

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town



School of Management Studies

---

---

# APIL-SV as a Predictor of Job Performance in a South African Financial Consulting Firm

---

---

## MASTERS THESIS

**SUPERVISOR:** Burger van Lill  
**STUDENT NAME:** Emma Jane Strachan  
**STUDENT NUMBER:** STREMM003

### PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is my own.
2. I have used the APA referencing guide for citation and referencing. Each contribution to, and quotation in this assignment from the work(s) of other people has been contributed, and has been cited and referenced.
3. This assignment is my own work.
4. I have not allowed, and will not allow, anyone to copy my work.

Signature:

Signed by candidate

Date: 25 April 2008

PREDICTIVE VALIDITY OF APIL-SV

APIL-SV as a predictor of job performance in a South African Financial  
Consulting Firm

Emma J Strachan

University of Cape Town

“To dream anything that you want to dream. That is the beauty of the human mind. To do anything that you want to do. That is the strength of the human will. To trust yourself to test your limits. That is the courage to succeed”

*Bernard Edmonds*

## Acknowledgements

The author wishes to thank the following people who allowed for the dream, and who have contributed to ensuring that the dream was achieved:

- Mr and Mrs Strachan, two loving parents who have been there to support me through the good times and the bad, to ensure that I could be all that I wanted to be, and who never stopped believing in me.
- Mrs Ferraz, for being there on long late nights, and providing me with the support through the tough times.
- Wiehann Rademan, who provided me with the data for my research topic
- And last but not least, to all of my friends and family who have been there this year, helping me in ways that they knew they could help, from statistics through to providing me with a sense of life at times when all seemed lost.

University of Cape Town

“APIL-SV AS A PREDICTOR OF JOB PERFORMANCE IN A SOUTH AFRICAN FINANCIAL/CONSULTING  
FIRM”

ABSTRACT

This paper aims to investigate the predictive validity of the Ability, Processing of Information and Learning Potential Short Version (APIL-SV) in predicting performance in a South African financial/consulting firm. The study uses the company's internal performance measure and a board written exam as the criterion measures. The paper also investigates whether there is any form of cultural bias in the APIL-SV and how many constructs the APIL-SV taps. The external performance measure sample consisted of 69 participants, and the general sample 2 877 participants with ages ranging from 16–56 respectively. The results of the statistical tests support the theoretical construct that the APIL-SV measures one construct and is a good predictor of performance if the criterion measure measures performance and not a different construct.

## Table of Contents

<b>ACKNOWLEDGEMENTS.....</b>	<b>III</b>
<b>ABSTRACT.....</b>	<b>IV</b>
<b>LIST OF TABLES .....</b>	<b>VI</b>
<b>LITERATURE REVIEW.....</b>	<b>I</b>
INTRODUCTION.....	1
HISTORY.....	2
PREDICTION OF JOB PERFORMANCE.....	3
COGNITIVE ABILITY AND JOB PERFORMANCE.....	4
COGNITIVE ABILITY AND LEARNING POTENTIAL.....	5
SOUTH AFRICA.....	5
VALIDITY.....	6
RELIABILITY.....	9
FAIRNESS.....	10
BIAS.....	10
PAST STUDIES ON THE APIL-B.....	11
RESEARCH QUESTIONS.....	12
<b>METHOD.....</b>	<b>13</b>
SAMPLE.....	13
<i>First group – factor analysis sample.....</i>	<i>13</i>
<i>Second group – performance on the Public Practice Examination sample.....</i>	<i>14</i>
DATA ANALYSIS.....	15
MEASURING INSTRUMENTS.....	15
<i>The predictor measure.....</i>	<i>15</i>
<i>The criterion measure.....</i>	<i>18</i>
PROCEDURE.....	19
<b>RESULTS.....</b>	<b>21</b>
PUBLIC PRACTICE EXAMINATION.....	21
FACTOR ANALYSIS SAMPLE.....	31
<b>DISCUSSION.....</b>	<b>40</b>
<b>REFERENCE.....</b>	<b>44</b>
<b>APPENDIX.....</b>	<b>48</b>

## List of Tables

TABLE 1.1 .....	13
TABLE 2.1 .....	21
TABLE 2.2 .....	22
TABLE 2.3 .....	23
TABLE 2.4 .....	23
TABLE 2.5 .....	24
TABLE 2.6 .....	24
TABLE 2.7 .....	24
TABLE 2.8 .....	25
TABLE 2.9 .....	26
TABLE 2.10 .....	26
TABLE 2.11 .....	27
TABLE 2.12 .....	28
TABLE 2.13 .....	29
TABLE 2.14 .....	30
TABLE 3.1 .....	31
TABLE 3.2 .....	31
TABLE 3.3 .....	32
TABLE 3.4 .....	32
TABLE 3.5 .....	33
TABLE 3.6 .....	33
TABLE 3.7 .....	34
TABLE 3.8 .....	35
TABLE 3.9 .....	36
TABLE 3.10 .....	38

## Literature Review

### *Introduction*

Assessments to predict a person's cognitive ability have been used globally for centuries, for example astrology, humorology and physiognomy (Cohen & Swerdlik, 2005; Foxcroft, Roodt & Abrahams, 2003). Astrology has been traced back to as early as the fifth century BC when it was used to describe personality traits and to predict the future (McReynolds, cited in Foxcroft *et al.*, 2003). McReynolds further discusses the use of humorology, also dating back to the fifth century BC, which uses human fluid to define a person's type of temperament. In the sixth century BC, physiognomy attempted to judge a person's characteristics by their external features. At the turn of the twentieth century assessments became grounded in the theories of psychology and personality, which led to an interest in creating cognitive ability assessments. Cognitive assessments are used to assist in making decisions about a predefined set of psychological procedures that are the traits required by the organisation. Cognitive assessments are used by trained psychologists to make inferences about whether or not a person is a fit with the job and the organisation (Hansen cited in Highhouse, 2002).

To be a successful predictor of job performance, the instrument used should measure one of two areas: it should measure some critical aspect of the task expected of the applicant should they be a successful employee, or it should measure a specific characteristic that is known to be the root cause of successful task performance (Rothstein & Goffin, 2006). Rothstein and Goffin (2006) emphasise the need for a clear demonstration of the predictor working as intended: satisfying certain psychometric requirements of congruency.

Rudner (1992) identifies three factors required for effective pre-employment testing. Firstly, there should be a correlation between test scores and job productivity; secondly, the percentage of applicants being hired in relation to the number applying for the position; and thirdly, the proportion of applicants that is classed as successful by a test. It is believed that, under the right conditions, pre-employment testing can result in improved corporate productivity. The use of assessments in selecting the correct employee can increase productivity by millions of dollars.

## *History*

Cognitive assessments first evolved during the First World War (Cohen & Swerdlik, 2005; Foxcroft *et al.*, 2003; Morris & Levinson, 1995), and were precursor to the development of personality tests. Tests were predominantly devised to determine military career suitability as either an officer or a soldier, based on a person's cognitive ability. The popularity of psychological assessments throughout history has been influenced by external factors such as politics, publicity and a country's leaders (Hogan, De-Fruyt & Rolland, 2006). In turn, psychological assessments have had periods of waxing and waning popularity.

In 1973, McClelland (cited in Barrett & Depinet, 1991) maintained that competency-based assessments should replace cognitive ability assessments. Cognitive ability assessments show strong correlations with school grades and academic performance as a social status result, but do not correlate strongly with other life outcomes. Although early studies of this phenomenon found positive correlations between cognitive ability and job performance, these studies have been found to suffer from methodological problems (Morris & Levinson, 1995). Although Super and Crites (1962, cited in Morris & Levinson, 1995) conclude that cognitive ability illustrates above and beyond what is required to learn a job, they show that it has no positive effect on the worker's success on that job. In contrast to this view, Schmidt and Hunter (1981, cited in Morris & Levinson, 1995) conclude that cognitive ability is indeed a great predictor of almost all occupational decisions. This is further supported in later research by Hunter and Hunter (1984) with a mean correlation range of success from 0.5 to 0.6. Ghiselli (1973) suggests that tests of cognitive ability have moderate validity when used to predict job performance, with the highest correlations being in protective occupations (Hiatt & Hargrave, 1988).

The new millennium has seen an increased interest in cognitive ability with regards to cognitive ability tests in batteries (Ree & Cattetta, 2002). Ree and Cattetta (2002) further elaborate that, as the complexity of the job increased, the validity of cognitive ability increased congruently.

Schultz (1984) found that validity coefficients will be lower when the focus is predominantly on one end of distribution measures, rather than on the whole range of measures (Schultz, 1984). Criterion measures that are focused on a specific topic create flaws that are likely to reduce validity levels, rather than raise them. Schultz (1984) mentions that ideologically separate predictors would lead to greater test

utility; however, in practice the extra expense is not feasible for many organisations. This results in organisations using one test to measure multiple domains (Murphy & Shiarella, 1997).

### *Prediction of job performance*

With history's focus on predicting a person's performance, and with organisations increasingly needing to rely on people to perform, modern psychometric assessments need to focus more on being able to predict potential job performance. Many researchers acknowledge this when evaluating job performance that does not deal with a single entity as models assume (Murphy & Shiarella, 1997). Borman and Motowildo (1997) propose two dimensions to this phenomenon, namely task performance and contextual performance, which is further supported by Murphy and Shiarella (1997).

Task performance evaluates an individual's learning on the task and the effectiveness with which that individual performs the designated activities (Borman & Motowildo, 1997; Murphy & Shiarella, 1997), as well as the individual's effectiveness in motivating themselves to perform the task effectively. This either directly or indirectly contributes to the core activities of the organisation (Borman & Motowildo, 1997; Murphy & Shiarella, 1997).

Contextual performance evaluates the behaviours exhibited by the individual that are not related to the key activities for their position (Borman & Motowildo, 1997; Murphy & Shiarella, 1997). These behaviours create and maintain contexts on social and organisational levels and act as catalysts for tasks and processes that contribute to the organisation's effectiveness in the market (Borman & Motowildo, 1997; Murphy & Shiarella, 1997). Borman and Motowildo (1997) further define this as an overall taxonomy of organisational citizenship behaviour, pro-social organisational behaviour and organisational courtesy, which is more likely to involve personality variables.

Studies have illustrated that this notion of multidimensional constructs has an Achilles heel (Borman & Motowildo, 1997). Many tools used to evaluate the performance of an individual are substantially influenced by the individual's contextual performance. The contextual performance rating will be inclined to take

into account the person's organisational citizenship behaviour and social behaviour rather than their actual task performance.

### *Cognitive ability and job performance*

Research shows that cognitive ability is considered to be the best predictor of job performance and training performance (Goldstein, Zedeck & Goldstein, 2002; Gottfredson, 2002; Murphy, 2002; Schmidt, 2002). Cognitive ability is perhaps the most studied human trait in psychology, which is why it is seen as a strong predictor across a wide range of settings and jobs, particularly if one emphasises criteria such as efficiency, performance and productivity in making selection choices between employees (Gottfredson, 2002; Murphy, 2002; Outtz, 2002; Schmidt & Hunter, 1998; Stenberg & Hedlund, 2002). Chan (1996 cited in Ree & Cattetta) has demonstrated that cognitive ability can predict future performance. Suggestions have been made that the cognitive ability of an individual becomes an important factor in influencing job performance as the nature of work becomes increasingly complex and unpredictable (Gottfredson, 1997).

Cognitive ability can be applied to jobs on all levels within an organisation as many job tasks contain an aspect of general problem solving (Gottfredson, 2002). Cognitive ability assesses the candidates' ability to deal with complexity and their ability to process information. The validity of cognitive ability derives from the intention to assess the operation of critical thinking skills and the candidates' fluid and crystallised intelligence. Coward and Sackett (1990) found that ability-performance relations exhibit a linear pattern, in that if a candidate performs low on a cognitive ability test then their job performance will be low, and vice versa. In a study conducted by Hunter and Hirsh (1987), they found that a correlation between cognitive ability and performance appears to be similar across jobs with degrees of considerable difference in content.

Schmidt and Hunter (1998) found that despite the heavy reliance on cognitive assessments by organisations, they only account for 20–25% of the variance in performance, leaving 75–80% of the variance in job performance unexplained. Schmidt and Hunter (1998) argue that cognitive ability has the strongest theoretical foundation as a predictor than many other areas of psychology. As jobs require an element of problem-solving ability, there needs to be some level of cognitive ability

that can be assessed accurately. Gottfredson (2002) found that the more complex and loaded the cognitive ability test was, the higher the test correlated with job performance, improving upon its predictive validity. Meta-analysis results of CATs have illustrated that operational validity ranges from .38 to .47 for overall job performance with CATs (Bertua, Anderson & Salgado, 2005).

Bertua *et al.* (2005) found that, in the UK, cognitive ability tests exhibit good predictive validity for overall job performance. In a study, cognitive ability assessments showed a .48 operational validity with a standard deviation of .24. The study illustrated that cognitive ability assessments are robust predictors of job performance across different occupations (Bertua *et al.*, 2005). A key finding to this study is that, contrary to previous research which failed to find validity generalisation for some tests, tests used in this study showed positive credibility values, indicating that validity can be generalised when predicting job performance.

#### *Cognitive ability and learning potential*

Schmidt (2002) maintains that the past two decades of research have proved that cognitive ability is essentially linked to a person's ability to learn. Gottfredson (2002) found that more complex items require an increase in mental manipulation in order for people to learn something or to solve a problem. In essence, Gottfredson describes an individual's cognitive ability as their ability to learn moderately complex to complex material as quickly and efficiently as possible.

#### *South Africa*

South Africa's psychometric history has mirrored the history of the United States of America (USA) and Europe (Foxcroft *et al.*, 2003) as a result of a significant western influence. In South Africa's history, apartheid played an influential role in the use of psychometric assessments, mirroring the western oppression of minorities. Psychological assessments were predominantly evolved to reflect a racially segregated society (Foxcroft *et al.*, 2003). In recent decades there have been worldwide efforts to reverse the prejudicial acts of the past.

As a result of South Africa's turbulent past and in compliance with the Health Professions Council of South Africa, the Employment Equity Act (1998) was amended in 1999 to include section 8. Section 8 emphasises that all psychological

tests should be scientifically proven to be valid and reliable, be able to be applied fairly to all employees, and to show no bias towards any group.

This has further promoted users' awareness of the legitimacy of the use of assessments for purposes such as recruitment, selection and identifying potential (Lopes, Roodt & Mauer, 2001). The current test, APIL-SV, has been proven by the author Taylor to comply with all the above aspects, and will be discussed in the sections that follow.

### *Validity*

A test is considered valid if it measures the construct that it states it measures (Cohen & Swerdlik, 2005; Foxcroft *et al.*, 2003). This is known as construct validity. Content validity, on the other hand, seeks to ascertain whether the content of the test covers a representative sample of the criterion being measured (Cohen & Swerdlik, 2005). Although not valid in psychometric terms, face validity refers to whether or not the test is actually measuring what it appears to be measuring (Cohen & Swerdlik, 2005). View points on validity have changed over recent years. Previously areas under validity were seen as separate entities, where as in recent times, Validity is seen as a construct where all entities relate to each other for a test to be valid (Huysamen, 2002).

Content validity is established by the researcher showing that the behaviours being sampled by the test are a representative sample of the attribute being measured (Murphy & DavidsHofer, 1991). The APIL-SV in essence aims to measure the attribute of a person's ability to process information and their learning potential. In this, the test is subdivided into four measureable dimensions: Conceptual Reasoning, Learning Rate, Level of Performance at the Conclusion of the Learning Exercise and Memory and Understanding of the Learning Material (Aprolab, 2004).

The Conceptual Reasoning dimension assesses the ability to reason abstractly and to recognise underlying structures, even if members of the category do not necessarily physically resemble one another (Aprolab, 2004). This relates to the Concept Formation Test in the battery, and is comprised of 33 items of a quasi-geometrical nature. The respondent is faced with a row of six diagrams, with one of

the diagrams being anomalous in that it does not follow the principle that all the other diagrams in the row follow (Aprolab, 2004). The test manual states that this type of activity is essential to all tasks that are more intellectually demanding, where a person will have to engage with new material and where established methods cannot be applied (APROLAB). This relates to the domain of ability, that is, the processing of information and learning potential, in that it examines the respondents' ability to process information by uncovering unwritten and undefined rules within the diagrams.

The dimension of Learning Rate is a dynamic measure, whereas the Conceptual Reasoning dimension is a static measure (Aprolab, 2004). This dimension is an index of improvement in performance that has resulted through practice, instruction and other learning opportunities (Aprolab, 2004). The test contains two symbols side by side, and is required to translate both symbols into two symbol-related words using an eight page dictionary (Aprolab, 2004). These symbols are categorised into four categories: clothing, transport, colours and numbers (Aprolab, 2004). The first symbol of the two that the respondent is given is always from the colours or numbers category, and the second is from the clothing or transport categories (Aprolab, 2004). The dictionary contains material from only one category, and there are only 12 stimuli per page, with each stimulus having two steps for each translation (Aprolab, 2004). Each of the four categories vary in difficulty, with the colours being the easiest to learn as there are associative elements that help the individual to cognise the material (Aprolab, 2004). The transport category, on the other hand, is the hardest category for the respondent to cognise as the respondent is required to make their own associations in the translation (Aprolab, 2004). Although a distinct construct, the Learning Rate construct is related to Conceptual Reasoning, as it requires respondents to engage and master new material and to become more efficient in processing the information in the material (Aprolab, 2004). This dimension relates back to the construct in that it examines the respondents' learning ability.

The Level of Performance at the Conclusion of the Learning Exercise score is reflected as a result of an initial practice, a lesson and a study period (Aprolab, 2004). If a respondent has performed well, it can be attributed to the mastery of the learning material: whereas a poor performance is due to no mastery of the learning material to

any significant level (Aprolab, 2004). In relating to the construct of ability, processing of information and learning potential, this dimension relates to the construct in that the respondents are expected to process information and to learn in two test sessions using the same symbols.

The final dimension, Memory and Understanding of the Learning Material, is described as a power test, in that it assesses the level of mastery of the learning material, with some items requiring the respondent to work with information that goes beyond what was contained in the learning task (Aprolab, 2004). For a respondent to obtain a high score that respondent should have almost or fully mastered the learning material (Aprolab, 2004). This dimension relates to the construct in that it assesses the learning of the respondent including the respondent's ability to process further information on top of what the respondent learned in the previous test.

In conclusion, the APIL-SV shows a relatively high degree of content validity, in that all the dimensions and items in the dimensions relate back to the content.

One of the most important areas of validity in psychometric terms is criterion validity, as it is a quantitative procedure that calculates correlation coefficients between the predictor and criterion, and can be evaluated in two ways (Cohen & Swerdlik, 2005). Concurrent validity is a method employed to evaluate a predictor's criterion validity, as it assesses the accuracy of the measure in identifying or diagnosing a specific skill or characteristic of the individual (Cohen & Swerdlik, 2005). The final criterion validity measure is predictive validity, which refers to the accuracy of the measure in predicting the future status or behaviour of the individual (Cohen & Swerdlik, 2005). However, when assessing criterion validity, the researcher needs to be aware of criterion contamination, which can manifest in various forms when ratings are based on judgements on the individual, known as the "halo effect" (Cohen & Swerdlik, 2005).

The final area of validity is construct validity, which evaluates the extent to which the test measures the theoretical construct that it is meant to measure (Cohen & Swerdlik, 2005). One method of assessing the construct validity of a test is to investigate correlations with similar tests. Another method is to use the statistical

technique of factor analysis, which analyses the interrelationships of variables to find a common variance between them.

To investigate the validity on the Concept Formation Test (CFT), Taylor (1995) gave the CFT 33 to first-year university students, who had been admitted into the university on criteria that were not based on Matric marks. Taylor correlated the marks from their CFT tests with marks on a course they took, which involved logical thinking skills and reasoning skills. The correlation was 0.44 ( $p = 0.012$ ). In a validity study on the Curve of Learning and Memory and Understanding tests, 110 participants from a beverage manufacturing company took part. The performance rating for the workers included aspects such as their capacity to learn new procedures and concepts, to understand why things happen in an organisation as a whole, initiative, capacity to plan and organise, their efforts at self-development, and their capacity to supervise themselves, with an overall rating also available. The results seemed to average correlations of 0.35. Taylor has conducted many validity studies on the two subsections with people from different backgrounds.

### *Reliability*

The reliability of a psychological test refers to the extent to which there is consistency in the measurement or that the measurement is repeatable (Cohen & Swerdlik, 2005). Cohen and Swerdlik (2005) go on to describe reliability as being dependent on the differences in measurements of tests taken at different times, which is referred to as a function of measurement error. The measurement needs to be consistent in what it is measuring in order to prove its dependability. According to Murphy and Davidshofer (1991), and other authors, reliability is the first requirement of a good measure.

Variance refers to the difference between what is expected, and that which actually occurs. In essence, variance is the proportion of total variance that is made up of true variance (Cohen & Swerdlik, 2005). True variance is used to refer to the unobservable variance of a whole population.

To measure reliability, people can use test-retest reliability estimates, split-half reliability estimates and inter-item consistency (Cohen & Swerdlik, 2005). Split-half reliability refers to the division of the tests into equal halves, and computing the correlation between these two tests (Cohen & Swerdlik, 2005).

Taylor (1995) conducted reliability studies on six samples from relatively different backgrounds. These reliability studies employed the split-half reliability method, which correlated at 0.86, illustrating the reliability of the instrument.

### *Fairness*

Test fairness refers to how the test is administered and used for individuals (Cohen & Swerdlik, 2005). Essentially, Cohen and Swerdlik (2005) define fairness as the extent to which a test is used in an impartial and equitable way. Test developers strive to ensure test fairness through the development of a test manual, which describes in detail how the test should be conducted in terms of times, materials needed, the type of room the candidate should be in and so forth.

### *Bias*

According to Cohen and Swerdlik (2005) bias has different meanings for different people. They found that, for the general public, bias means prejudice and preferential treatment, whereas for judges bias refers to a test that is designed to be too difficult for one group in comparison to another. Cohen and Swerdlik (2005) define test bias as a factor that is built into the test naturally and that systematically prevents any form of accurate measurement.

History has prevailed, and in both in the western world and South Africa, results show that the black population scores lower in cognitive ability tests than the white population (Jensen, 1980). Despite past societal norms where men were treated as superior to women, Jensen (1998) found that men and women from equal backgrounds tend to have similar average overall scores of cognitive ability. Jensen did find however that performance between gender groups could differ slightly on some subtests of cognitive ability tests.

Researchers have highlighted the fact that cultural bias exists in cognitive tests, which can be attributed to history's treatment of different cultures (Foxcroft *et al.*, 2003). A study by Roth, Bevier, Bobko, Switzer and Tyler (2001) confirms that there seems to be a substantial black-white gap between cognitive ability tests and job performance. Chung-Yan and Cronshaw (2002) found that bias could have an impact on the organisation, as black workers who should have been selected for the job based

on their job performance ratings would have been eliminated incorrectly due to scores on cognitive ability tests.

Murphy (2002) raises a contentious issue stating that, if organisations use a cognitive ability assessment as a predictor of future performance, this may have a substantially adverse impact on people of ethnic/racial groups in applying for employment opportunities. This is further supported by a discussion by Schmidt (2002) where people of ethnic/racial groups are at a disadvantage in acquiring employment that is a result of the mean score difference between racial groups in cognitive ability testing. Kehoe (2002) suggests that, in order to counteract the difference in mean scores attributed to racial groups, psychologists should explore different methods for modifying cognitive tests specifically to reduce these group differences. Decorte (1999) suggests using a weighting system that allows employers to determine a desirable balance between utility and increasing workforce diversity.

A European meta-analysis study conducted by Salgado and Anderson (2002) examines differences between the Spanish and English cultures. The Spanish culture is predominantly centred on collectivism; whereas the English culture is predominantly individualistic. The study illustrates that cultural differences may be less influential on criterion validity than previously believed; illustrating other factors that impact on test results creating the population group differences. This study discovered that cognitive ability tests are independent of national boundaries, suggesting cross-national validity generalisation.

Taylor (1995) conducted a study on 66 individuals with obvious African surnames, with 400 individuals from other races classed as the white subsample. Both races' slope values were similar. In the investigation, a  $t$ -value of 0.033 was obtained with a 462df, which is non-significant. Taylor believes that this was related to the small black subsample size.

#### *Past studies on the APIL-B*

Owing to a lack of available published literature on the APIL-SV, it would seem that no investigations have been conducted on this psychometric tool. Therefore all comparisons made in this study will be related to the APIL-B.

Lopes *et al.* (2001) conducted a study to ascertain the predictive validity of the APIL-B in a financial institution. The sample consisted of 235 successful job applicants in the company. They found that the APIL-B is a useful instrument for

predicting how the individual will perform within the organisation. This would seem to illustrate that the APIL-B is a reliable instrument when it is applied to job applicants within a financial industry setting. They found that the APIL-B does measure construct effectively, but they refrain from stating whether the tool is valid for selection purposes. They found that African and coloured groups consistently scored lower in comparison to Indians and whites; however, they found that with the available evidence assessors did not discriminate against ethnic groups. They recommend the use of a larger sample group to explore potential bias in a future study.

Twigge, Theron, Steele and Meiring (2005) conducted a study using the APIL-B as a predictor of performance of students in an MBA programme consisting of 101 participants. The correlation between the predictor and the criterion illustrated moderate levels of correlation ( $r = 0.416$ ,  $p = 0.000$ ). Further investigation into the regression of the study showed that approximately 17% of criterion variance can be attributed to the candidates' performance on the APIL-B. The sample size used for this study proved to be inadequate for determining any racial bias.

In an unpublished study by Makgoatha (n.d.), the author set out to discover whether the APIL-B predicts job performance in a global consulting firm and assess whether bias exists towards different population groups. The study found that there was a difference in mean scores of approximately one standard deviation. However, further analysis showed that individuals who performed at the same level on the APIL-B had the same predicted performance level, despite their racial membership. This illustrates that the APIL-B can predict job performance equally across race groups.

#### *Research questions*

1. Is the APIL-SV a valid predictor of potential future performance of individuals within a financial/consulting industry?
2. Is there a difference in scores on the APIL-SV depending on gender and race?
3. Does the APIL-SV have construct validity?

## Method

### *Sample*

All applicants in the sample participated in the APIL-SV assessment on applying to a global financial consulting organisation based in South Africa. Owing to the nature of data collected, there are two sample sets for analysing the test. The samples have been collected since the beginning of 2006.

#### *First group – factor analysis sample*

The first sample consists of applicants who applied to the organisation for various reasons from bursaries through to executive positions in all the business units in the organisation. The sample consists of 2 877 applicants, with an age range from 16–56 years. The range of education in this sample set is from Grade 11 through to a master's qualification, with the majority of the applicants holding a matric (34%) or bachelor's degree (34%).

Table 1.1

*Descriptive statistics used for factor analysis examining the distribution between gender and race in the sample*

		Distribution of frequency	Cumulative percentage of distribution
Gender	Female	1446	50.3
	Male	1431	49.7
	Total	2877	100.0
Race	Black	1168	40.6
	Coloured	127	4.4
	Indian	630	21.9
	White	952	33.1
	Total	2877	100.0

---

The second data sample, which is used to determine whether the instrument predicts performance and learning ability, can be divided into two further subsamples.

*Second group - performance on the Public Practice Examination sample*

The second subsample consists of 69 applicants who have sat the IRBA Public Practice Exam with an age range of 18–33 years (with the mean age being 23.25 years, and the mode being 22 years). These applicants work in areas such as auditing, general, training, and UCT students.

Table 1.2

*Descriptive statistics for the sample used for to examine the effectiveness of the APIL-SV in predicting performance on the public practice exam, examining the distribution between gender, race and education in the sample*

		Distribution of sample between groups	Cumulative percentage distribution	
Gender	Female	29	42.0	
	Male	40	58.0	
	Total	69	100.0	
Education	Chartered Accountant	1	1.4	
	CTA	5	7.2	
	Degree	39	56.5	
	Honours	13	18.8	
	Masters	1	1.4	
	Matric	9	13.0	
	Diploma	1	1.4	
	Total	69	100.0	
	Race	Black	10	14.5
		Coloured	1	1.4
Indian		23	33.3	
White		35	50.7	
Total		69	100.0	

### *Data analysis*

With the use of SPSS, the following statistical analyses will be conducted on the data obtained: descriptive statistics (e.g. means, standard deviations and ranges), correlation coefficients, discriminate analysis, and factor analysis of the different samples.

### *Measuring instruments*

#### *The predictor measure*

The APIL-B was designed to ascertain whether employees can be developed to fill a higher managerial role or specialised skills (APROLAB, 2004). The APIL-SV is a shortened version of the APIL-B, and has been condensed into about two hours (Aprolab, 2004). The APIL-SV comprises the following areas:

- Fluid Intelligence Test

The content of the Fluid Intelligence Test is identical to the normal APIL, and is known as the Concept Formation Test (CFT). The test contains 33 items of a quasi-geometrical nature. The task of the candidates is to identify the inconsistent diagram in the pattern. In identifying this pattern, the candidate has exhibited his/her ability to understand the underlying concept to the question, which the other five diagrams share. The candidates can score a minimum of 0 points to a maximum of 33 points. In the current data samples the minimum score was 3 with the maximum reaching 33 points.

- Curve of Learning Test

The material found in the APIL-SV Curve of Learning Test is the same as to that of the original APIL; however, the difference is designed to mirror real-life training situations (Aprolab, 2004). The APIL-SV mirrors training sessions by presenting candidates with material and providing them with training on the material, which they are then tested on, much like a training session in an organisation where the candidate is expected to know the material after the training session. The APIL-B consists of four sessions with study periods in between without a training scenario; the APIL-SV Session 2 contains

more items than the standard 30 items. In Session 1 of the APIL-SV the candidate has 12 minutes to answer 30 double items; in the last session, the candidate has nine minutes to answer 70 double items.

The candidate is presented with two symbols side by side, which require a double translation into a simple word, using the supplied eight-page dictionary. The symbols are divided into four categories: clothing, transport, colours and numbers. Each page in the dictionary contains material from one category, with is part of the double translation creating two steps in the translation.

The four categories vary in difficulty, where colours and clothing have a logical associative element, which makes it easier for the candidate to recognise the material. The numbers category offers a greater challenge as the candidate has to work out how to relate an intricate geometrical structure to a given number. The transport category is the hardest as it requires the candidates to make their own associations for the second step of the translation with little relevance to the final meaning.

In applying a statistical analysis to this test to gain an understanding of the data, all the answers from the two study sessions are added together to yield a COL Total (COLtot) score. The minimum an applicant can score is 0 points to a maximum of 200 points: the range for the current data samples is 38 to 200. To acquire the curve of learning the total points from the first session is subtracted from the total points of the first session. This is known as the COL difference (COLdif) score, where candidates' scores range from -42 to 109 with the current data samples.

- **Memory and Understanding Test**

The Memory and Understanding Test is intended to uncover the individual's level of knowledge regarding the material in the dictionary. The test in the APIL-SV is the same as the APIL-B except that one existing item has been replaced with a more demanding item, with two additional difficult items. These changes were implemented due to the lesson enabling the candidate to do better than before, therefore preventing the test from becoming too easy for candidates.

On the memory and understanding tests, the minimum a candidate can score is 0 points to a maximum score of 28 points.

It has been suggested that the APIL-SV is "easier" than the APIL-B as it provides candidates with more structure (Aprolab, 2004). In response to this respondents are given more work to do in the APIL-SV than in the APIL-B. As extra items, the developer has included additional difficult items on the Memory and Understanding test, and on the Concept Formation Test (CFT).

The APIL-SV assesses the following dimensions within the candidate's ability (Aprolab, 2004):

- **Conceptual Reasoning or Fluid Intelligence**

This is the candidate's ability to reason abstractly and to recognise any underlying structures when objects in the category do not necessarily physically resemble each other. The candidate is required to identify underlying rules in order to classify objects presented to them. This tends to be a static measure in that a person's conceptual reasoning can rarely be altered to improve performance.

The use of this subtest is seen to be essential for tasks which are more intellectually demanding, where the candidate will have to grapple with novel material and where current procedures cannot be applied.

- **Learning Rate**

Learning rate is a dynamic measure, unlike the static measure of conceptual reasoning. Learning rate is an index of improvement of performance as a result of given practice, instruction and consolidation.

Although this is a distinct construct, it is related to conceptual reasoning, as conceptual reasoning is required for the candidate to master new material and to become efficient in processing the material.

- **Level of Performance at the Conclusion of the Learning Exercise**

This score reflects the candidate's attained level of performance with the learning material as an output of the initial practice, a lesson and a study period within a limited amount of time.

- Memory and Understanding of the Learning Material

Unlike the above score, the Memory and Understanding of the Learning Material score is a pure power test, in that it assesses the person's level of mastery of the learning material. The final items in this test require the candidate to work with information that goes beyond what was given in the learning task.

#### *The criterion measure*

The criterion measure that is applied to a small number of candidates is based on their ability to pass the exam set by the Independent Regulatory Board for Auditors (IRBA). The IRBA exists to register public accountants and auditors, to regulate training of public accountants and auditors, and to administer related matters. All information discussed in the following paragraphs regarding the PPE was sourced from the IRBA (2007). Each of the candidates writes the Public Practice Examination (PPE), as it is a final requisite for qualifying as a Chartered Accountant of South Africa (CA(SA)) should the candidate have done their training contract (articles) in a public practice. These in a sense are the traditional articles offered at an auditing firm, as opposed to the financial management articles offered by some of the larger financial institutions. The exam can be used to become a Registered Auditor (RA) in South Africa, however most people do it with the sole purpose of attaining the CA(SA) qualification.

The exam content is based approximately on 40% auditing, with the balance being split between financial accounting, management accounting and taxation, as seen in the curriculum in the appendix. The exam is considered to be less complex than the South African Institute of Chartered Accountants (SAICA) Qualifying Exam (QE) Board 1.

In order to be considered as having completed the PPE, candidates need to have completed an appropriate undergraduate degree, and can then complete the CTA, which permits candidates to write the SAICA QE1. The SAICA QE1 exam has to be completed before candidates are eligible to write the PPE; however, it should be noted that the PPE cannot be attempted until candidates have completed at least 18 months of their training contract.

Once qualified, candidates are entitled to register as members of the IRB and SAICA, entitling them to use the designations RA and CA(SA) respectively. It allows the candidate to attest audit information on financial information and publications.

The job that is reserved for people with this qualification is that of an auditor; however, one has to be registered with the IRBA in order to express an opinion on the validity of any financial information.

The reason for the performance variable being given to the researcher in a dichotomous format is that this is how the SAICA board releases the results, that is, as either a pass or a fail.

### *Procedure*

The data collected for the present study were obtained over a period of approximately 18 months as secondary data from a consulting company, which offers this type of assessment to the financial consulting organisation. Should data have been collected over a longer period of time, the sample size would have been bigger, however, the researcher would then have needed to ensure that all performance data collected to assess the predictive validity were of the same time period. Owing to the fact that this is secondary data, it is unknown whether the candidates had had previous exposure to the test.

The test was administered over a period of two hours, in a well-lit, well-ventilated, quiet room in the client's building, which was made off-limits to employees. When candidates arrived for assessment, the desks were prepared with pencils and an eraser, with the answer booklets already laid out.

When the candidates arrive they are asked to fill in the biographical information on the first page of the answer booklet and then the psychologist goes through the instructions for each section with the candidates.

The first test for the CFT is administered in 40 minutes, with 33 questions to answer. In session 1 of the Curve of Learning Test (COL), the respondent receives an additional booklet called the dictionary. The candidate has 12 minutes to answer 30 double questions. Thereafter the administrator gives a COL lesson, and hands out an additional booklet called the "APIL Lesson Book" which condenses the eight-page Dictionary into four pages. The administrator goes through the lesson book, whilst the candidates follow the lesson. After each section in the lesson book, candidates are

given 30 seconds to study the material. At the end of the last section, the administrator gives the candidates five minutes for self-study of the given material. At the end of the study session, the lesson book is taken away from the candidates. Session 2 of the COL requires the candidates to answer 70 double questions in nine minutes, after which they are stopped and the administrator collects the dictionaries. The final test, the Memory and Understanding test, requires the candidates to answer 28 cognitive questions in 12 minutes without the use of the dictionary. This concludes the APIL-SV.

This data was then given to the researcher in conjunction with the data obtained from the organisation on the performance of candidates who were subsequently hired as a result of this test

University of Cape Town

## Results

### *Public Practice Examination*

The data sample used for the PPE was collected over the past year on the performance on the board exam of internal employees in the consulting company.

Table 2.1

*Descriptive statistics of scores examining the subtests of the APIL-SV*

	<u>N</u>	<u>Minimum</u>	<u>Maximum</u>	<u>Mean</u>	<u>Std. Deviation</u>	<u>Skewness</u>	<u>Std. Error</u>	<u>Kurtosis</u>	<u>Std. Error</u>
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Error	Statistic	Error
CFT	69	12	33	24.17	4.920	-.458	.289	-.723	.570
Memory Correct	69	12	27	21.93	2.992	-1.099	.289	1.033	.570
Col Total	69	42	198	112.55	35.305	.224	.289	-.597	.570
Col Diff	69	-10	106	40.70	24.240	.245	.289	-.361	.570

Table 2.2

*Group statistics between dichotomous variables for the PPE.*

	Pass/Fail	Mean	Std. Deviation	Valid N
CFT	Pass	25.33	4.259	46
	Fail	21.87	5.421	23
Memory Correct	Pass	22.43	2.664	46
	Fail	20.91	3.397	23
Col Total	Pass	120.09	34.725	46
	Fail	97.48	32.093	23
Col Diff	Pass	46.39	22.844	46
	Fail	97.48	32.093	23
Total	CFT	24.17	4.920	69
	Memory Correct	21.93	2.992	69
	Col Total	112.55	35.305	69
	Col Diff	40.70	24.240	69

Owing to the performance criterion being a dichotomous variable, a discriminant analysis was conducted on the data. None of the data was excluded from processing the analysis, as no data was missing discriminating values, missing or out-of range group codes, or both elements combined. Table 2.2 illustrates the group statistics of the mean and standard deviation between the performance ratings.

Analysis of variance (ANOVA) was used in the discriminant analysis (Table 2.3): all subtests prove to be significant, but the CFT and COL Diff are significantly more significant: COLdif ( $F(67)= 8.454, p= .005$ ), CFT ( $F(67)= 8.391, p= .005$ ), COLtot ( $F(67)= 6.827, p= .11$ ) and the Memory: Correct ( $F(67)= 4.151, p= .46$ ) which allows us to reject the null hypothesis and conclude that the subtests of the

APIL-SV have the ability to determine whether the applicant will succeed in the PPE exam.

Table 2.3

*ANOVA on the PPE Sample*

	Wilks'				
	Lambda	F	df1	df2	Sig.
CFT	.889	8.391	1	67	.005
Memory Correct	.942	4.151	1	67	.046
COL Total	.908	6.827	1	67	.011
COL Diff	.888	8.454	1	67	.005

Table 2.4 shows factors that were entered into the analysis, further illustrating the only subtest that minimises Wilks' Lambda is the COLdif with a score of 0.888 respectively.

Table 2.4

*Variables entered/removed during ANOVA on the PPE Sample*

Step	<u>Entered</u>			<u>Wilks' Lambda</u>					
	Statistic	df2	Sig.	Exact F					
				Statistic	df1	df2	Sig.	Statistic	df1
1	COLdif	.888	i	1	67.000	8.454	1	67.000	.005

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- a Maximum number of steps is 8.
- b Minimum partial F to enter is 3.84.
- c Maximum partial F to remove is 2.71.
- d F level, tolerance, or VIN insufficient for further computation.

Table 2.5 shows variables which are not in the analysis at each step. At step zero, none of the variables were included in the analysis. At step one, COLdif was entered. When COLdif was entered as a variable, the Wilks' Lambda and Tolerance of each subtest are lowered.

Table 2.5  
*Variables in the analysis of the PPE sample*

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	CFT	1.000	1.000	8.391	.889
	Memory Correct	1.000	1.000	4.151	.942
	COLtot	1.000	1.000	6.827	.908
	COLdif	1.000	1.000	8.454	.888
1	CFT	.845	.845	3.177	.847
	Memory Correct	.926	.926	1.465	.869
	COLtot	.274	.274	.059	.887

Tables 2.6 and 2.7 illustrate the percentage of the variance accounted for by the one discriminant function generated. Table 2.6 illustrates the variance of the discriminant function, which accounts for 100% of the variance of each function.

Table 2.6  
*Eigenvalues for the PPE Sample*

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.126(a)	100.0	100.0	.335

a. First 1 canonical discriminant functions were used in the analysis.

Table 2.7  
*Wilks' Lambda for the PPE Sample*

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.888	7.903	1	.005

The test of the function in table 2.7 illustrates that the function is significant ( $p = .005$ ). Because there are only two classification categories (pass/fail), only one discriminant function was generated.

The output from the structure matrix illustrates that the association between the variables Memory Correct and CFT and the discriminant function is very small. There is a high level of association between the discriminant function and the variable COLdif. The table also shows a fairly high level of association between the discriminant function and the variable COLtot. However, the variables COLtot, CFT, and Memory Correct were not used in the analysis.

Table 2.8  
*Structure matrix of the APIL-SV for the dichotomous PPE sample*

	Function
	1
COLdif	1.000
COLtot(a)	.852
CFT(a)	.394
Memory	
Correct(a)	.273

Pooled within-groups correlations between discriminating variables and standardised canonical discriminant functions

Variables ordered by absolute size of correlation within function.

a This variable not used in the analysis.

The analysis of the Functions at group centroids, with unstandardised canonical discriminant functions evaluated group means. This illustrates how distinct the two groups are in their reflection; however, they do seem to remain a little close, as seen in table 2.9.

Table 2.9

*Group centroids for the dichotomous PPE sample*

	<u>Function</u>
Pass/Fail	1
Pass	.248
Fail	-.495

Unstandardised canonical discriminant functions  
evaluated at group means

Table 2.10 represents the classification of results and provides an indication of the success rate for the predictions of membership of the criterion group's variables. The table indicates that there is a 59.4% likelihood of the original grouped cases being correctly classified.

Table 2.10

*Classification of results for the dichotomous PPE sample*

		<u>Predicted group</u>			<u>Total</u>
		<u>membership</u>			
	<u>Pass/Fail</u>	<u>Pass</u>	<u>Fail</u>		
Original	Count	Pass	27	19	46
		Fail	9	14	23
	%	Pass	58.7	41.3	100.0
		Fail	39.1	60.9	100.0

a 59.4% of original grouped cases correctly classified.

Whilst table 2.10 shows that people who are likely to fail the PPE exam are most accurately classified with 60.9% of the cases correct. On the other hand, only 58.7% of the pass cases are correct, which suggests that 41.3% of the cases that passed the PPE could have failed the exam. The opposite, however, is seen in the fail group, where 39.1% of the cases could have passed the exam. Even though there are a great number of correct predictions in the pass and fail groups, the small difference between the correct and incorrect predictions is too small to state whether or not the APIL-SV has a high level of predictive validity, although it does predict.

Differences between the presence of the subtests and the candidate's performance level in the PPE exam were analysed using *t*-tests. Table 2.11 illustrates

the group statistics, whereas table 2.13 illustrates the Independent Samples Test, using Levene's test for equality of variances to ascertain which *t*-test for equality of means the researcher should use.

Table 2.11

*T-test group statistics for performance*

	Pass/Fail	N	Mean	Std. Deviation	Std. Error Mean
CFT	Pass	46	25.33	4.259	.628
	Fail	23	21.87	5.421	1.130
Memory Correct	Pass	46	22.43	2.664	.393
	Fail	23	20.91	3.397	.708
Col Total	Pass	46	120.09	34.725	5.120
	Fail	23	97.48	32.093	6.692
Col Dif	Pass	46	46.39	22.844	3.368
	Fail	23	29.30	23.350	4.869

The results from the *t*-tests in table 2.13 assumed that all variances were equal as the *p*-values for *F* are greater than 0.05. The results found that overall the COL subtests have the greatest influence on predicting the candidates' ability to pass the PPE. COLtot has the greatest impact on the candidates' ability to pass or fail the PPE exams ( $M = 22.609$ ) ( $t(67) = 2.613$ ;  $p = .011$ ). COLdif ( $M = 17.087$ ) has the second highest impact on the candidates' ability to pass the PPE exams ( $t(67) = 2.908$ ;  $p = .005$ ). The candidates' score on the CFT also reveals significant insight into the candidates' ability to perform on the PPE ( $M = 3.457$ ) ( $t(67) = 2.897$ ;  $p = .005$ ). The *t*-tests reveal that the Memory Correct subtest does have an impact on the candidates' ability to pass the PPE exam ( $M = 1.522$ ); however, this is not as significant as the above subtests ( $t(67) = 2.037$ ;  $p = .046$ ) discussed.

Table 2.12

*T-test group statistics for gender regarding the PPE sample*

	Gender	N	Mean	Std. deviation	Std. error mean
CFT	Female	29	22.90	4.716	.876
	Male	40	25.10	4.914	.777
Memory Correct	Female	29	22.21	2.664	.495
	Male	40	21.73	3.226	.510
Col Total	Female	29	113.24	29.847	5.542
	Male	40	112.05	39.155	6.191
Col Diff	Female	29	38.90	20.428	3.793
	Male	40	42.00	26.846	4.245

A further study in table 2.14 illustrated that in predicting PPE performance, the APIL-SV showed no gender bias between males and females taking the APIL-SV and their performance on the PPE, despite the differences in mean as seen in table 2.13 specifically, with a difference in means on the COLdif where males ( $M=42.00$ ) scored a higher mean than females ( $M=38.90$ ). Inclusive of a difference of means regarding the CFT subtests where males ( $M=25.10$ ) scored higher than females ( $M=22.90$ ). The small sample size did not allow for a further examination on racial bias to be conducted using any form of statistical measure.

Table 2.13

*T-test results for the PPE performance variable*

		<u>Levene's test for equality</u>		<u>t-test for equality of means</u>						
		<u>of variances</u>								
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
CFT	Equal variances assumed	1.994	.163	2.897	67	.005	3.457	1.193	1.075	5.838
	Equal variances not assumed			2.673	35.994	.011	3.457	1.293	.834	6.079
Memory Correct	Equal variances assumed	3.329	.073	2.037	67	.046	1.522	.747	.031	3.013
	Equal variances not assumed			1.879	35.949	.068	1.522	.810	-.121	3.164
Col Total	Equal variances assumed	.079	.780	2.613	67	.011	22.609	8.653	5.337	39.880
	Equal variances not assumed			2.683	47.360	.010	22.609	8.426	5.661	39.556
Col Diff	Equal variances assumed	.338	.563	2.908	67	.005	17.087	5.877	5.357	28.817
	Equal variances not assumed			2.886	43.252	.006	17.087	5.920	5.150	29.024

Table 2.14

*T-test results for the gender variable for the PPE sample*

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
	Equal variances not assumed			-.545	66.813	.587	-3.103	5.693	-14.467	8.260
CFT	Equal variances assumed	.065	.799	-1.870	67	.066	-2.203	1.178	-4.556	.149
	Equal variances not assumed			-1.882	61.894	.065	-2.203	1.171	-4.544	.137
Memory Correct	Equal variances assumed	1.441	.234	.658	67	.513	.482	.733	-.981	1.944
	Equal variances not assumed			.678	65.794	.500	.482	.711	-.937	1.901
Col Total	Equal variances assumed	1.774	.187	.137	67	.891	1.191	8.673	-16.121	18.504
	Equal variances not assumed			.143	66.800	.886	1.191	8.309	-15.395	17.778
Col Diff	Equal variances assumed	2.143	.148	-.522	67	.603	-3.103	5.944	-14.967	8.760

*Factor analysis sample*

As mentioned above the sample used for the factor analysis study comprises all applicants to the consulting company over the past two years, with table 3.1 representing the descriptive statistics of the subtests.

Table 3.1

*Factor analysis sample descriptive statistics*

	<u>N</u>	<u>Minimum</u>	<u>Maximum</u>	<u>Mean</u>	<u>Std.</u> <u>deviation</u>	<u>Skewness</u>	<u>Std.</u> <u>Error</u>	<u>Kurtosis</u>	<u>Std.</u> <u>Error</u>
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
CFT	2877	3	33	20.66	5.926	-.341	.046	-.597	.046
Memory Correct	2877	0	28	19.68	4.389	-.858	.046	.749	.046
COL tot	2877	0	200	97.37	36.567	.336	.046	-.441	.046
COL dif	2877	-16	109	30.07	23.209	.344	.046	-.557	.046

Table 3.2 illustrates levels of correlation between the subtests, which all show high levels of correlation. As seen in the results above, there is a high level of correlation between COLdif and COLtot 0.875, with the lowest levels of correlation found between COLdif and Memory Correct 0.579. However, with both extremes being taken into consideration, the average correlation between subtests is 0.665, which still proves to be extremely high for tests.

Table 3.2

*Correlation matrix of the Factor Analysis Sample*

		Memory			
		CFT	Correct	COL Tot	COL dif
Correlation	CFT	1.000			
	Memory Correct	.632	1.000		
	COL Tot	.654	.679	1.000	
	COL dif	.569	.579	.875	1.000

Table 3.3  
*Communalities*

	Initial	Extraction
CFT	1.000	.668
Memory Correct	1.000	.688
COLtot	1.000	.871
COLdif	1.000	.775

Extraction Method: Principal Component Analysis.

The communalities in table 3.3 illustrate that 87.1% of the variance of the COLtot subtest is accounted for by the factors; whereas only 66.8% of the variance in the Memory Correct is accounted for by the factors. The average amount of variance accounted for by one factor (table 3.4) is 75.037%, which influences the results in the subtests.

Table 3.4  
*Total variance of the factor analysis sample explained*

Component	<u>Initial eigenvalues</u>			<u>Extraction sums of squared loadings</u>		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.001	75.037	75.037	3.001	75.037	75.037
2	.519	12.987	88.024			
3	.367	9.186	97.210			
4	.112	2.790	100.000			

Extraction Method: Principal Component Analysis.

The component matrix illustrated in table 3.5 shows that the psychometric test is in fact measuring one component that it is meant to, that is, cognitive ability.

Table 3.5

*Component matrix of the APIL-SV in the factor analysis sample*

	<u>Component</u>
	1
CFT	.817
Memory	.830
Correct	.933
COL Tot	.933
COL dif	.880

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

The relatively low residual levels in table 3.6 illustrate a better fit between their reproduced correlations and the original correlations, such as Memory Correct and CFT (-0.046).

The *t*-test was employed to investigate potential gender differences in the candidates' scores on the subtests. Table 3.7 illustrates the group statistics for the subtests.

Table 3.6

*Reproduced correlations for the factor analysis sample of the APIL-SV*

		CFT	Memory Correct	COLtot	COLdif
Reproduced Correlation	CFT	.668(b)			
	Memory Correct	.678	.688(b)		
	COLtot	.763	.774	.871(b)	
	COLdif	.719	.730	.821	.775(b)
Residual(a)	CFT				
	Memory Correct	-.046			
	COLtot	-.109	-.095		
	COLdif	-.150	-.151	.053	

Extraction method: Principal Component Analysis.

a. Residuals are computed between observed and reproduced correlations. There are 5 (83.0%) nonredundant residuals with absolute values greater than 0.05.

b. Reproduced communalities

Table 3.7

*T-test gender group statistics for the factor analysis sample*

	Gender	N	Mean	Std. deviation	Std. error mean
CFT	Female	1446	20.32	5.791	.152
	Male	1431	20.99	6.043	.160
Memory Correct	Female	1446	19.98	4.061	.107
	Male	1431	19.38	4.679	.124
COL Tot	Female	1446	100.03	35.222	.926
	Male	1431	94.67	37.699	.997
COL diff	Female	1446	30.19	22.634	.595
	Male	1431	29.95	23.783	.629

Table 3.8 illustrates the Independent Samples Test, using Levene's test for equality of variances, to ascertain which *t*-test for equality of means the researcher should use. The results from table 3.8 do not assume that all variances are equal, as the *p*-value for *F* is less than 0.05. The *t*-test illustrates a potential gender bias in the test; the only test that shows no gender bias is the COLdif subtest. The three remaining subtests all illustrated high levels of significance when related to gender differences. The COLtot test, on the other hand, seems to have a greater significant influence of the performance of gender groups in the test ( $t(2857.491) = 3.937$ ;  $p = .000$ ). There is a significant difference between the gender performance on the Memory Correct subtest ( $t(2810.868) = 3.726$ ;  $p = .000$ ) and the CFT test ( $t(2866.969) = 3.052$ ;  $p = .002$ ).

Table 3. 8

*Gender t-test on the factor analysis sample*

		<u>Levene's test for equality</u>		<u>t-test for equality of means</u>						
		<u>of variances</u>								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
CFT	Equal variances assumed	4.770	.029	-3.052	2875	.002	-.674	.221	-1.106	-.241
	Equal variances not assumed			-3.052	2866.969	.002	-.674	.221	-1.106	-.241
Memory Correct	Equal variances assumed	32.008	.000	3.729	2875	.000	.609	.163	.289	.929
	Equal variances not assumed			3.726	2810.868	.000	.609	.163	.288	.929
COL Tot	Equal variances assumed	6.565	.010	3.938	2875	.000	5.356	1.360	2.689	8.023
	Equal variances not assumed			3.937	2857.491	.000	5.356	1.361	2.688	8.024
COL diff	Equal variances assumed	6.456	.011	.279	2875	.780	.242	.866	-1.455	1.939
	Equal variances not assumed			.279	2864.718	.780	.242	.866	-1.456	1.939

Based on the results of the ANOVA from the company performance sample, an ANOVA study was conducted on the sample for the factor analysis, which provided a large enough sample of applicants. The results in table 3.11 show a significant difference between groups on the subtests of the APIL-SV ( $p = 0.000$ ,  $p < 0.05$ ), with each subtest illustrating F ratios. In concurrence with the performance sample results, the COLtot reflects higher levels of inequality between racial groups ( $F(2.2873) = 395.086$ ,  $p = 0.000$ ).

Table 3.9

*ANOVA investigating racial differences on the APIL-SV for the factor analysis sample*

		Sum of squares	df	Mean square	F	Sig.
CFT	Between groups	23486.253	3	7828.751	290.170	.000
	Within groups	77513.079	2873	26.980		
	Total	100999.33	2876			
Memory Correct	Between groups	10571.038	3	3523.679	225.836	.000
	Within groups	44826.956	2873	15.603		
	Total	55397.994	2876			
COLtot	Between groups	1123166.9	3	374388.970	395.086	.000
	Within groups	2722494.9	2873	947.614		
	Total	3845661.8	2876			
COLdif	Between groups	213288.81	3	104429.605	242.772	.000
	Within groups	1235834.4	2873	430.155		
	Total	1549123.2	2876			

Table 3.10 illustrates a further breakdown of the inequalities of the scores between the different population groups using Turkey HSD for multiple comparisons. The subtest CFT illustrates the greatest difference between the white and black populations on the scores of the subtest ( $M = 6.632$ ,  $SD = .227$ ,  $p = .000$ ), with the lowest difference being between Indians and coloureds respectively ( $M = 1.480$ ,

$SD = .505, p = .018$ ). Once again on the Memory Correct subtest the population groups with the largest disparity in points is between white and black ( $M = 4.388, SD = .172, p = .000$ ), whereas the least disparity can be found between the coloured and black populations ( $M = 1.207, SD = .369, p = .006$ ).

The COL subtests continue to illustrate the discrepancy in scores between different racial population groups. COLtot illustrates the highest inconsistency of scores between white and black ( $M = 46.063, SD = 1.344, p = .000$ ), with the least disparity between Indians and coloureds ( $M = 8.586, SD = 2.994, p = .022$ ). The COLdif illustrates a score gap between population groups, particularly continuing to illuminate the disparity between white and black ( $M = 24.335, SD = .906, p = .000$ ). The COLdif subtest is the only subtest in the battery which shows no significant levels of racial discrimination between one group, coloureds and Indians ( $M = 4.870, SD = 2.017, p = .075$ ).

University of Cape Town

Table 3.10

*Tukey HSD on the factor analysis sample for the APIL-SV*

Dependent Variable	(I) Race	(J) Race	Mean difference		Sig.	95% confidence interval	
			(I-J)	Std. error			
CFT	Black	Coloured	-2.325(*)	.485	.000	-3.57	-1.08
		Indian	-3.804(*)	.257	.000	-4.46	-3.14
		White	-6.632(*)	.227	.000	-7.21	-6.05
	Coloured	Black	2.325(*)	.485	.000	1.08	3.57
		Indian	-1.480(*)	.505	.018	-2.78	-.18
		White	-4.307(*)	.491	.000	-5.57	-3.05
	Indian	Black	3.804(*)	.257	.000	3.14	4.46
		Coloured	1.480(*)	.505	.018	.18	2.78
		White	-2.827(*)	.267	.000	-3.51	-2.14
	White	Black	6.632(*)	.227	.000	6.05	7.21
		Coloured	4.307(*)	.491	.000	3.05	5.57
		Indian	2.827(*)	.267	.000	2.14	3.51
Memory Correct	Black	Coloured	-1.207(*)	.369	.006	-2.16	-.26
		Indian	-2.822(*)	.195	.000	-3.33	-2.33
		White	-4.388(*)	.172	.000	-4.83	-3.94
	Coloured	Black	1.207(*)	.369	.006	.26	2.16
		Indian	-1.625(*)	.384	.000	-2.61	-.64
		White	-3.191(*)	.373	.000	-4.14	-2.22
	Indian	Black	2.822(*)	.195	.000	2.33	3.33
		Coloured	1.625(*)	.384	.000	.64	2.61
		White	-1.555(*)	.203	.000	-2.08	-1.03
	White	Black	4.388(*)	.172	.000	3.94	4.83
		Coloured	3.191(*)	.373	.000	2.22	4.14
		Indian	1.555(*)	.203	.000	1.03	2.08
COL Tot	Black	Coloured	-15.675(*)	2.876	.000	-23.05	-8.26
		Indian	-24.230(*)	1.522	.000	-28.15	-20.33
		White	-46.063(*)	1.344	.000	-49.52	-42.61
	Coloured	Black	15.675(*)	2.876	.000	8.26	23.05
		Indian	-8.586(*)	2.994	.022	-16.28	-.89

		White	-30.408(*)	2.908	.000	-37.88	-22.93
	Indian	Black	24.240(*)	1.522	.000	20.33	28.15
		Coloured	8.586(*)	2.994	.022	.89	16.28
		White	-21.823(*)	1.581	.000	-25.89	-17.76
	White	Black	46.063(*)	1.344	.000	42.61	49.52
		Coloured	30.438(*)	2.908	.000	22.93	37.88
		Indian	21.823(*)	1.581	.000	17.76	25.89
COL diff	Black	Coloured	-6.319(*)	1.938	.006	-11.30	-1.34
		Indian	-11.189(*)	1.025	.000	-13.82	-8.55
		White	-24.335(*)	.906	.000	-26.66	-22.01
	Coloured	Black	6.319(*)	1.938	.006	1.34	11.30
		Indian	-4.870	2.017	.075	-10.06	.32
		White	-19.017(*)	1.959	.000	-23.05	-12.98
	Indian	Black	11.189(*)	1.025	.000	8.55	13.82
		Coloured	-4.870	2.017	.075	-3.32	10.06
		White	-13.147(*)	1.065	.000	-15.88	-10.41
	White	Black	24.335(*)	.906	.000	22.01	26.66
		Coloured	19.017(*)	1.959	.000	12.98	23.05
		Indian	13.147(*)	1.065	.000	10.41	15.88

\* The mean difference is significant at the .05 level.

## Discussion

The objective of the research was to investigate several areas of the APIL-SV. It would appear that this is the first study to be conducted on the APIL-SV. The first topic of investigation was to evaluate whether the APIL-SV was a valid predictor for a global financial/consulting organisation to use to ascertain the potential future performance of employees.

When used to predict the performance of employees on the PPE exam, the APIL-SV shows moderate to high levels of inter-test correlation. When correlation between the PPE exam and APIL-SV was explored, levels of correlation were average; however, this could be attributed to the small sample size obtained for this dataset. It should also be noted that the low correlation levels can be attributed to the range restriction of using only successful employees from the APIL-SV results. Nonetheless, these low correlation levels concur with the correlation levels in the study conducted by Twigge *et al.* (2005).

The canonical discriminant was used to discover whether the APIL-SV is a good predictor for the PPE. Unfortunately, the results were not as high as had been hoped for, with 59.4% accuracy of prediction of the rating. The APIL-SV has been proven to show higher levels of accuracy for this form of analysis than those of the APIL-B in the study conducted by Lopes *et al.* (2001) which scored 36.6% accuracy of the ratings predicted. Lopes *et al.* (2001) further suggest that, in a social science context, this is a relatively good indication of the ability of the APIL to predict job performance.

Although the researcher was given a secondary performance sample to use for the analysis, this sample was not used, due to the following reasons. Firstly, no information regarding how the performance measure is conducted was given. Secondly, even though the sample size was over 100 people, these people were not in the same level of job, nor in the same department. Therefore one could not analyse the data, due to criterion contamination.

From the past studies it is clear to see that APIL-B has the ability to predict future performance of employees. Similarly, the performance of employees who wrote the PPE and the results of the APIL-SV illustrate the construct validity of the test as opposed to the APIL-B, which further shows that it has predictive validity

owing to the high correlation levels. In general, from using the PPE sample, one could say that the APIL-SV has predictive validity and illustrates construct validity in that the current test scores correlate with previous test scores on the APIL-B.

The second part of the study was to discuss potential performance differences on the APIL-SV of scores. It is known and researched that psychometric tests in general contain a certain level of bias, particularly towards previously disadvantaged population groups. Owing to the small sample size of candidates writing the PPE exam, this dataset could not be used to investigate potential racial bias in the subtests. Therefore, the data sample used to investigate this research question was the factor analysis sample, containing 2 877 participants.

T-tests were used to ascertain any potential levels of gender discrimination that exist between the subtests. Owing to the sample sizes, only the factor analysis sample could uncover any gender bias in all the subtests. The t-tests from the factor analysis sample reveal that there are significant levels of gender difference in the scores of the subtests. The results from the factor analysis sample concur with Jensen's (1998) findings, showing significant levels of difference between male and female scores on the subtests, with males scoring consistently higher.

The first dataset from the performance sample consistently illustrates a greater racial bias between the white population and the coloured population on all subtests (CFT ( $M=6.030$ ,  $p=.003$ ), Memory Correct ( $M=3.981$ ,  $p=.005$ ), COLtot ( $M=46.085$ ,  $p=.000$ ) and COLdif ( $M=31.974$ ,  $p=.000$ )). However, this data sample also illustrated levels of bias between the white population and the coloured population on subtests (CFT ( $M=5.912$ ,  $p=.000$ ), Memory Correct ( $M=3.200$ ,  $p=.000$ ), COLtot ( $M=41.922$ ,  $p=.000$ ) and COLdif ( $M=22.712$ ,  $p=.000$ )). The two population groups that showed no significant bias levels in the test were the black population and the coloured population groups respectively (CFT ( $M=0.118$ ,  $p=1.000$ ), Memory Correct ( $M=0.789$ ,  $p=.923$ ), COLtot ( $M=4.162$ ,  $p=.984$ ) and COLdif ( $M=9.264$ ,  $p=.000$ )). In essence, this illustrates that there is little racial bias between the black population and the coloured population, and that there seems to be racial bias in the white population group performing better at these cognitive ability tests than the black and coloured populations. However, this is a small data sample from which to make such assumptions, and therefore requires analysis of the factor analysis sample. The larger sample illustrates a similar pattern in general racial bias, but in comparison to the performance sample there was found to be greater general

disparity between the white population and black population (CFT ( $M = 6.632$ ,  $p = .005$ ), Memory Correct ( $M = 4.388$ ,  $p = .000$ ), COLtot ( $M = 46.063$ ,  $p = .000$ ) and COLdif ( $M = 24.335$ ,  $p = .000$ )) respectively. Similar differences are found between the white and coloured population groups (CFT ( $M = 4.307$ ,  $p = .000$ ), Memory Correct ( $M = 3.181$ ,  $p = .000$ ), COLtot ( $M = 30.408$ ,  $p = .000$ ) and COLdif ( $M = 18.017$ ,  $p = .000$ )). Unlike the performance sample the population groups that show the least amount of bias between them are the Indian and coloured population groups.

The data from both samples support historical findings (Chung-Yan & Cronshaw, 2002; Jensen, 1980; Roth *et al.*, 2001) that despite awareness of test biasness, psychologists have still not been able to produce a psychometric test that holds no bias. The results from the study further support the findings by Lopes *et al.* (2001) where scores from the coloured and black population groups were consistently lower than the average scores of the sample as a whole, with the white population group scoring consistently higher. This contradicts previous research by Salgado and Anderson (2002), where cultural differences were found to be influential. In addition, research by Makgothu (n.d.) which focuses on the APIL-B specifically, found there to be no levels of racial bias. Kehoe (2002) suggests that researchers should explore different methods for reducing group differences by modifying cognitive tests. Decorte (1999) indicates that the introduction of a weighting system to gain a favourable racial balance in the utility of tests could be used.

The APIL-SV like the APIL-B subtests, is used to create a global score, for which the psychologist can select a disadvantaged adjustment factor depending on the guidelines given for adjustment according to the candidate's history and current status (Aprolab, 2004). However the disadvantaged adjustment factor does not have any influence on the raw data of the subtests.

The final objective of the study was to ascertain the construct validity of the APIL-SV through the use of a factor analysis. The results of the factor analysis concur with the conclusions of Lopes *et al.* (2001) from the inter-correlations. The factor analysis shows that the APIL-SV has construct validity, in that it is measuring an element that it was intended for, which is cognitive ability.

This study investigated the predictive validity of the APIL-SV in a financial/consulting environment in line with South African legislation. Owing to the small size of the PPE sample, conclusions made for the statistical analysis cannot be

generalised to the population within the organisation or for the APIL-SV in general. A concern with the data gathered for the PPE sample is the levels of education given by the candidates, as it is a prerequisite that candidates should have completed their first SAICA QE1, and 18 months of training before writing the PPE. One would not expect a qualified Chartered Accountant to rewrite the PPE, nor would a person with just a Matric be eligible to write the PPE. It is therefore suggested that further studies with a larger sample be conducted on the predictive validity of the APIL-SV on PPE performance. This will allow the researcher to gather the biographical details of the candidates, and be involved in the testing of the candidates.

In order to avoid the limitations discussed above, it is recommended that, if possible, the study should use primary data collected by the researcher. This would allow the researcher to have greater control of the data and the type of data collected.

The current study illustrates the correlations between the subscales and performance even though the levels of correlation are low. However, this issue has been discussed. As Jensen (1998) acknowledges that gender differences exist in a subtest, but not in the overall score, valuable understanding from the APIL-SV could be gained if there was further investigation into the global scores as these are what the client is given on which to base their decisions.

To conclude, the study uncovered racial and gender differences, however did not explore this further. It was found that on an external performance measure the APIL-SV subtests show relatively high levels of correlation in that they predict how the majority of the candidates will perform on the PPE. In this study, which consisted of a data sample of 2 877 subjects, it was conclusively found that one factor accounts for 75% of variance in the performance on the APIL-SV. Owing to the design of the APIL-SV, it is therefore deduced that this one factor that it seeks to assess is the person's cognitive ability.

## Reference

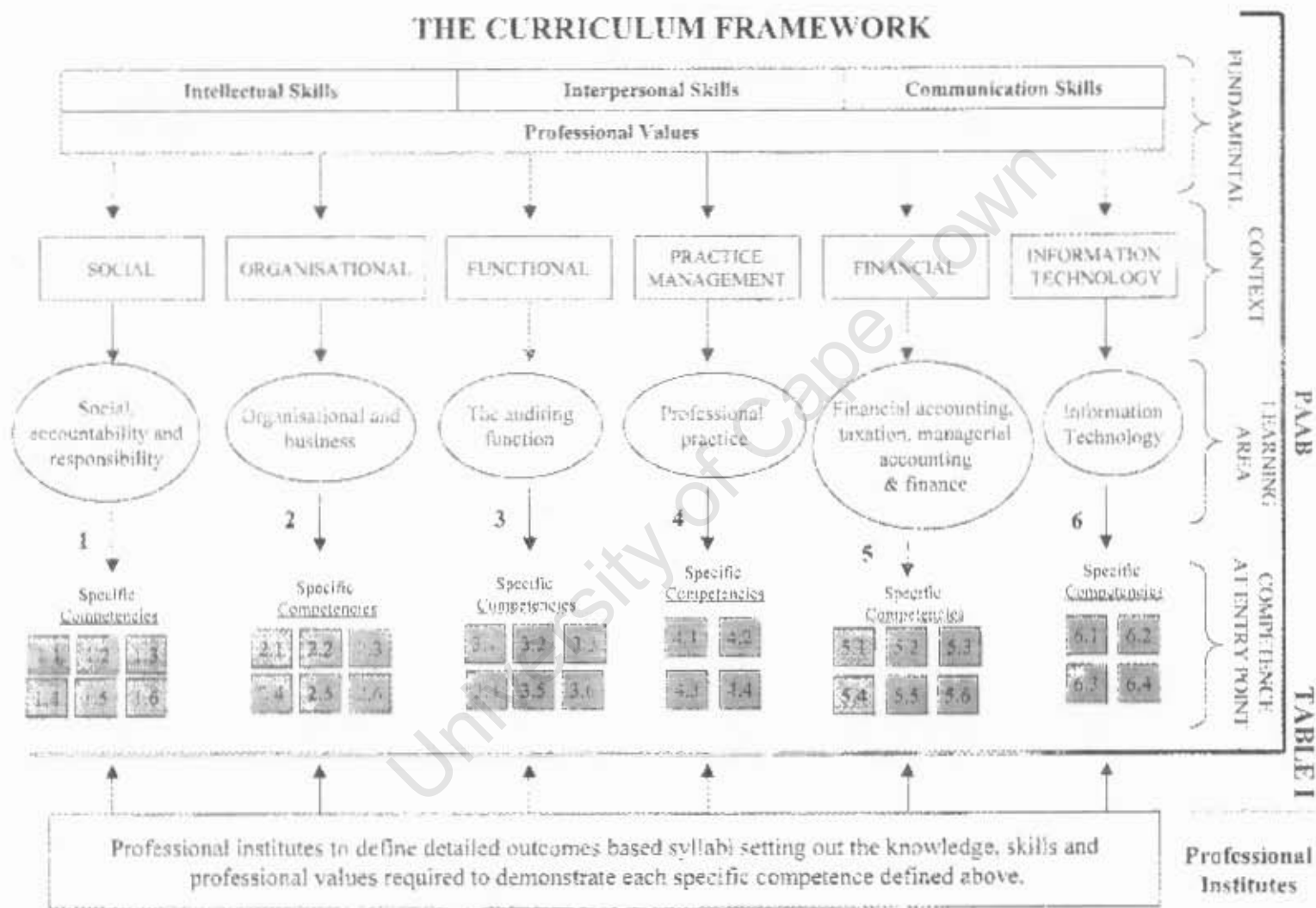
- Aprolab. (2004). *User's manual for APIL Short Version (SV) Battery*. Highlands North: Aprolab.
- Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, *46*, 1012–1024.
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, *10*, 99–109.
- Bertua, C., Anderson, N., & Salgado, J.F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organisational Psychology*, *78*, 387–409.
- Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurements (6th ed.)*. New York: McGraw-Hill.
- Coward, W. M., & Seakett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology*, *75*, 297–300.
- Decorte, W. (1990). Weighing job performance predictors to both maximize the quality of the selected workforce and control the level of adverse impact. *Journal of Applied Psychology*, *84*, 695–702.
- Dickens, W. T. (to be published). Cognitive ability. In S. Durlauf (Ed.), *The new Palgrave dictionary of economics*.
- Employment Equity Act (1998). No. 55. Republic of South Africa.
- Foxcroft, C., Roodt, G., & Abraham, F. (2003). Psychological assessment: A brief retrospective overview. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment* (pp. 11–33). Oxford: Oxford University Press.
- Ghiselli, E. F. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, *26*, 461–477.
- Goldstein, H. W., Zedeck, S., & Goldstein, J. L. (2002). g: Is this your final answer? *Human Performance*, *15*, 123–142.
- Gottfredson, I. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.
- Gottfredson, I. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, *15*, 25–46.

- Hiatt, D. & Hargrave, G.E. (1988) Predicting job performance problems with psychological screening. *Journal of Police Science and Administration*, 16, 122–125.
- Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology*, 55, 353–396.
- Hogan, R. T., De-Fruyt, F., & Rolland, J. P. (2006). The validity of personality measures for personnel selection: Using personality psychology to solve business problems. *Psychologie-Francaise*, 51, 245–264.
- Hunter, J. E., & Hirsh, R. F. (1987). Application of meta-analysis. In C. L. Cooper & I. T. Robertson (Eds.), *Review of industrial psychology* (Vol. 2, pp. 321–357). New York: Wiley.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.
- Huysamen, G.K. (2002). The relevance of the new APA standards for educational and psychological testing in South Africa. *South African Journal of Psychology*, 32(2), 26–33.
- Independent Regulatory Board for Auditors. (2007). *Independent Regulatory Board for Auditors* retrieved November 15, 2007, from <http://www.irba.co.za/home.asp>
- Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.
- Jensen, A. R. (1998) *The g factor: The science of mental ability*. Westport Conn.: Praeger.
- Kehoe, J. F. (2002). General mental ability and selection in private sector organisations: A commentary. *Human Performance*, 15, 97–106.
- Lopes, A., Roodt, G., & Mauer, P. (2001). The predictive validity of the APIL-B in a financial institution. *Journal of Industrial Psychology*, 27, 61–69.
- Makgoatha, J. (no date) *The predictive validity APIL-B: Financial/Consulting Environment*. Unpublished thesis. South Africa.
- Morris, R. W., & Levinson, E. M. (1995). Relationship between intelligence and occupations adjustment and functioning: A literature review. *Journal of Counselling and Development*, 73, 503–514.
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance*, 15, 173–186.

- Murphy, K.R., & Davidshofer, C.O. (1991). *Psychological Testing: Principles and Applications*. Prentice-Hall: New Jersey
- Murphy, K. R., & Shirella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology*, *51*, 823–854.
- Ottz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance*, *15*, 161–171.
- Ree, M.J., & Cattetta, T.R. (2002). g2f. *Human Performance*, *15*, 3–23.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S. III, & Tyler, P. (2001). Ethnic differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, *54*, 297–330.
- Rothstein, M.G., & Geffin, R.D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, *16*, 155–180
- Rudner, L. M. (1992). Pre-employment testing and employee productivity. *Public Personnel Management*, *21*, 133–150.
- Salgado, J. F., & Anderson, N. (2007). Cognitive and gma testing in the European Community: Issues and evidence. *Human Performance*, *15*, 75–96.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187–210.
- Schmidt, F. L. & Hunter, J. F. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.
- Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, g, and work psychology. *Human Performance*, *15*, 142–160
- Taylor, T. R. (1995). *APII*. Auckland Park: Aprolab.
- Twigge, L., Theron, C., Steel, H., & Mairing, D. (2005). A psychometric investigation into the use of an adaptation of the ghiselli predictability index in personnel selection. *SA Journal of Industrial Psychology*, *31*, 18–30.

University of Cape Town

## THE CURRICULUM FRAMEWORK



Appendix