterion (Gregory, 2000; Levy, 2006; Viswesvaran & Ones, 2005). These are –

- Relevance – the extent to which the actual criterion measure is related to the major elements of the job, or the ultimate criterion as discussed below.

It is thus the overlap between what is measured and what was intended to be measured;

- Reliability – the scores obtained should be consistent, meaning that the measure should not give drastically different results when used for the same employees at close time intervals. The validity coefficient is constrained by the reliability of both the test and the criterion;

- Sensibility – the extent to which the actual criterion measure can discriminate among effective and ineffective employees;

- Practicality – the ease with which a measure can be administered;

- Fairness – the extent to which the criterion measure is perceived by employees to be just and reasonable;

- Free of contamination – the exclusion of sources of variance in a measure that was not intended in the theoretical conceptualization. Contamination occurs in that outcomes could be due to factors beyond the control of the individuals; and

- Free of deficiency – criterion deficiency refers to the extent to which the outcomes assess may not take into account important aspects of individual job performance.

Performance ratings continue to be the most often used criterion measure for personnel research applications. The most appropriate criterion measure for the validity of intelligence, special aptitude and personality measures is actual job performance (Foxcroft & Roodt, 2001). Levy (2006, p 91) defines performance as 'actual on-the-job behaviours that are relevant to the organisation's goals'. The ultimate criterion encompasses all aspects of performance that define

success on the job. Unfortunately, this makes the ultimate criterion complex, resulting in the use of the actual criterion, the best real-world representative of the ultimate criterion (Levy, 2006).

The various types of criteria have been formed based on the time span of performance considered or on what is included in the measure of performance. Types of criteria include immediate, intermediate and ultimate criteria (Viswesvaran & Ones, 2005). The ultimate criterion represents the total worth of the individual to the organisation over the entire career span, while the immediate criterion is a measure of individual job performance at a particular point in time. The intermediate criterion summarises performance over a period of time. Various authors, as cited in Viswesvaran & Ones (2005), go further to distinguish between maximal and typical performance measures. Maximal performance is what an individual can do if highly motivated, whereas typical performance is what an individual is likely to do in a typical day.

2.4.1 Job performance as a criterion measure

Job performance refers to 'scalable actions, behaviours and outcomes that employees engage in or bring about that are linked with and contribute to organisational goals (Viswesvaran & Ones, 2005, p 354). These are used in selection to validate predictors. According to these authors, the methods of measuring job performance can be classified into two broad categories, namely organisational records or objective criteria and subjective evaluations.

Objective criteria are taken from organisational records, for example, number of days absent per year, number of products produced per day (productivity), number of accidents per year and number of minutes late per month. These criteria are not supposed to involve any subjective judgments or evaluations (Levy, 2006; Viswesvaran & Ones, 2005).

From the examples given, it is clear that objective criteria are best applied to non-managerial jobs, but managerial and other higher-level jobs do not lend themselves to this type of criteria. For these jobs, subjective criteria are more appropriate.

Subjective criteria refer to performance measures based on judgments or evaluations (Levy, 2006). These include ratings of employees by other employees such as supervisors, co-workers and subordinates. The use of performance ratings by supervisors as a criterion measure is discussed in the next section.

Apart from objective and subjective criteria, the use of contextual performance as a criterion domain also emerged. According to Levy (2006) this refers to activities performed by employees that help maintain the broader organisational, social, and psychological environment in which the technical core operates. Borman and Motowidlo, as cited in Levy (2006), suggest that the inclusion of contextual performance as part of the criterion is heightened by changes in the workplace such as global competition and an increased emphasis on teams, as it has a role to play in determining organisational effectiveness.

Performance is best represented by multiple criteria. At least two alternatives, focusing on the creation of a composite criterion, are used by organisations to make a decision based on this multidimensional information. By creating a composite criterion, one index of performance is obtained by combining the multiple criteria. In this instance equal weightings are assigned to the criteria. Another approach is to weight the criteria differently according to the importance to the specific decision. The manner in which such weightings are assigned is a value judgment made by the organisation based on its goals and on research conducted by the organisation (Levy, 2006).

Schmidt and Kaplan as cited in Levy (2006) argue that the use of separate criteria versus the use of a composite criterion is an issue of purpose. According to them multiple criteria should be examined in the light of predictor variables without combining them into a composite score if the primary goal is psychological understanding. If the primary goal is decision-making with an economic focus, they propose that criteria be combined into one measure.

Empirical research suggests that psychometric properties are not affected by issues such as the number of scale points of a measure (Viswesvaran & Ones, 2005). Consistency across and within raters is improved by providing a common frame of reference as to what each scale point refers to.

Typically, in traditional organisations, the supervisors of employees provide the ratings. In recent years, 360-degree feedback systems have become popular. Such systems comprise rating assessments done by the ratee himself or herself, supervisors, subordinates, peers, and customers or clients. When ratings are used for selection purposes, self-ratings are mostly inappropriate (Viswesvaran & Ones, 2005).

Studies cited in Viswesvaran and Ones (2005) found in a review of 1 506 validation studies that 63 percent of the studies used ratings as the criterion measurement method. Of these studies, 93 percent used supervisor ratings. A study by Bernardin and Beatty as cited in Viswesvaran and Ones (2005) estimates that over 90 percent of the ratings used in the literature are supervisory evaluations.

Conway and Huffcutt (1997) investigated the inter-rater reliabilities within sources and correlations between sources of subordinate, supervisor, peer, and self-ratings of job performance. It was found that supervisors showed the highest mean reliability (0,50), when compared to the 0,30 of subordinates and 0,37 of peers.

Mean correlations between sources were low for subordinate ratings (0,22 with supervisor, 0,22 with peer and 0,14 with self-ratings) and for self-ratings (0,22 with supervisors and 0,19 with peer ratings), while the mean supervisor-peer correlation was higher at 0,34. Both reliabilities and correlations between sources tended to be higher for non-managerial and lower complexity jobs.

A study by Harris and Schaubroeck (Viswesvaran and Ones; 2005) reported that the correlation between peer and supervisor ratings or overall job performance was 0,62. Subsequent studies by Viswesvaran and Ones suggested that the inter-rater reliability of supervisor ratings is 0,52 and that of peer ratings is 0,42. These values suggest that the convergence is much lower than the 0,62 reported by Harris and Schaubroek.

The collection of ratings from different sources for the assessment of job performance in personnel selection will result in a more comprehensive sampling of the domain of performance and subsequently, a more valid and reliable assessment. User acceptability may also be enhanced by using the multiple sources in validation (Viswesvaran & Ones, 2005).

2.4.2 Supervisory ratings

As mentioned, supervisor ratings are most commonly used as the criterion.

Various factors can have an influence on overall supervisory ratings. Hunter, as cited in Borman, White and Dorsey (1995) evaluated causal relationships between cognitive ability, job knowledge (paper-and-pencil test scores), task proficiency (work-sample scores), and overall supervisory performance ratings. Results showed that ratee job knowledge had the largest direct path to the ratings, approximately three times as large as the ratee task proficiency. Cognitive ability had an indirect influence on the ratings, primarily through its

effect on knowledge acquisition. A subsequent study by Schmidt, Hunter and Outerbridge (Borman et al., 1995) showed similar results, with experience having a moderate effect on ratings, primarily through its influence on job knowledge.

A study by Borman, Whith, Pulakos and Oppler (Borman et al., 1995) concluded that technical proficiency and rate of disciplinary problem behaviour had the strongest direct effects on the supervisory ratings.

Thus, supervisory ratings appear to be substantially based on ratee technical performance (job knowledge and technical proficiency). 13 percent to 17 percent of the variance in supervisory ratings can be accounted for by ratee ability, job knowledge and technical proficiency (Borman et al., 1995).

Criterion-related validity coefficients can be significantly increased by providing raters with a training programme. Such a training programme, which teaches people behaviour observation skills in recording accurately the frequency with which they have seen the employee perform the requirements of the job, has shown to decrease rating error (Pursell et al., 1980). The programme developed by Latham, Wexly and Pursell (Pursell et al., 1980) incorporates three basic principles of learning that is necessary for bringing about a relatively permanent change in behaviour, namely, active participation, knowledge of results, and practice. The training stresses the necessity of basing a numerical rating on the frequency with which the rater has observed the employee engage in critical job behaviours. The programme also requires participants to brainstorm as a group specific ways of ensuring that a rating is not contaminated by factors irrelevant to job performance.

2.5 INTEGRATION

Staffing is a key strategic opportunity for enhancing companies' competitive advantage. Organisations that can better attract, retain and select talent should

be able to ensure their competitiveness and ultimate survival. The selection process, as only one of two ways of ensuring that employees have the abilities to perform their work, holds positive effects for both the organisation and the individual.

The purpose of a selection programme is the identification of the best individuals for a job. Through job analysis a competency framework is determined against which to assess these individuals. Defining the job in terms of duties, responsibilities and outputs the use of appropriate tests in selection is also determined. Job analysis is therefore the starting point and it provides a base for the other steps in the selection process.

In selection, psychometric tests are used as predictors to forecast performance on the job or some other work-related criterion. These tools play an important role in significantly improving the selection process for both new entrants and internal promotions, as well as in staff development practices. Assessment centres, ability assessment, the measurement of learning potential and situational judgment tests are tools presented to aid in this process and have been shown to be related to job performance.

It is imperative that organisations embark on validation studies in order to offer them an insight into whether or not people who perform well during the assessment process tend to be successful on the job; that is, that practices and tools measure what they intend to measure. As personnel selection procedures are used to predict future performance or other work behaviour, criterion-related validation is suggested as more appropriate.

In the research, the core competencies for the first-line supervisor position were determined by means of a job analysis. A concurrent validation study was then conducted utilising ability and learning potential assessment tools, situational judgment tests and assessment centres as the predictors and supervisory ratings

of job performance as the criterion. Validity and reliability of the instruments were checked and correlations were conducted to determine if a relationship existed. Details of the empirical study and the validity of the instruments are presented in Chapter 3.

REMARK

In concluding this chapter, the following theoretical aims as captured in Section 1.3.2 have been fulfilled:

- To conceptualise selection, job analysis, psychometric assessment and validity. (Section 2.2, Section 2.1.1, Section 2.1.2, Section 2.2, Section 2.2.1, Section 2.3, Section 2.3.1.1, Section 2.3.1.2, and Section 2.3.1.3).
- To conceptualise the use of assessment centres, ability tests, tests of learning potential and situational judgment tests in predicting work performance. (Section 2.2.1.1, Section 2.2.1.2, Section 2.2.1.3 and Section 2.2.1.4).
- To conceptualise the development of a criterion measure. (Section 2.4).
- To integrate the aspects of selection, work performance and validation. (Section 2.5).

2.6 CHAPTER SUMMARY

In this chapter the selection process was discussed, with specific reference to job analysis and psychometric testing. Test validation and the different types of validity were presented. The development of a criterion measure was discussed and the chapter was concluded with an integration and consolidation of the discussions. In Chapter 3 that follows, the empirical study is discussed.