Mode of administration and the stability of the OPQ32n -

Comparing Internet (controlled) and Paper-and-pencil (supervised) administration.

by

Gustav Holtzhausen

submitted for the partial fulfilment of the requirement for the degree of

Master of Commerce

in the subject

Industrial Psychology

at the

University of Pretoria

Supervisor: Prof. Johan Basson

Table of Contents

CHAPTER 1:

BACKGROUND AND PURPOSE

Over the last decade the Internet has opened up a whole new way of communicating, finding or research information and collecting data. It has also opened up a whole new set of opportunities for advancing the science of psychometrics and the technology of testing (Bartram, 2001). For example, administering tests over the Internet allows for greater efficiency and lower costs (i.e. no printing costs), can reduce the possibility of missing data (i.e. applicants did not complete the full questionnaire), allows for applicants to be tested from diverse locations and at different times, and has the potential for immediate scoring (Bartram, 2001; Harris, 1999).

As a result, more and more test publishers are now turning to the Internet to deliver questionnaires and in particular, personality questionnaires, to clients and test users and those who are considering or using the Internet as delivery platform, are challenged with ethical, fairness and best practice issues.

Furthermore assessing people over the Internet is a fairly new concept in South Africa and little information and research exists about its feasibility and fairness and for this reason many psychologists and practitioners are reluctant to use the Internet as medium for test administration.

The Occupational Personality Questionnaire (OPQ32), in particular, is one of the most commonly used occupational personality questionnaires in the world today. It was designed to give information on individual styles or preferences at work and for use in a whole range of assessment and development applications. Its paper-and-pencil and PC-based formats have proven its reliability and validity beyond doubt and many organisations locally and internationally use the OPQ with confidence. The OPQ32 is now available for use via the Internet and similar

questions could be raised about its fairness and stability when candidates complete the questionnaire via the Internet.

As with other questionnaires on the Internet, the OPQ32 can be administered via many alternative modes of delivery, which ranges from open uncontrolled administration to more stringent controlled supervised administration. Only after the publication of encouraging research will administration via the Internet be considered a viable alternative to paper-and-pencil and PC-based administration. Furthermore, the effect of each mode of administration on the stability of the questionnaire needs to be studied in isolation and positive results of one mode of administration should not necessarily be generalised to the other.

Therefore, it is becoming increasingly important to study the effects of the Internet and the mode of administration on the fairness and stability (properties) of occupational assessments when being administered via the Internet.

As a result, this study aims to investigate if it is fair to use the Occupational Personality Questionnaire (OPQ32n) via the Internet by studying its stability when administered in a controlled unsupervised Internet environment.

To achieve this objective it is considered necessary to study the Internet-based administration of the OPQ32 from six different perspectives:

2.1 Fairness

The concept of fairness in psychological testing and personnel decisionmaking is nothing new in South Africa. Mauer (2000) indicates that various sets of legislation attempts to ensure that personnel decisions derived from psychological assessments are fair and controlled.

Mauer (2000) suggests that we consider, the Constitution of the RSA, the Labour Relations Act, The Employment Equity Act and the Health Professions Act when considering fairness.

But not only is fairness a legal issue it is also a social issue where group differences and equitable treatment of all test takers should be considered and a psychometric issue, where psychological assessments or the decisions derived from them cannot be considered fair if we cannot rely on the information produced by them. Specific reference should be made to reliability and validity of psychological tests.

Fairness should also incorporate procedural aspects of the psychological assessment process. Zieky (2002) states that the fairness concept is too complex that no single statistical method can prove that tests are fair but the best way to ensure test fairness is to build fairness into the development, administration, and scoring processes.

Therefore, tests administered over the Internet would considered to be unfair if the legal, social, psychometric and procedural aspects are neglected.

2.2 Personality as determinant of job success

For many years the importance of general ability as a predictor of job success has been accepted and is no longer disputed. This, however, is not the case for personality questionnaires. Only after the landmark paper of Barrick and Mount (1991) this view has changed.

As result, a large body of evidence, including a number of meta-analytical studies, have indicated that measurement of personality should also be considered to supplement predictions of job performance. This view is supported

by a number of world leaders in occupational assessment and a number of occupational personality questionnaires are now being used, not only to make selection decisions, but also to aid personnel decisions in general.

2.3 The OPQ32 as a valid and reliable personality questionnaire

The OPQ series of personality questionnaires were designed to give information on individual styles or preferences at work and for use in a whole range of assessment and development applications.

The OPQ32 specifically, is an occupational model of personality, which describes 32 dimensions or scales of people's preferred or typical style of behaviour at work. It is used by many organisations locally and internationally. Its sound development, reliability and validity should be beyond doubt.

2.4 Media and modes of administration of personality questionnaires

The media by which personality questionnaires can be administered to participants varies. Some examples are:

- Paper and Pencil;
- Palm-top computers;
- Personal Computers; or
- Over the Internet.

The increase in computerised and Internet testing has led to research into the equivalence of different media of administration. Although these studies have delivered mixed results, in most cases results were positive (Clarke, 2000; Donovan, Drasgow & Probst, 2000; Neuman & Baydoun, 1998). In fact Lievens and Harris (2003) note, "initial evidence seems to indicate that measurement equivalence between web-based and paper-and-pencil tests is generally

established". As Bartram (2002) points out "while the general consensus for self-report inventories is that the medium of administration does not have an impact, it is not clear in many cases whether the field studies reported in the literature also confound differences in supervision".

Little is therefore known about how the lack of supervision (mode of administration) will impact on the equivalence of the questionnaire.

2.5 The impact of the Internet on assessment practices

Extensive body of evidence in the literature highlights the advantages and vast possibilities that the Internet offers to test publishers. However, these advantages are worthless if ethical and best practice considerations are ignored.
In particular:

- Performance – questionnaires should not fail half-way or be unable to cope with the demand of test takers;
- Security – only the correct people should have access;
- Privacy – test results should remain confidential at all times;
- Fairness – do all potential test takers have equal access to the Internet;

2.6 Measures of stability

Stability can be measured in many different ways. This study will incorporate three different methods to investigate stability:

- The test reliability (Cronbach Alpha) in supervised paper-and-pencil and controlled Internet-based administration;
- The second analysis compares the effect sizes (*d* statistics) of the two groups;
- Finally the effect of controlled Internet-based administration on scale intercorrelations is examined using Structural Equation Modelling with EQS.

In the event that the results of the stability measures of the OPQ32 remains within the internationally accepted norms, it will be considered fair to use the OPQ32 via the Internet in a controlled unsupervised way.

CHAPTER 2:
THEORETICAL BACKGROUND AND REVIEW OF CURRENT LITERATURE

In Chapter 1 we indicated that the aim of this study is to investigate if it is fair to use the Occupational Personality Questionnaire (OPQ32n) via the Internet by studying its stability when administered in a controlled unsupervised Internet environment. We also indicated that in order for us to achieve this objective, we need to consider it from six different perspectives, namely:

- Fairness in personnel decisionmaking and psychological assessment;
- Personality and determinant of job success;
- The Occupational Personality Questionnaire (OPQ32n);
- Media and modes of administration of personality questionnaires;
- The impact of the Internet on assessment practices; and
- Measures of stability.

Therefore, in this chapter, each perspective will be studied in more detail by reviewing the current literature on the topic.

2.1 Fairness in personnel decisionmaking and psychological assessment

Fairness is not a new concept in South Africa and for many years have the fairness of psychological assessment been under the spotlight due to the perceived disparate treatment it caused certain groups of test takers. However, many psychologists and test publishers have proven beyond doubt that occupational assessment can fairly and effectively be applied in the workplace if the important aspects of fairness have been sufficiently covered. In reviewing the literature, four main components of fairness are consistently highlighted. As a

result, the use of psychological assessments should be benchmarked against each component to evaluate its compliance. The components are:

2.1.1 Fairness legislation

Mauer (2000) indicates that various sets of legislation attempts to ensure that personnel decisions derived from psychological assessments are fair and controlled:

- **The Constitution of the Republic of South Africa (Act 108 of 1996)**
  Mauer (2000) stated that the Constitution is considered to be the *"supreme law of the Republic"* [2], thus *"law or conduct inconsistent with it is invalid, and the obligations imposed by it must be fulfilled"* (section 2). The Bill of Rights in chapter 2 of the Constitution is considered to be *"the cornerstone of democracy in South Africa. It enshrines the rights of all people in our country and affirms the democratic values of human dignity, equality and freedom"*.

  Mauer (2000) appropriately continues by focusing our attention to Section 9 of the Constitution, which deals with equality. *"Everyone is equal before the law…"* (subsection 1) and *"No person may unfairly discriminate directly or indirectly against anyone…"* (subsection 2).

  Our right to privacy is also controlled by the Constitution. Section 14 indicates *"Everyone has the right to privacy, which includes the right not to have –*
  (a) their person or home searched;
  (b) their property searched;
  (c) their possessions seized; or
  (d) the privacy of their communications infringed"
- **The Labour Relations Act (66 of 1995)**
  The LRA, as it is commonly referred to, is deemed to be an instrument in terms of which labour stability is pursued, the rights of organised labour are recognised and where the legislative framework to safeguard workers from

possible exploitation by employers is developed (Mauer, 2000). It should be read in combination with the Constitution of the country and the purpose of the act is clearly stated as *"…to advance the economic development, social justice, labour peace and the democratisation of the workplace…"*

In addition, the LRA makes specific reference in Schedule 7 to *Unfair Labour Practices.* All of which are very relevant in the field of psychological assessment and personnel decisionmaking. Unfair Labour Practices are considered to be:

(a) Unfair discrimination of an employee or applicant employee;
(b) Unfair conduct of an employer relating to promotion, demotion or training;
(c) The unfair suspension of an employee; or
(d) The refusal to re-employ a former employee in terms of any agreement.

? **The Employment Equity Act (55 of 1998)**

The purpose of the Act is to achieve equity in the workplace by:
(a) Promoting equal opportunity through the elimination of unfair discrimination, and
(b) Implementing affirmative action measures to redress the disadvantages in employment experienced by designated groups.

In addition, Section 6(2) of the Act determines that *"it is not unfair discrimination to –*
(a) take affirmative actions measures consistent with the purpose of this Act; or
(b) …exclude or prefer any person on the basis of an inherent requirement of the job"
Employers who uses processes or procedures of discrimination other that these stated above could be challenged. This could include psychological assessments and based on Section 8 of the Act *"Psychological testing and*

other similar assessments of an employee are prohibited unless the test or assessment being used –

(a) has been scientifically shown to be valid and reliable;

(b) can be applied fairly to all employees;

(c) is not biased against any employee or group.

- **The Health Professions Act (56 of 1974)**

  Mauer (2000) states that in terms of the Health Professions Act, a Professional Board for Psychology has been established, and this Board has subsequently established a Psychometrics committee. This Committee recommended that it would constitute unprofessional conduct on the part of a registered psychologist to use tests that had not been classified by the Committee.

  One of the main criteria for classification is the availability of prudent and systematic research evidence that the instrument being used is unbiased, valid and reliable so that the producer, supplier and user has sound grounds to defend deductions made from it.

2.1.2 Fairness as a social responsibility

Not only is fairness a legal issue it is also a social issue and very difficult to site one single definition for a concept so wide, therefore SIOP (2003) views fairness in four possible ways:

- **Group differences** should not be as a result of test bias (systematic error that differentially affects the performance of different groups). As Thorndike (1971) purposefully stated *"The presence (or absence) of differences in means scores between groups…tells us nothing directly about fairness."* The average height of men differs from the average height of women. This does not prove that the measurement yardstick is biased;
- **Equitable treatment** of all test takers. There is also consensus on a need for equitable treatment in test administration;

- Test takers have an **equal opportunity to learn** the subject matter; and
- The **lack of predictive bias** where a common regression line can be used to describe the predictor-criterion relationship for all subgroups.

2.1.3 Psychometrics and fairness

Psychological assessments or the decisions derived from them **cannot** be considered fair if we cannot trust the information produced by them. Specific reference should be made to reliability and validity of psychological tests.

- **Reliability** is concerned with the degree of consistency or agreement between two sets of independently derived scores (Cascio, 1998) In practice, reliability may serve one or both of two purposes:
  (1) to describe the precision of a particular procedure as a measuring instrument, or
  (2) to evaluate the consistency of performance on the procedure by the examinees.
  In this study we will be particularly interested in the evaluation of the consistency or stability of performance on the procedure by the examinees or test takers.
- Cascio (1998) claims the methods of **validation** refers to the degree to which the evidence supports inferences that are made from the scores and revolve around two issues:
  (1) *what* a test or other procedure measures (i.e., the hypothesized underlying trait or construct), and
  (2) *how well* it measures (i.e., the relationship between scores from the procedure and some external criterion measure).

  Although the validity of the OPQ32n is not under review it is fair to say that the validity will be negatively influenced if we discover the instrument to be unstable or inconsistent under different modes of administration.

2.1.4 The procedural aspects of fairness

Lastly fairness should incorporate procedural aspects of the psychological assessment process. Zieky (2002) states that the fairness concept is too complex that no single statistical method can prove that tests are fair but the best way to ensure test fairness is to build fairness into the development, *administration,* scoring, feedback and decision-making processes.

The Association for Assessment in Counselling (2003) also sites procedural issues as critical in ensuring fairness. Procedural consideration includes, but is not limited to, job analysis to determine critical success factors, training and qualification requirements of test users, consideration of test properties, test administration, test scoring and interpretation, and so on. Both test properties (stability) and test administration (modes of administration) is the focus of this study.

To conclude, fairness is a universal concept, which includes legal, social, psychometric and procedural issues and it would be considered unfair to neglect or even ignore any one of these. We could however not study all aspect simultaneously, since this would be nearly impossible. For this reason, this study aims to study at least two very important components of fairness, namely test stability (psychometric) and test administration mode and medium (procedural). In the event that the mode or medium of test administration impacts on test stability, we could conclude that we would treat test takers unfairly if they were to take the OPQ32n over the Internet.

2.2 Personality as a determinant of job success

Many different assessment methods are used to determine job success and to aid in improving personnel decisions. But, many test users are unsure of which assessments to use and if personality questionnaires are in fact useful within the broader personnel decision-making context.

For many years the importance of general ability as a predictor of job success has been accepted and is no longer disputed. This, however, is not the case for personality questionnaires. The need for personality measures in personnel selection has been driven by industry and Bartram (2004) states that in many cases has research followed practice. Thus, many personnel practitioners have included personality measures as part of their assessment procedures.

Only after the landmark paper of Barrick and Mount (1991) this view has changed. As result, a large body of evidence, including a number of meta-analytical studies, have indicated that measurement of personality should also be considered to supplement predictions of job performance. Research has also indicated that the validity of both general ability and personality is now generalisable (Barrick & Mount, 1991; Digman, 1990; Matthews, 1997). There is now very extensive literature on this topic and the validity of personality attributes for predicting job performance is well supported.

Furthermore, the use of ability test tend to result in large group differences, with blacks scoring from 0.5 to 1.0 SD lower on average than whites (Roth, Bevier, Bobko, Switzer & Tyler, 2001) compared to personality measurements which tend to have small differences between black and white subgroups with effect sizes in general smaller than 0.2 (Bartram, Brown, Kriek & Joubert, 2003). Therefore personality measures have the advantage of reducing adverse impact (Bartram, 2004).

It is now confirmed that personality measures tend to be independent of measures of ability (Bartram, 2004) and can add to the overall validity of any selection battery and the Occupational Personality Questionnaire (OPQ32n) is no different.

2.3 The Occupational Personality Questionnaire (OPQ32n)

The following overview describes the OPQ32 and its development in more detail as described in the OPQ32 Manual and User's Guide (1999):

The OPQ series of personality questionnaires were designed to give information on individual styles or preferences at work and for use in a whole range of assessment and development applications.

The OPQ32 specifically, is an occupational model of personality, which describes 32 dimensions or scales of people's preferred or typical style of behaviour at work. It breaks personality down into three domains, namely:

- Relationships with people;
- Thinking style;
- Feeling and emotions;

and joined by a potential fourth

- Dynamism.

In table A, detail on the scale descriptions are provided. For each scale, the extreme low and high scores are defined and sub-clustering is also indicated.

The OPQ32 model of personality is measured by two questionnaires: OPQ32n (normative) and OPQ32i (ipsative) and for the purposes of this study, the normative version was used.

The OPQ32 questionnaires is based on appropriate work-related and are designed for use by Human Resource, training, management development and personnel specialists as well as psychologists. The OPQ32 is particularly appropriate for use with professional and managerial groups, although the content of the scale deals with personality characteristics important to a wide variety of roles: (OPQ32 Manual and User's Guide, 1999).

The development of the OPQ was guided by five criteria:

- Designed specifically for the world of work;
- Avoids clinical or obscure psychological constructs;
- Comprehensive in terms of personality scale measures;
- For use by HR professionals as well as psychologists;
- Based on sound psychometric principles.

The OPQ32 is applicable to a number of different HR challenges:

- Selection;
- Development;
- Succession Planning;
- Career Guidance;
- Restructuring;
- Counselling.

The OPQ32n is a normative questionnaire by which respondents rate themselves on a 1 to 5 scale, ranging from Strongly Disagree (1) through to Strongly Agree (5). The questionnaire consists of 230 statements and the majority of candidates complete the questionnaire in 35-45 minutes.

TABLE A - OPQ32 SCALE DESCRIPTIONS (© SHL Group, plc)

RELATIONSHIPS WITH PEOPLE			
Rarely pressures others to change their views, dislikes selling, less comfortable using negotiating	Persuasive	Enjoys selling, comfortable using negotiation, likes to change other's views	Influence
Happy to let others take charge, dislikes telling people what to do, unlikely to take the lead	Controlling	Likes to be in charge, takes the lead, tells others what to do, takes control	
Hold back from criticising others, may not express own view, unprepared to put forward own opinion	Outspoken	Freely expresses opinions, makes disagreement clear, prepared to criticise others	
Accepts majority decisions, prepared to follow the consensus	Independent Minded	Prefers to follow own approach, prepared to disregard majority decisions	
Quiet and reserved in groups, dislikes being the centre of attention	Outgoing	Lively and animated in groups, talkative, enjoys attention	Sociability
Comfortable spending time away from people, values time spent alone, seldom misses the company of others	Affiliative	Enjoys other's company, likes to be around people, can miss the company of others	
Feels more comfortable in less formal situations, can feel awkward when firs meeting people	Socially Confident	Feels comfortable when firs meeting people	
Makes strengths and achievements known, talks about personal success	Modest	Dislikes discussion achievements, keeps quiet about personal success	Empathy

Prepared to make decisions without consultation, prefers to make decisions alone	Democratic	Consults widely, involves others in decision-making, less likely to make decisions alone	
Selective with sympathy and support, remains detached form others' personal problems	Caring	Sympathetic and considerate towards others, helpful and supportive, gets involved in other's problems	

THINKING STYLE			
Prefers dealing with opinions and feeling rather than facts and figures, likely to avoid using statistics	Data Rational	Likes working with numbers, enjoys analysing statistical information, bases decisions on facts and figures	Analysis
Does not focus on potential limitations, dislikes critically analysing information, rarely looks for errors of mistakes	Evaluative	Critically evaluates information, looks for potential limitations, focuses on upon errors	
Does not question the reason for people's behaviour, tends not to analyse people	Behavioural	Tries to understand motives and behaviour, enjoys analysing people	
Favours changes to work methods, prefers new approaches, less conventional	Conventional	Prefers well established methods, favours a more conventional approach	Creativity and

Prefers to deal with practical rather than theoretical issues, dislikes dealing with abstract concepts	Conceptual	Interested in theories, enjoys discussing abstract concepts	
More likely to build on than generate ideas, less inclined to be creative and inventive	Innovative	Generates new ideas, enjoys being creative, thinks of original solutions	
Prefers routine, is prepared to do repetitive work, does not seek variety	Variety Seeking	Prefers variety, tries out new things, likes changes to regular routine, can become bored by repetitive work	
Behaves consistently across situations, unlikely to behave differently with different people	Adaptable	Changes behaviour to suit the situation, adapts approach to different people	
More likely to focus upon immediate than long-term issues, less likely to take a strategic perspective	Forward Thinking	Takes a long-term view, sets goals for the future, more likely to take a strategic perspective	Structure
Unlikely to become preoccupied with detail, less organised and systematic, dislikes tasks involving detail	Detail Conscious	Focuses on detail, likes to be methodical, organised and systematic, may become preoccupied with detail	
Sees deadline as flexible, prepared to leave some tasks unfinished	Conscientious	Focuses on getting things finished, persists until the job is done	

Not restricted by rules and procedures, prepared to break rules, tends to dislike bureaucracy	Rule Following	Follows rules and regulations, prefers clear guidelines, finds it difficult to break rules	

FEELINGS AND EMOTIONS			
Tends to feel tense, finds it difficult to relax, can find it hard to unwind after work	Relaxed	Finds it easy to relax, rarely feels tense, generally calm and untroubled	Emotion
Feels calm before important occasions, less affected by key events, free from worry	Worrying	Feels nervous before important occasions, worries about things going wrong	
Sensitive, easily hurt by criticism, upset by unfair comments or insults	Tough Minded	Not easily offended, can ignore insults, may be insensitive to personal criticism	
Concerned about the future, expects things to go wrong, focuses on negative aspects of a situation	Optimistic	Expects thing will turn out well, looks to the positive aspects of a situation, has an optimistic view of the future	
Wary of other's intentions, finds it difficult to trust others, unlikely to be fooled by people	Trusting	Trusts people, sees others as reliable and honest, believes what others say	
Openly expresses feelings, finds it difficult to conceal feelings, displays emotion clearly	Emotionally Controlled	Can conceal feelings from others, rarely displays emotion	

Likes to take things at a steady pace, dislikes excessive work demands	Vigorous	Thrives on activity, likes to keep busy, enjoys having a lot to do	Dynamism
Dislikes competing with others, feels that taking part is more important than winning	Competitive	Has a need to win, enjoys competitive activities, dislikes losing	
Sees career progression as less important, looks for achievable rather than highly ambitious targets	Achieving	Ambitious and career-centred, likes to work to demanding goals and targets	
Tends to be cautious when making decisions, likes to take time to reach conclusions	Decisive	Makes fast decisions, reaches conclusions quickly, less cautious	

Each of the 32 scales of the OPQ32n has an actual score that is reflected on a sten scale and can be interpreted as single dimensions, or scales can be linked within one section or sub-section, or even across sections. Individual or linked scales indicate a respondent's preferred personality style in work context, not abilities. A specific personality style is not in itself good or bad, but appropriate or inappropriate depending on circumstances. On some scales, a higher score is indicative of a more positive outcome of the scale, whilst with other scales, a lower score indicates a more favourable description within work context parameters.

Various international and local norm groups exist for interpretation

2.3.1 Reliability and validity

2.3.1.1 Reliability

Reliability sets the upper limit for validity. Reliable questionnaires ensure fairness. Therefore, an important aim of the OPQ32 development programme was to ensure high reliabilities for scales.

The **internal consistencies** of two groups are reported in the OPQ32 Manual and User's Guide. In both cases are high levels of internal consistency achieved:

- the OPQ32n scales for a trailing group in the UK (N = 1228) range between 0,90 and 0,70 with a median of 0,84.
- For a general population group in the UK (N = 2028) it ranged from 0,87 to 0,65 with a median of 0,79.

Various studies in South Africa have been conducted which reports similar results. One such study, Study nr. R038, (2002), reported internal consistencies ranging between 0.69 and 0.88 with only one scale below 0.70 (Kriek & Joubert, 2002). Twenty scales' alpha coefficients exceeded 0.80 and 11 ranged between 0.70 and 0.79.

Test-retest reliability was also established. The reliabilities range after one month from 0,64 to 0,91 with a median of 0,79. This shows that the OPQ32n scores can be expected to remain stable over this period.

2.3.1.2 Validity

As with most properly designed questionnaires it is important to study the various forms of validity.

In terms of **construct validity**, it was found that the intercorrelations for the OPQ32n ranged from -0,51 to 0,56. This suggests generally a high degree of independence for the scales, despite the large number of relatively narrow scales included. Of the 32 scales in the OPQ32n, 77% of scale pairs share less than 10% common variance. However, there are a few pairs of scales that are highly related.

To further prove **construct validity**, factor analysis was also performed. Five factors were extracted explaining 48% of the total variance in the data set. In interpreting the factors, it is useful to consider the 'Big 5' model of personality of McCrae and Costa (McCrae & Costa, 1987) and the six factor SHL Images questionnaire.

Internationally, many studies have been conducted to verify the criterion validity of the OPQ32. For example, the relationship of the OPQ32 and the following studies were examined:

- The OPQ Concept Model;
- OPQ Factor Model;
- The SHL Motivation Questionnaire;
- 16PF;
- NEO PI-R;
- MBTI;
- Occupational Personality Profile (OPP);
- Peer Ratings and self-reports.

The clear patterns of relationships found in all the studies provide strong support for the validity of the OPQ32 questionnaire.

To conduct **predictive and concurrent validity** studies are notoriously difficult due to criterion contamination. Even under these difficult

circumstances, SHL has conducted a number of predictive and concurrent validity studies proving the OPQ can effectively be used to predict job success. In four studies conducted in the UK, hypothesised OPQ32 scales were compared to actual OPQ32 scales predicting job performance. The detail of the significant relationships is tabled in the OPQ32 Manual and User's Guide. Overall the studies clearly indicated that the OPQ makes a strong contribution to predicting job performance (OPQ32 Manual and User's Guide, 1999).

To support this evidence in SA, one concurrent validity study conducted in South Africa the effectiveness of the SA Wechsler Adult Intelligence Scale (SA WAIS), the Career Path Appreciation (CPA), the OPQ32, the Minnesota Multiphasic Personality Inventory – 2 (MMPI – 2) and the Sixteen Personality Factor Questionnaire (16PFi) in selection of Supervisors in the food production industry was assessed (Kriek and Joubert, 2001). It was concluded that the SA WAIS and OPQ32 make the strongest contribution to the hierarchical regression even after various alternative combinations were attempted.

2.4 Media and modes of administration of personality questionnaires

The media by which personality questionnaires can be administered to participants varies. Some examples are:

- Paper and Pencil;
- Palm-top computers;
- Personal Computers; or
- Over the Internet.

The increase in computerised and Internet testing has led to research into the equivalence of different media of administration. Although these studies have delivered mixed results, in most cases results were positive (Clarke, 2000; Donovan, Drasgow & Probst, 2000; Neuman & Baydoun, 1998). In fact Lievens and Harris (2003) note, “initial evidence seems to indicate that measurement equivalence between web-based and paper-and-pencil tests is generally established”. As Bartram (2002) points out “while the general consensus for self-report inventories is that the medium of administration does not have an impact, it is not clear in many cases whether the field studies reported in the literature also confound differences in supervision”.

Little is therefore known about how the lack of supervision will impact on the equivalence of the questionnaire. When administering tests to test takers regardless of the medium of delivery, four modes of test administration can be defined (Bartram, 2001).

- **Open Mode.** These are conditions where there is no means of identifying the test taker and there is no human supervision. Examples of this include books of tests available from bookshops, tests that can be accessed openly on the Internet without any requirement for test taker registration.

- **Controlled Mode.** This is similar to the Open Mode in that no human supervision of the test session is assumed. However, the test is only made available to known test takers. For the Internet this is controlled through the requirement for the test taker to be provided with a logon username and password.

- **Supervised Mode.** For this mode, a level of human supervision is assumed, whereby the identity of the test taker can be authenticated and test-taking conditions validated. This Mode also provides a better level of control over

dealing with unexpected problems or issues. For Internet testing, this mode is achieved by requiring the test administrator to login the candidate and to confirm that the testing was completed correctly at the end of the session.

- **Managed Mode.** This is a mode where a high level of human supervision is assumed and there is also control over the test-taking environment. For computer-based testing this is achieved through the use of dedicated testing centres. The organisation managing the testing process can define and assure the performance and specification of equipment in test centres. They can also generally exercise more control over the competence of the staff.

Both the medium (Internet vs. paper-and-pencil) and the mode of administration (controlled versus supervised) are of particular interest to this study, since it potentially affects the fairness of the test process (see section of fairness). Furthermore, Bartram (2001) identifies the six main functions of test administrators, which will directly impact on the supervised mode of administration. Little is currently known what the effect on the test stability will be if these functions are absent. These six functions are:

- **Authenticating** the identity of the test taker (i.e. establishing who is actually taking the test);
- Establishing a **positive rapport** with the test taker (i.e. making sure that an appropriate climate is created for the test taking session and that the test taker is not unduly anxious);
- Ensuring that **instructions** regarding standardised conditions are followed (e.g. making sure that timing conditions are adhered to, that calculators or other aids are used or not as instructed);
- Dealing with **unexpected conditions** or problems that arise prior to or during the administration process (managing problems with equipment, hardware, disruptions during the test session, test taker disabilities etc);

- **Validating the test results** (i.e. ensuring that the results obtained are what they appear to be, and were the product of the authenticated test taker operating unaided);
- Ensuring that **test materials** are kept **secure** (i.e. making sure that no copies of test booklets or items are removed by the test takers).

Although some research is available on the medium of delivery, limited research is available regarding what the impact of the mode of administration, or when a test administrator is not present, will be on the stability of the questionnaire. One initial study conducted on the OPQ32i by Bartram and Brown (2002) concluded that web-based controlled administration appear to have comparable psychometric properties to paper-and-pencil supervised data in terms of reliability and relationships between scales, which is encouraging as it implies there is no distortion to the instrument itself. This study will aim to examine the OPQ32n in this regard.

2.5 The impact of the Internet on assessment practices

The Internet has opened up a whole new set of opportunities for advancing the science of psychometrics and the technology of testing (Bartram, 2001). In addition, the increased use of computers and the interconnectivity of the World Wide Web offer many possibilities for improving the practice of personnel recruitment, selection and development. For example, administering tests over the Internet allows for greater efficiency and lower costs (i.e. no printing costs), can reduce the possibility of missing data (i.e. applicants did not complete the full questionnaire), allows for applicants to be tested from diverse locations and at different times, and has the potential for immediate scoring (Bartram, 2001; Harris, 1999).

The field of recruitment has experienced the largest expansion and strongest development in the online arena. A survey (Reed, 2002) revealed that 78% of graduate recruiters prefer their applicants to apply online rather than on a paper-based form compared to only 44% in 1999 (Park, 1999) and the field of selection and development is following closely behind.

2.5.1 Advantages of Internet assessments

Extensive body of evidence in the literature explains the advantages and possibilities that the Internet offers to test publishers. It is considered relevant to discuss some of the most important advantages, since these will impact on how we view testing in the future.

Since the aim of this study is to investigate if the OPQ32 remains stable if administered over the Internet, these advantages will become worthless if we cannot deliver reliable tests over the Internet with intact psychometric properties.

Some of these advantages are:

- Positive candidate response
  1) Candidates are **generally positive** with the selection process and view the company at which they were applying with, more positively (Wroe, 2000);
  2) In a survey conducted by Ferguson and Glennon (2001) it was ascertained that 74% of test takers **preferred online testing** to paper-and pencil testing. They cited ease of use, access, speed, convenience and ease of submission as some of the reasons;
  3) Job seekers are **willing to spend time applying** for jobs online. In a study conducted, over 60% of job seekers are willing to spend more than 30 minutes or more answering online questions about their skills and capabilities (Handler & Hunt, 2002).

? Access

1) Web-based testing offers global **24/7 access** to testing, immediate scoring and more limited need for test administrators leading to convenient, cost effective and efficient testing (Jones, 1998);
2) Globally **access to the Internet is increasing** rapidly. One study forecasts that by next year there will be over 717,000 million users worldwide (CyberAtlas, 1999). Forecasts for Africa however is not even 4% of this projection;
3) The Internet is one of the **main ways that people look for jobs**. In the USA alone over 52 million people have used the Internet in 2002 to find job related information (Handler & Hunt, 2002);
4) In the same study conducted (Ferguson & Glennon, 2001) **more and more** applicants have the opportunity to **gain Internet access from work** (55%) or academic institutions (38%);
5) The Internet enables test publishers to have the same questionnaire available in **multiple languages** for global use. International organisations can now assess candidates in multiple countries in the appropriate time zone;
6) It is widely accepted that candidates or test takers should have the opportunity to **practice tests**. The Internet provides the opportunity to get Practice Tests to candidates, efficiently and inexpensively.

? HRIS Integration

1) Internet-based assessments is leading to a seamlessly **integrated personnel assessment platform** which link from recruitment to selection, development performance management and succession planning, which typically had to be managed on paper based systems (Jones & Higgins, 2002);
2) **Communication** between employers and candidates are vastly improved with automated email responses and notifications which results in more

effective management of candidates through the selection and assessment process;

3) The Internet increases the **ability to monitor** and approve the effectiveness of staffing methods by linking selection data with other data such as employee performance and tenure;
4) Many large human resources information systems like **SAP and PeopleSoft** are now teaming up with test publishers to ensure that assessment results can be integrated into these HR platforms. The Internet is particularly useful for this.

? Simplified and cost effective maintenance and data collection
1) Internet-based systems are **easier to maintain** (e.g. test norms, test updates and software upgrades are only updated on one server vs. to all test users Jones & Higgins, 2002);
2) **Data collection** for norms and research purposes are drastically simplified with web-based technology due to a single server application. (Bartram, 2001);

? Improved mechanisms
1) Although **Item Response Theory or Adaptive Testing** are not unknown concepts and have been available to test publishers for many years, the Internet now provides for the opportunity to implement adaptive tests with ease due to single server maintenance;
2) It is now possible to truly **automate the assessment process**. For example candidates can be tasked with one assessment first (hurdle) and if successful be tasked to complete another;
3) **Automated scoring** mechanisms of Internet platforms can take the human error element out of scoring.

? Reduce costs

1) Other than the advantages already mentioned which in most cases lead to cost saving, **the need for test administrators** in lower stakes assessments, like development, can be reduced thus making tests more accessible and less costly. The cost of test administrators tends to increase assessment costs dramatically;

? Decentralised decision-making

1) A survey conducted in the UK (IPD, 1999) that more and **more line managers are involved** in the selection process. This trend has continued and is similar in South Africa, which have experienced major restructuring in especially human resources departments. The Internet enables access to job analysis and assessment technology and the reports of these test and questionnaires to be delivered directly to line managers. These reports should however be **interpreted reports** based on inherent job requirements. The OPQ32n has number of interpreted competency-based outputs like the Management Competency Profile, which can be used by line managers directly.

2.5.2 Challenges and the need for fairness, best practice and ethical guidelines

Together with new technology a number of challenges are born. These need to be considered and studied in order to establish the Internet as a reliable and effective alternative recruitment and selection mechanism. It is fair to comment that in the years to come a body of evidence will emerge to study the affect of the Internet on assessment and guidelines for best practice will emerge as the test fraternity is more established in the field of Internet assessment.

Let us study some of these challenges and best practice concerns in more detail:

? Performance

1) The Internet could be **compared with** the early days of the **telephone**. Not everyone had a connection, the system did not always work and connections could be problematic at times. Although still a major concern in South Africa and other developing countries, the Internet is now experiencing increasing levels of reliability and dependability (Bartram, 2001);
2) Testing on the Internet should at least have two characteristics (Bartram, 2001). One, it should provide the means of controlling the **timing** and delivery and two, it should **not fail mid-way** through the test. In the study conducted by Ferguson and Glennon (2001) 75% of candidates were concerned with the Internet connection failing;
3) The Internet system should also be able to **cope with the demand of test takers**. For example, if a large number of test takers require to do the test simultaneously, the system must be robust enough to cope with the demand;
4) Test should be designed in a format that, in the event of a **breakdown in connectivity**, the test should be seamlessly resumed from the point at which the break occurred;
5) **Consistency of appearance** is also a main concern. Global Internet browser packages are not always consistent and test publishers are considering alternative way in which to deliver tests to end-users.

? Security

1) The danger of **security breeches** is high. Test users can pass these materials on to people who are not authorise to use them (Bartram, 2001);
2) Protecting the publisher's **intellectual property**, controlling access and distribution of tests and keeping scores and rules confidential are all high on the agenda of test publishers and psychologists (Bartram, 2001);

3) Although extremely difficult in circumstances where test administrators are not available, the **test taker's identity** must be confirmed. Some alternatives exist namely, user names and passwords and limiting access to specific machines on the Internet only (Bartram 2001). The Ferguson & Glennon study (2001) determined that 50% of other applicants were concerned about applicants acquiring assistance. In fact 13% indicated they would be tempted to obtain assistance if the assessment was for a job application and 7% actually gained help;
4) Tests need to be updated regularly. Due to test being in the public domain, the content of tests can be copied and studied. As said before, Adaptive Testing could be the answer for this problem.

? Privacy

1) Test publishers and test users should study and understand the risks associated with **controlling access to test results and reports**, the legal issues relating to data protection, privacy and storage (Bartram, 2001);
2) In most cases data is kept centrally and test users could be concerned about **data protection**. Test publishers should aim to secure test results as banks would money (Bartram, 2001). In the Ferguson & Glennon study (2001) an alarming 81% of respondents did not strongly agree that the Internet is a confidential way of transferring assessment information. This indicates a serious need for education;
3) As previously discussed, one of the advantages of the Internet is global reach. This however posts the challenge that not all countries share the same **data protection laws** and test publishers should be aware of local laws in this regard.

? Fairness

1) All groups applying for a position should have **equal access** to the Internet and the assessments offered. This is however far from realistic. Most development in infrastructure is currently concentrated in the US, Europe and Asia Pacific and developing countries are still far behind. As a result, candidates could have to bear a costs to use a third party Internet vendor like the popular Internet Café, which could limit access;
2) Although discussed in more detail in section 1.4, the use of **test administrators** and assessment on the Internet will be studied in more detail in the years to come. The main distinguishing factor is the type of assessment that is carried out. With high stakes assessments, like selection or restructuring, preliminary views are that test administrators should be present. With lower stakes assessments, like development, it seems that consensus exists among practitioners that the role of test administrators are less important. In the Ferguson & Glennon study, 89% of applicants indicated they would prefer not to have an administrator present;
3) Test taker **honesty** within the context of web-based assessment is a heavily debated topic. Honesty is not just a matter of supervision but refers to the complete assessment process (Bartram, 2001). One possible solution to honesty could be an honesty policy which candidates sign before taking a web-based test. This honesty policy could include aspect like identification, the fact that the web-based assessment results will be compared to results for later supervised and controlled assessments and so on;
4) Two options exist for **candidate feedback** when referring to web-based assessments. Candidates can receive feedback over the Internet or could receive feedback from a trained professional. Both have its advantages and disadvantages and discussion is outside of the scope of this study.

However, test users should consider the assessment scenario and what design fits the assessment requirements best;

5) As discussed in various sections of this document, **test quality** is of the utmost importance when constructing tests for the Internet. Tests, which prove to be unreliable when administered over the Internet, will prove to be unfair and negatively impact on the perception of web-based assessments. It will become a standard for test publishers to publish the **equivalence results** of their web-based tests (Buchanan, date unknown);
6) There are also a number of other factors that could influence web-based assessments. For example, the **level of computer skill** or level of comfort with computers could have an impact;
7) **The race to create** web-based assessments may result in improperly designed and constructed tests and questionnaires or insufficient job performance dimensions being used to define performance dimensions;
8) **On-line support** for test takers should also be considered. In a study conducted by Price and Patterson (2003) 85% of test takers believed that support was inadequate because it was difficult to obtain contact details, help files were to general or some even did not offer any support. Time zones for support should also be considered.

SHL Group plc, the test publisher of the OPQ32n, delivers its questionnaires via its SHLSolutions Internet platform. It is considered a secure, robust and leading edge Internet platform and in its continuous development provides for most of the fairness and best practice concerns raised about the Internet as test administration platform (Bartram, 2001).

2.5.3 Global Internet assessment trends

With globalisation, international companies are looking for new ways to implement assessment practices on a worldwide scale and adopt common selection and recruitment practices across different countries (Bartram, 2004).

Because of its advantages, many international test publishers turning to the Internet to deliver their instruments and some of these organisations are:

- SHL;
- Pan testing;
- The Psychological Corporation;
- Pearson Reid London House;
- Ramsay;
- Hogan.

In most cases these organisations are converting their paper & pencil and existing computer-based assessments onto the Internet. Not only is the Internet being considered to deliver tests to candidates, it provides a new medium for distribution of test materials, reports, manuals, and for the automated collection of data (Bartram, 2001). Many test publishers see the Internet as a possible source of support for the test user, i.e. user group discussion forums, test updates, norm distribution to only name a few.

The Internet is however opening the field of adaptive testing, which is based on item-response theory. Adaptive differs from any other form of testing in that the item selection and test length vary based on the way the test taker answers the questions (Jones & Higgins, 2002). Adaptive testing is however outside the scope of this study and we can expect a body of research to be published on this subject.

Not only is the Internet being considered or used for recruitment and selection but many more HR applications are considered (Pfenninger, Klion & Wenzel, 2003), like:

- Reference checks;
- Identification of leadership potential;
- Succession planning;
- Job analysis;
- Organisational surveys;
- Outplacement;
- Business simulations;
- Linking recruitment, assessment and development technology with human resources information systems;
- Performance management;

Technological advances within the online world will, for years to come, shape the way we deliver human resources applications to test takers, applicants and incumbents, wireless technology like WAP (Wireless Application Protocol), the open global standard for providing wireless Internet access from handheld devices typically used in mobile phones, to mention only one.

UMTS (universal mobile telephony services) will carry an average of 144,000 bps compared to the current 9,600 bps of current mobile phones. This will ensure faster more efficient access to the Internet and email (Bartram, 2000).

Data convergence in the IT world is enjoying a large part of the development attention. Bartram (2000) comments that technologies are converging to the point where it is difficult to distinguish between PC's, TVs and telephones. Soon we will be accessing the Internet and downloading our emails on our TV sets via our digital decoder. This leaves possibilities for test publishers beyond our imagination.

2.6 Measures of stability

As said before, the advancement to web-based assessment raises a number of challenges. The main focus of this research is test stability, therefore, three forms of stability will be considered. The intention is to mirror the stability measures of the OPQ32i study conducted by Bartram and Brown (2002) to ascertain if the OPQ32n and OPQ32i produce comparable results.

2.6.1 The test reliability (Cronbach Alpha) in supervised paper-and-pencil and controlled Internet-based administration

Reliability is the extent to which a test is repeatable and yields consistent scores. All measurement procedures have the potential for error, so the aim is to minimize it. An observed test score is made up of the true score plus measurement error.

The goal of estimating reliability (consistency) is to determine how much of the variability in test scores is due to measurement error and how much is due to variability in true scores. Measurement errors are essentially random: a person's test score might not reflect the true score because they were sick, anxious, in a noisy room, etc.

Reliability can be improved by:

- Getting **repeated measurements** using the same test; and
- Getting **many different measures** using slightly different techniques and methods.

There are several types of reliability:

- Test-retest reliability
  The test-retest method of estimating a test's reliability involves administering the test to the same group of people at least twice. Then the first set of scores is correlated with the second set of scores. Correlations range between 0 (low reliability) and 1 (high reliability)

- Alternate forms
  Administer Test A to a group and then administer Test B to same group. Correlation between the two scores is the estimate of the test reliability.

- Split-half reliability
  Relationship between half the items and the other half of the items within the same assessment device.

- Internal consistency
  Internal consistency is commonly measured as Cronbach's Alpha (based on inter-item correlations). An Alpha of 0 is considered low and 1 considered high. The greater the number of similar items, the greater the internal consistency. For this reason some tests seem to be repeating the same type question (Item) in different ways. Generally, an Alpha of 0.80 is considered as a reasonable benchmark for measures of ability and between 0.70 and 0.80 for personality questionnaires (Gliem & Gliem, 2003; Field, 2001).

2.6.2 The second analysis compares the effect sizes (*d* statistics) of the two groups

To effectively assess meaningful outcomes, a metric other than statistical significance testing is useful, one that measures the magnitude of a result, rather than the probability that the result is due to chance (Cook, 1999)

One of the limitations of traditional statistical significance testing is that it is highly dependent on sample size (the opportunity for achieving statistically significant results increase as sample size increases and vice versa). Therefore two studies, with similar results but different sample sizes, may yield very different outcomes when statistical significance testing is applied (Cook, 1999).

The most commonly used metric to evaluate the magnitude ofan outcome is "effect size" (Cohen, 1988). In this case *d* statistics is used for measuring the effect size difference between two means.

Effect sizes are generally considered as small (effect size = 0.2), medium (effect size = 0.5) and large (effect size = 0.8) corresponding to correlations of 0.1, 0.3 and 0.5 respectively (Cohen, 1988).

2.6.3 Finally the effect of controlled Internet-based administration on scale intercorrelations is examined using Structural Equation Modelling with EQS

Structural equation modeling (SEM) grows out of and serves purposes similar to multiple regression, but in a more powerful way which takes into account the modeling of interactions, nonlinearities, correlated independents, measurement error, correlated error terms, multiple latent independents each measured by multiple indicators, and one or more latent dependents also each with multiple indicators. SEM may be used as a more powerful alternative to multiple regression, path analysis, factor analysis, time series analysis, and analysis of covariance. That is, these procedures may be seen as special cases of SEM, or, to put it another way, SEM is an extension of the general linear model (GLM) of which multiple regression is a part (Rigdon, 2004).

Advantages of SEM compared to multiple regression include more flexible assumptions, use of confirmatory factor analysis to reduce measurement error by

having multiple indicators per latent variable, the attraction of SEM's graphical modeling interface, the desirability of testing models o verall rather than coefficients individually, the ability to test models with multiple dependents, the ability to model mediating variables, the ability to model error terms, the ability to test coefficients across multiple between-subjects groups, and ability to handle difficult data (Rigdon, 2004).

The SEM is a family of statistical techniques, which incorporates and integrates factor analysis and path analysis, and centres around two steps, namely validating the measurement model and fitting the struc tural model. The former is accomplished primarily through confirmatory factor analysis, while the latter is accomplished primarily through path analysis with latent variables. A measurement model, as its name implies, is about measurement and data collection. A structural model specifies how well some variables could predict some other variables. Because prediction involves relationships, it could be viewed as a regression model. Also, since the relationships form a "chain" or a "path," it is also known as path model. When these two models are combined, they form a structural equation model (Rigdon, 2004).

LISREL, AMOS and EQS are three popular statistical packages for conducting SEM. LISREL popularised SEM in sociology and the social sciences and is still the package of reference in most articles about structural equation modeling. EQS is considered more user friendly as some of the other packages and tests the full range of SEM and is becoming popular as an easier way of specifying structural models (Rigdon, 2004).

A comparison of covariance structures will be carried out using SEM with EQS. The fit will be tested of a model that evaluates each pair of scale intercorrelations and each scale variance, to be equal across the two samples. There are a few statistics that measure how adequately the hypothesized model describes the

sample data (Bartram, 2002). The comparative fit index (CFI) (Bentler, 1990) ranges from zero to 1.00 and provides a measure of complete covariation in the data. Although a value >0.90 was originally considered representative of a well-fitting model, a revised cut-off value close to 0.95 has recently been advised (Hu & Bentler, 1999). Another fit measure is the root mean square error of approximation (RMSEA). This index has recently been recognized as one of the most informative criteria in covariance structure modelling; values less than 0.05 indicate good fit (Byrne, 2001).

2.7 Conclusion

It is apparent from the literature that personality questionnaires are now widely accepted to contribute significantly to personnel decisions. It is also apparent that the OPQ32 is a widely used and accepted personality questionnaire in the world today and that the publisher, SHL Group plc, has provided for most of the fairness and best practice concerns regarding Internet assessment. It is also evident for self-report inventories that the medium of administration does not have an impact on the equivalence of the questionnaire.

However, there is insufficient research to confirm if the mode of delivery has an impact on the stability of Internet-based questionnaires. As a result, this study will investigate the impact of the mode of delivery on the stability of the OPQ32n by comparing the test reliability, the effect sizes and scale intercorrelations (equivalence) of two groups namely supervised paper-and-pencil and controlled Internet-based administration groups.

CHAPTER 3: RESEARCH ARTICLE

MODE OF ADMINISTRATION AND THE STABILITY OF THE OPQ32N:

Comparing Internet (controlled) and paper-and-pencil (supervised) administration.

GC HOLTZHAUSEN
Department of Human Resources Management
University of Pretoria

ABSTRACT

This paper reports on the stability of the OPQ32n when administered in controlled web-based conditions over the Internet. The fairness of Internet assessments in South Africa is a strongly debated and since most test publishers consider web-based testing, it is important that web-based questionnaires remain psychometrically stable. This study compares 322 supervised paper-and-pencil respondents with 322 controlled web-based respondents, who completed the OPQ32n in South Africa. The results indicate that the two samples have comparable psychometric properties in terms of reliability and covariance structures. The results also indicate that that only a small difference exists between the scale means of the two groups. The evidence, therefore, supports the use of the OPQ32n for controlled web-based administration conditions.

OPSOMMING

Hierdie navorsing doen verslag oor die stabiliteit van die OPQ32n wanneer dit of onder gekontroleerde omstandigehede oor die Internet afgeneem word. Die regverdigheid van sielkundige toetsing oor die Internet is huidiglik onder die soeklig en aangesien verskeie toetspubliseerders die Internet oorweeg as 'n alternatief vir papier-en-potlood toetsing, is dit krities om te bepaal of Internet gebasseerde vraelyste stabiel bly. Die studie vergelyk 322 Internet gebasseerde toets afnemers met 322 papier-en-potlood afnemers. Die resultate het bevind dat die twee steekproewe vergelykbare psigometriese eienskappe het in terme van betroubaarheid en ekwivalensie en dat daar slegs klein verskille is tussen die skaal gemiddeldes is van die twee groepe. Dit bleik dus duidelik dat die OPQ32n gemaklik in gekontroleerde Internet gebasseerde omstandighede gebruik kan word.

Over the last decade the Internet has opened up a whole new way of communicating, finding or research information and collecting data. It has opened up a whole new set of opportunities for advancing the science of psychometrics and the technology of testing (Bartram, 2001). For example, administering tests over the Internet allows for greater efficiency and lower costs (i.e. no printing costs), can reduce the possibility of missing data (i.e. applicants did not complete the full questionnaire), allows for applicants to be tested from diverse locations and at different times, and has the potential for immediate scoring (Bartram, 2001; Harris, 1999).

As a result, more and more test publishers are now turning to the Internet to deliver questionnaires to clients and test users and those who are considering or using the Internet as delivery platform, are challenged with ethical, fairness and best practice issues. Furthermore assessing people over the Internet is a fairly new concept in South Africa and little information and research exists about its feasibility and fairness. It therefore becoming increasingly important to study the effects of the Internet on the fairness and properties of occupational assessments.

Fairness in personnel decisionmaking and psychological assessment

The concept of fairness in psychological testing and personnel decisionmaking is nothing new in South Africa. Mauer (2000) indicates that various sets of legislation attempts to ensure that personnel decisions derived from psychological assessments are fair and controlled:

The Constitution of the Republic of South Africa (Act 108 of 1996)

Mauer (2000) stated that the Constitution is considered to be the *"supreme law of the Republic"*, thus *"law or conduct inconsistent with it is invalid, and the obligations imposed by it must be fulfilled"* (section 2). The Bill of Rights in chapter 2 of the Constitution is considered to be *"the cornerstone of democracy in South Africa. It enshrines the rights of all people in our country and affirms the democratic values of human dignity, equality and freedom"*.

Mauer (2000) appropriately continues by focusing our attention to Section 9 of the Constitution, which deals with equality. *"Everyone is equal before the law…"* (subsection 1) and *"No person may unfairly discriminate directly or indirectly against anyone…"* (subsection 2).

The Labour Relations Act (66 of 1995)

The LRA, as it is commonly referred to, is deemed to be an instrument in terms of which labour stability is pursued, the rights of organised labour are recognised and where the legislative framework to safeguard workers from possible exploitation by employers is developed (Mauer, 2000).

It makes specific reference in Schedule 7 to *Unfair Labour Practices.* All of which are very relevant in the field of psychological assessment and personnel decisionmaking. Unfair Labour Practices are considered to be:
(a) Unfair discrimination of an employee or applicant employee;
(b) Unfair conduct of an employer relating to promotion, demotion or training;
(c) The unfair suspension of an employee; or
(d) The refusal to re-employ a former employee in terms of any agreement.

The Employment Equity Act (55 of 1998)

The purpose of the Act is to achieve equity in the workplace by:
(a) Promoting equal opportunity through the elimination of unfair discrimination, and
(b) Implementing affirmative action measures to redress the disadvantages in employment experienced by designated groups.

In addition, Section 6(2) of the Act determines that *"it is not unfair discrimination to –*
(a) take affirmative actions measures consistent with the purpose of this Act; or
(b) …exclude or prefer any person on the basis of an inherent requirement of the job"
Employers who uses processes or procedures of discrimination other that these stated above could be challenged. This could include psychological assessments and based on Section 8 of the Act *"Psychological testing and other similar assessments of an employee are prohibited unless the test or assessment being used –*

(a) has been scientifically shown to be valid and reliable;
(b) can be applied fairly to all employees;
(c) is not biased against any employee or group.

The Health Professions Act (56 of 1974)

Mauer (2000) states that in terms of the Health Professions Act, a Professional Board for Psychology has been established, and this Board has subsequently established a Psychometrics committee. This Committee recommended that it would constitute unprofessional conduct on the part of a registered psychologist to use tests that had not been classified by the Committee.

One of the main criteria for classification is the availability of prudent and systematic research evidence that the instrument being used is unbiased, valid and reliable so that the producer, supplier and user has sound grounds to defend deductions made from it.

Not only is fairness a legal issue it is also a social issue and very difficult to site one single definition for a concept so wide, therefore SIOP (2003) views fairness in four possible ways:

- Group differences should not be as a result of test bias (systematic error that differentially affects the performance of different groups).
- Equitable treatment of all test takers. There is also consensus on a need for equitable treatment in test administration;
- Test takers have an equal opportunity to learn the subject matter; and
- The lack of predictive bias where a common regression line can be used to describe the predictor-criterion relationship.

In addition, psychological assessments or the decisions derived from them *cannot* be considered fair if we cannot trust the information produced by them. Specific reference should be made to reliability and validity of psychological tests.

Lastly fairness should incorporate procedural aspects of the psychological assessment process. Zieky (2002) states that the fairness concept is too complex that no single statistical method can prove that tests are fair but the best way to ensure test fairness is to build fairness into the development, administration, and scoring processes.

The Association for Assessment in Counselling (2003) sites procedural consideration as, but is not limited to, job analysis to determine critical success factors, training and qualification requirements of test users, consideration of test properties, test administration, test scoring and interpretation, and so on.

To conclude, fairness is a universal concept, which includes legal, social, psychometric and procedural issues and it would be considered unfair to neglect or even ignore any one of these. We could however not study all aspect simultaneously, since this would be nearly impossible. For this reason, this study aims to study at least two very important components of fairness, namely test stability (psychometric) and test administration mode and medium (procedural).

In the event that the mode or medium of test administration impacts on test stability, we could conclude that we would treat test takers unfairly if they were to take tests over the Internet.

Personality as a determinant of job success

For many years the importance of general ability as a predictor of job success has been accepted and is no longer disputed. This, however, is not the case for personality questionnaires. Only after the landmark paper of Barrick and Mount (1991) this view has changed.

As result, a large body of evidence, including a number of meta-analytical studies, have indicated that measurement of personality should also be considered to supplement predictions of job performance. Research has also indicated that the validity of both general ability and personality is now generalisable (Barrick & Mount, 1991; Digman,

1990; Matthews, 1997). There is now very extensive literature on this topic and the validity of personality attributes for predicting job performance is well supported. This view is supported by a number of world leaders in occupational assessment and a number of occupational personality questionnaires are now being used not only to make selection decisions but also to aid personnel decisions in general.
In addition, the use of ability test tend to result in large group differences, with black samples scoring from 0.5 to 1.0 SD lower on average than white samples (Roth, Bevier, Bobko, Switzer & Tyler, 2001) compared to personality measurements which tend to have small differences between black and white subgroups with effect sizes in general smaller than 0.2 (Bartram, Brown, Kriek & Joubert, 2003). Therefore personality measures have the advantage of reducing adverse impact (Bartram, 2004).

It is now confirmed that personality measures tend to be independent of measures of ability (Bartram, 2004) and can add to the overall validity of any selection battery and the Occupational Personality Questionnaire (OPQ32n) is no different.

Modes of administration of personality questionnaires

The media by which personality questionnaires can be administered to participants varies. Some examples are:

- Paper and Pencil;
- Palm-top computers;
- Personal Computers; or
- Over the Internet.

Bartram (2002) points out “while the general consensus for self-report inventories is that the medium of administration does not have an impact, it is not clear in many cases whether the field studies reported in the literature also confound differences in supervision”. Little is therefore known about how the lack of supervision will impact on the equivalence of questionnaires.

When administering tests to test takers, regardless of the medium of delivery, four modes of test administration or supervision can be defined (Bartram, 2001).

- **Open Mode.** These are conditions where there is no means of identifying the test taker and there is no human supervision.
- **Controlled Mode.** This is similar to the Open Mode in that no human supervision of the test session is assumed. However, the test is only made available to known test takers.
- **Supervised Mode.** For this mode, a level of human supervision is assumed, whereby the identity of the test taker can be authenticated and test-taking conditions validated.
- **Managed Mode.** This is a mode where a high level of human supervision is assumed and there is also control over the test-taking environment.

Both the medium (Internet vs. paper-and-pencil) and the mode of administration (controlled versus supervised) are of particular interest to this study, since it potentially affects the fairness of the test process. Furthermore, Bartram (2001) identifies the six main functions of test administrators, which will directly impact on the supervised mode of administration. Little is currently known what the effect on the test stability will be if these functions are absent. These six functions are:

- Authenticating the identity of the test taker;
- Establishing a positive rapport with the test taker;
- Ensuring that instructions regarding standardised conditions are followed;
- Dealing with unexpected conditions or problems that arise prior to or during the administration process;
- Validating the test results;
- Ensuring that test materials are kept secure.

One initial study conducted on the OPQ32i by Bartram and Brown (2002) concluded that web-based controlled administration appear to have comparable psychometric properties to paper-and-pencil supervised data in terms of reliability and relationships between scales, which is encouraging as it implies there is no distortion to the instrument itself. This study will aim to examine the OPQ32n in this regard.

The impact of the Internet on assessment practices

The Internet has opened up a whole new set of opportunities for advancing the science of psychometrics and the technology of testing (Bartram, 2001) and offer many possibilities for improving the practice of personnel recruitment, selection and development. Extensive bodies of evidence in the literature explain the advantages and possibilities that the Internet offers to test publishers.

Some of these advantages are:

- Positive response – In general candidate responses are positive towards assessments on the internet (Wroe, 2000; Ferguson & Glennon, 2001; Handler & Hunt, 2002);
- Access – Web-based testing offers global 24/7 access to testing and immediate scoring (Jones, 1998). Access is increasing rapidly (CyberAtlas, 1999) and the Internet is one of the main ways that people look for jobs. (Handler & Hunt, 2002). The Internet also enables test publishers to have the same questionnaire available in multiple languages for global use and candidates or test takers should have the opportunity to practice tests;
- HRIS Integration – Internet-based assessments are leading to a seamlessly integrated personnel assessment platform (Jones & Higgins, 2002) with the potential to vastly improve the communication between employers and candidates. In addition, the Internet increases the ability to monitor and approve the effectiveness of staffing methods by linking selection data with other data such as employee performance and tenure;
- Data collection – Internet-based systems are easier to maintain (e.g. test norms, test updates and software upgrades are only updated on one server vs. to all test users (Jones & Higgins, 2002) and data collection for norms and research purposes are drastically simplified (Bartram, 2001);
- Improved mechanisms – the potential for implementing Item Response Theory or Adaptive Testing is now better than ever due to single server maintenance. It is now possible to truly automate the assessment process;

- Decentralised decision-making - more and more line managers are involved in the selection process (IPD, 1999). This trend has continued and is similar in South Africa, which have experienced major restructuring in especially human resources departments. The Internet enables access to job analysis, assessment technology and reports to be delivered directly to line managers.

These advantages will become worthless if personality questionnaires are found to be unstable when administered over the Internet.

Together with new technology a number of challenges are born. These need to be considered and studied in order to establish the Internet as a reliable and effective alternative recruitment and selection mechanism. It is fair to comment that in the years to come a body of evidence will emerge to study the affect of the Internet on assessment and guidelines for best practice will emerge as the test fraternity is more established in the field of Internet assessment.
Let us study some of the fairness and best practice concerns in more detail:

- Performance - testing on the Internet should at least have two characteristics (Bartram, 2001). One, it should provide the means of controlling the timing and delivery and two, it should not fail mid-way through the test. It should also be able to cope with the demand of test takers and in the event of a breakdown in connectivity, the test should be seamlessly resumed from the point at which the break occurred;
- Security - Test users can pass test materials on to people who are not authorised to use them (Bartram, 2001) It also necessary to protect the test publisher's intellectual property, controlling access and distribution of tests and keeping scores and rules confidential, all of which are all high on the agenda of test publishers and psychologists (Bartram, 2001);
- Privacy - test publishers and test users should study and understand the risks associated with controlling access to test results and reports, the legal issues relating to data protection, privacy and storage (Bartram, 2001). In the Ferguson & Glennon study (2001) an alarming 81% of respondents did not strongly agree that the Internet is a confidential way of transferring assessment information. This indicates a serious need for education;

- Fairness - Groups being assessed should have equal access to the Internet and the assessments offered. Test taker honesty within the context of web-based assessment is a heavily debated topic. Bartram, (2001) offers one possible solution to honesty in the form of a signed honesty policy by candidates. Two options exist for candidate feedback when referring to web-based assessments. Candidates can receive feedback over the Internet or could receive feedback from a trained professional. Both have its advantages and disadvantages and discussion is outside of the scope of this article. However, test users should consider the assessment scenario and what design fits the assessment requirements best. Test quality is of the utmost importance when constructing tests for the Internet. Tests, which prove to be unreliable when administered over the Internet, will prove to be unfair and negatively impact on the perception of web-based assessments.

Aim

The aim of this study is to establish if the controlled web-based administration of the OPQ32n impacts on the stability of the questionnaire.

METHOD

Sample

Two sample groups of managers completed the OPQ32n for development purposes. The first sample consisted of managers from different industry sectors and various job functions. 245 were males and 77 females with a mean age of 36.69. The ethnic composition was 36.65% Africans, 3.42% Coloureds, 4.35% Indians, 54.97% Whites and 0.61% were from other ethnic origins. All managers in this group had at least grade 12 or more.

The second sample consisted of managers in the mining industry across various job functions and consisted of 258 males and 64 females with a mean age of 36.88. The ethnic composition was 32.30% Africans, 3.72% Coloureds, 6.21% Indians, 50.31% Whites and 7.46% were from other ethnic groups. As with the first sample, all managers in this group had at least grade 12 or more.

Measurement instrument

All respondents completed the OPQ32n for development purposes.

The OPQ series of personality questionnaires were designed to give information on individual styles or preferences at work and for use in a whole range of assessment and development applications.

The OPQ32 specifically, is an occupational model of personality, which describes 32 dimensions or scales of people's preferred or typical style of behaviour at work. It breaks personality down into three domains, namely:

- Relationships with people;
- Thinking style;
- Feeling and emotions;

and joined by a potential fourth

- Dynamism.

The OPQ32 model of personality is measured by two questionnaires: OPQ32n (normative) and OPQ32i (ipsative) and for the purposes of this study, the normative version was used by which respondents rate themselves on a 1 to 5 scale, ranging from Strongly Disagree (1) through to Strongly Agree (5). The questionnaire consists of 230 statements and the majority of candidates complete the questionnaire in 35-45 minutes.

The OPQ32 questionnaires are based on appropriate work-related and are designed for use by human resource, training, management development and personnel specialists as well as psychologists. The OPQ32 is particularly appropriate for use with professional and managerial groups, although the content of the scale deals with personality characteristics important to a wide variety of roles: (OPQ32 Manual and User's Guide, 1999).

The internal consistencies of two groups are reported in the OPQ32 Manual and User's Guide. In both cases are high levels of internal consistency achieved:

- the OPQ32n scales for a trailing group in the UK (N = 1228) range between 0,90 and 0,70 with a median of 0,84.
- For a general population group in the UK (N = 2028) it ranged from 0,87 to 0,65 with a median of 0,79.

Various studies in South Africa have been conducted which reports similar results. One such study, Study nr. R038, (2002), reported internal consistencies ranging between 0.69 and 0.88 with only one scale below 0.70 (Kriek & Joubert, 2002). Twenty scales' alpha coefficients exceeded 0.80 and 11 ranged between 0.70 and 0.79.

Test-retest reliability was also established. The reliabilities range after one month from 0,64 to 0,91 with a median of 0,79. This shows that the OPQ32n scores can be expected to remain stable over this period.

Satisfactory validity is reported in both international and South African studies.

Procedure

The first sample group completed the paper-and-pencil version of the OPQ32n in standard supervised test conditions.

The second sample group of managers completed the OPQ32n on the Internet. Special care was taken to consider all fairness and ethical aspects of web-based assessment. Regarding Performance, the on-line version of OPQ32 can be continued at a later stage if a break in communication occurs. Security was also considered by means of controlled access whereby all respondents were issued with a unique user name and password to log on. To protect the Privacy of each candidate the data is kept on a secure server and limited to the purpose it was attended for. Only authorised users can access the data. All participants are skilled in the use of the Internet, had access to the Internet and are provided with on-line support.

At first the second sample was considerably larger that the first sample (N = 616) and to ensure that biographical information does act as a moderator variable the samples were compared for significant differences. The size differences in the sample could be attributed to the white and coloured respondents. As a result a simple random sample was selected from the White and Coloured respondents reducing this sample to 322. Thereafter significant differences between gender, age and ethnic origin were considered. A t-test was conducted on the continuous data namely age and for the categorical data (gender and ethnicity) chi-square analysis was conducted. No significant differences are reported. Thus the sample groups are considered to be homogeneous.

One aspect, which could act as moderator variable, is industry sector. The second sample consists of manager in the mining sector only. So in the event that the stability of the OPQ32n is influenced by controlled web-based administration, it could be attributed to industry sector.

Statistical analysis

The analysis of stability involved comparing the supervised paper-and-pencil sample with the controlled web-based administration sample in three ways.

? In the first analysis test the internal consistency in supervised paper-and-pencil and controlled web-based will be considered.

  Reliability is the extent to which a test is repeatable and yields consistent scores. All measurement procedures have the potential for error, so the aim is to minimize it. The goal of estimating reliability (consistency) is to determine how much of the variability in test scores is due to measurement error and how much is due to variability in true scores. Internal consistency is commonly measured as Cronbach's Alpha (based on inter-item correlations). The greater the number of similar items, the greater the internal consistency.

? In the second analysis the Effect Sizes (*d*-statistic) of the supervised paper-and-pencil administration is compared with the web-based controlled administration.

To effectively assess meaningful outcomes, a metric other than statistical significance testing is useful, one that measures the magnitude of a result, rather than the probability that the result is due to chance (Cook, 1999)
The most commonly used metric to evaluate the magnitude of an outcome is "effect size" (Cohen, 1988). In this case *d* statistics is used for measuring the effect size difference between two means.

? In the third and last analysis of stability a comparison of covariance structures is carried out using Structural Equation Modelling with EQS. The fit is tested of a model that evaluates each pair of scale intercorrelations and each scale variance to be equal across the two samples.

RESULTS

The test reliability (ALPHA coefficient) in supervised paper-and-pencil and controlled web-based administration

Table 1 compares the APLHA coefficients and Standard Error of Measurement (SEm). Generally, an Alpha of 0.80 is considered as a reasonable benchmark for measures of ability and between 0.70 and 0.80 for personality questionnaires (Gliem & Gliem, 2003; Field, 2001).

ALPHA coefficients for the supervised paper-and-pencil sample ranges between 0.69 and 0.88 and the Social Desirability scale 0.66. Only one scale is below the 0.70 mark. ALPHA coefficients for the controlled web-based sample ranges between 0.71 and 0.91 and the Social Desirability scale 0.66.

Table 1.

	Supervised Paper-and-Pencil			Controlled Web-based		
	N = 322			N = 322		
Scale	**Alpha**	**SD**	**SE m**	**Alpha**	**SD**	**SE m**
Persuasive	0.82	4.76	2.02	0.83	4.67	1.92
Controlling	0.80	3.79	1.69	0.72	3.42	1.82
Outspoken	0.75	4.56	2.27	0.77	4.65	2.22
Independent Minded	0.73	4.73	2.46	0.71	4.48	2.43
Outgoing	0.85	5.13	1.96	0.88	5.42	1.87
Affiliative	0.80	5.27	2.33	0.82	5.41	2.30
Socially Confident	0.82	4.54	1.95	0.79	4.24	1.96
Modest	0.86	4.86	1.84	0.87	4.94	1.79
Democratic	0.77	4.65	2.22	0.73	4.35	2.25
Caring	0.76	4.58	2.26	0.71	4.16	2.23
Data Rational	0.84	4.43	1.77	0.84	4.56	1.82
Evaluative	0.69	3.60	2.01	0.71	3.56	1.92
Behavioural	0.81	4.33	1.87	0.85	4.88	1.86
Conventional	0.79	5.08	2.31	0.80	4.86	2.17
Conceptual	0.80	4.84	2.17	0.80	4.78	2.12
Innovative	0.85	4.46	1.72	0.85	4.27	1.64
Variety Seeking	0.73	4.59	2.38	0.74	4.13	2.12
Adaptable	0.80	4.78	2.12	0.80	4.56	2.06
Forward Thinking	0.82	3.86	1.62	0.84	4.00	1.58
Detail Conscious	0.81	5.51	2.40	0.79	5.45	2.47
Conscientious	0.70	3.42	1.86	0.79	3.80	1.73
Rule Following	0.88	5.33	1.84	0.87	5.14	1.87
Relaxed	0.81	4.52	1.98	0.85	4.76	1.81
Worrying	0.87	5.05	1.79	0.91	5.36	1.61
Tough Minded	0.85	5.08	1.99	0.86	5.16	1.90
Optimistic	0.83	4.77	1.94	0.82	4.44	1.88
Trusting	0.83	5.40	2.25	0.83	5.13	2.11
Emotionally Controlled	0.80	4.48	2.02	0.84	4.70	1.89
Vigorous	0.80	4.23	1.87	0.76	3.55	1.73
Competitive	0.85	5.15	1.99	0.84	4.78	1.90
Achieving	0.77	4.25	2.04	0.75	4.34	2.15
Decisive	0.78	4.40	2.04	0.78	4.27	1.99
Social Desirability	0.66	4.51	2.64	0.63	4.24	2.56
Average	0.80	4.63	2.05	0.80	4.56	1.99

For the web-based sample 15 scale reliabilities are higher than the paper-and-pencil sample, 5 remained the same and 12 somewhat lower. The average ALPHA coefficient is exactly the same for both samples. The average standard error of measurement (SE m) for the web-based sample (1.99) is slightly better (smaller) than the paper-and-pencil

sample (2.05). This is due to differences in the Standard Deviations (SD) for the two conditions.

As a result, it is clear from the analysis that controlled web-based administration does not compromise scale reliabilities.

Comparison of the Effect Sizes (*d*-statistic) of the supervised paper-and-pencil administration and the web-based controlled administration.

The two samples are compared using Effect Sizes (*d*-statistics). Mathematically, *d* reflects the number of standard deviations that the mean of the second sample is offset from the mean of the first sample. The effect size indicates the extent to which the two samples do not overlap due to the impact of the difference in the means.

Effect sizes are generally considered as small (effect size = 0.2), medium (effect size = 0.5) and large (effect size = 0.8) corresponding to correlations of 0.1, 0.3 and 0.5 respectively (Cohen, 1988).

Descriptive statistics as well as the effect sizes for both groups are shown in Table 2. From the data it can be deducted that there are only small differences in the scores of the two samples. The highest *d*-statistic is still a small *d of* 0.34 indicating that there is between a 78.7% and almost 100% overlap of the constructs between the supervised paper-and-pencil and controlled web-based groups.

Table 2.

	Supervised Paper-and-Pencil				Controlled Web-based				Combined Group				Effect sizes
	N = 322				N = 322				N=644				
Scale	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD	
Persuasive	23.68	10.00	32.00	4.76	22.68	9.00	32.00	4.67	23.18	9.00	32.00	4.74	0.24
Controlling	25.46	12.00	32.00	3.79	24.72	10.00	32.00	3.42	25.09	10.00	32.00	3.62	0.21
Outspoken	24.87	10.00	36.00	4.56	24.27	10.00	34.00	4.65	24.57	10.00	36.00	4.61	0.24
Independent Minded	21.29	10.00	34.00	4.73	21.82	6.00	36.00	4.48	21.55	6.00	36.00	4.61	-0.14
Outgoing	21.62	9.00	32.00	5.13	20.08	8.00	32.00	5.42	20.85	8.00	32.00	5.33	0.11
Affiliative	23.47	9.00	36.00	5.27	21.55	7.00	34.00	5.41	22.51	7.00	36.00	5.42	0.18
Socially Confident	23.48	13.00	32.00	4.54	22.64	13.00	32.00	4.24	23.06	13.00	32.00	4.41	0.14
Modest	17.86	8.00	32.00	4.86	19.31	8.00	31.00	4.94	18.59	8.00	32.00	4.95	0.01
Democratic	26.49	10.00	36.00	4.65	26.17	8.00	36.00	4.35	26.33	8.00	36.00	4.50	0.11
Caring	26.40	11.00	36.00	4.58	26.23	14.00	35.00	4.16	26.31	11.00	36.00	4.37	-0.21
Data Rational	25.32	8.00	32.00	4.43	24.18	9.00	32.00	4.56	24.75	8.00	32.00	4.53	0.10
Evaluative	27.39	14.00	36.00	3.60	26.89	14.00	36.00	3.56	27.14	14.00	36.00	3.59	0.13
Behavioural	27.39	11.00	36.00	4.33	26.45	11.00	36.00	4.88	26.92	11.00	36.00	4.63	-0.02
Conventional	16.29	4.00	35.00	5.08	15.54	4.00	28.00	4.86	15.91	4.00	35.00	4.99	-0.18
Conceptual	25.04	11.00	36.00	4.84	25.00	11.00	36.00	4.78	25.02	11.00	36.00	4.81	0.34
Innovative	25.12	13.00	32.00	4.46	24.55	9.00	32.00	4.27	24.83	9.00	32.00	4.37	0.12
Variety Seeking	26.93	10.00	36.00	4.59	27.86	12.00	36.00	4.13	27.40	10.00	36.00	4.39	-0.29
Adaptable	21.88	9.00	32.00	4.78	22.11	9.00	32.00	4.56	22.00	9.00	32.00	4.67	-0.07
Forward Thinking	26.30	14.00	32.00	3.86	25.83	12.00	32.00	4.00	26.06	12.00	32.00	3.93	-0.02
Detail Conscious	25.75	9.00	36.00	5.51	24.82	7.00	36.00	5.45	25.29	7.00	36.00	5.50	0.18
Conscientious	26.77	16.00	32.00	3.42	25.90	14.00	32.00	3.80	26.34	14.00	32.00	3.63	0.25
Rule Following	19.49	8.00	32.00	5.33	20.13	9.00	32.00	5.14	19.81	8.00	32.00	5.24	0.18
Relaxed	23.07	10.00	32.00	4.52	22.34	12.00	32.00	4.76	22.70	10.00	32.00	4.65	0.22
Worrying	19.51	8.00	30.00	5.05	20.45	8.00	30.00	5.36	19.98	8.00	30.00	5.22	0.04
Tough Minded	21.43	9.00	32.00	5.08	20.23	8.00	32.00	5.16	20.83	8.00	32.00	5.15	0.14
Optimistic	29.55	7.00	36.00	4.77	29.00	12.00	36.00	4.44	29.27	7.00	36.00	4.61	0.12
Trusting	22.08	6.00	33.00	5.40	23.54	9.00	34.00	5.13	22.81	6.00	34.00	5.32	-0.12
Emotionally Controlled	20.56	8.00	31.00	4.48	20.88	9.00	31.00	4.70	20.72	8.00	31.00	4.59	0.29
Vigorous	28.68	14.00	36.00	4.23	28.75	13.00	36.00	3.55	28.72	13.00	36.00	3.90	0.42
Competitive	19.05	8.00	32.00	5.15	18.25	8.00	32.00	4.78	18.65	8.00	32.00	4.98	0.22
Achieving	29.45	12.00	36.00	4.25	28.29	12.00	36.00	4.34	28.87	12.00	36.00	4.33	-0.29
Decisive	19.20	9.00	32.00	4.40	19.29	9.00	29.00	4.27	19.25	9.00	32.00	4.33	0.07
Social Desirability	19.01	5.00	32.00	4.51	18.31	10.00	31.00	4.24	18.66	5.00	32.00	4.38	-0.05

Equivalence of covariance structures using Structural Equation Modelling with EQS

A comparison of covariance structures was carried out using Structural Equation Modelling with EQS. The fit was tested of a model that evaluated each pair of scale intercorrelations and each scale variances, to be equal across the two samples.

There are a few statistics that measure how adequately this hypothesized model describes the sample data (Bartram, 2002). The comparative fit index (CFI; Bentler, 1990) ranges from zero to 1.00 and provides a measure of complete covariation in the data. Although a value >0.90 was originally considered representative of a well-fitting model, a revised cut-off value close to 0.95 has recently been advised (Hu & Bentler, 1999). Another fit measure is the root mean square error of approximation (RMSEA). This index has recently been recognized as one of the most informative criteria in covariance structure modelling; values less than 0.05 indicate good fit (Byrne, 2001).

The model tested had 529 degrees of freedom; CFI was 0.961 with RMSEA 0.026, indicating an exceptionally good fit. This analysis suggests that the pattern of relationships between scales can be considered the same for controlled web-based and supervised paper-and-pencil groups as the model fits equally well for both samples, across and within different modes of administration.

DISCUSSION

The results indicate that controlled web-based sample and supervised paper-and-pencil sample have comparable psychometric properties in terms of reliability and covariance structures. This implies there is no distortion to the instrument itself.

Furthermore the results also point out that that only a small difference exists between the scale means of the two groups. In fact scales overlap between 78.7 and nearly 100%, indicating, at most, a small difference between the means. This implies that the same norms can also be used for both conditions.

It is also important to note that industry sector had no impact on the data, which implies it did not act as a moderator variable.

The evidence reported here supports the use of the OPQ32n for controlled web-based administration conditions.

REFERENCES

Association for Assessment in Counseling, 2003. *Responsibilities of Users of standardized tests (RUST), 3rd ed.*

Barrick, M.R., Mount, M.K., 1991. *The Big Five personality Dimensions and job performance: A meta-analysis.* Personnel Psychology, 44, 1-25.

Bartram, D. & Brown, A., 2002. *Mode of administration and the stability of the OPQ32i.* Paper presented at the ITC Conference on Computer-bases testing and the Internet, Winchester.

Bartram, D., 2000. *Internet Recruitment and Selection: Kissing Frogs to find Princes.* International Journal of Selection and Assessment, Vol. 8(4) 261-274.

Bartram, D., 2001. *The impact of the Internet on testing for recruitment, selection, and development.* Paper presented at the 4th Industrial and Organisational Psychology Conference, Australia.

Bartram, D., 2004. *Assessment in Organisations.* Applied psychology: An International Review, 53 (2), 237-259

Bartram, D., Brown, A., Kriek, H.J., Joubert, T., 2003. *The similarity of the OPQ32n scales for subgroups in South Africa.* Unpublished.

Bentler, P.M. (1990). *Comparative fit indexes in structural models.* Psychological Bulletin, 107(2), 238-246

Buchanan, T., date unknown. *Online Personality Assessment: Equivalence of traditional and WWW personality measures.* [Online] Available from www.gor.de [Downloaded: 2004-04-10].

Byrne, B.M. (2001). *Structural Equation Modeling with AMOS. Basic concepts, applications and programming*. New Jersey: Lawrence Erlbaum Associates.

Cascio, W.F. *Applied Psychology in Human Resource Management, 4th ed.* Prentice-Hall, Inc.

Clarke, D., 2000. *Evaluation of a networked self-testing program.* Psychological Reports, 86, 127-128

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*, *2nd ed.* New York: Lawrence Erlbaum.

Cook, L.B., 1999. *Effective uses of "effect size" statistics to demonstrate business value.* Quirk's Marketing Research Review.

CyberAtlas (1999a) *Worldwide Internet users to pass 500 million early in the next century.* [Online] Available from http://.cyberatlas.internet.com [Downloaded: 2004-04-01]

Digman, J.M., 1990. *Personality structure: Emergence of the five-factor model.* Annual review of Psychology, 41, 417-440.

Donovan, M.A., Drasgow, F., & Probst, T.M., 2000. *Does computerizing paper-and-pencil job attitude scales make a difference? New IRT Analyses offer insight. Journal of applied psychology,* 85,305-313.

Ferguson, N., Glennon, R., 2001. *The candidates experience of the online OPQ.* SHL, London.

Field, A., 2001. *Overview of reliability analysis.* Unpublished.

Gliem, J.A., Gliem, R.R., 2003. *Calculating, Interpreting and reporting Cronbach's Alpha reliability Coefficient for Likert-type scales.* Paper presented at the 2003 Research to Practice Conference, USA.

Handler, C.A., Hunt, S.T., 2003. *Buyer's Guide to Web-based Screening & Staffing Assessment Systems.* Rocket-Hire, New Orleans.

Harris, M.M., 1999. *Practice Network.* The Industrial-Organizational Psychologist, 36, 89-93.

Hu, L. & Bentler, P.M. (1999). *Cut off criteria for fit indexes in covariance structural analysis: Conventional criteria versus new alternatives.* Structural Equation Modeling, 6(1), 1-55.

Jones, J.W., (1998). *Personnel Testing: A manager's guide.* Crisp Publications, California.

Jones, J.W., Higgins, K.D., 2002. *Mega trends in personnel testing: A practitioner's perspective.* Unpublished.

Kriek, H.J. & Joubert, T., 2001. *The effectiveness of the SA WAIS, CPA, the OPQ32, the MMPI – 2 and 16PFi in the selection of Supervisors.* Study no V019.

Kriek, H.J. & Joubert, T., 2002. *The internal consistency of the OPQ32n for a mixed occupational group in South Africa*, Study no R038.

Lievans, F. & Harris, M.M., 2003. *Research on Internet Recruitment and Testing: Current Status and Future Directions.* In I. Robertson, & C. Cooper (Eds), The International Review of Industrial and Organizational Psychology, Chichester, England: Wiley.

Matthews, G., 1997. *The Big Five as a framework for personality assessment.* International Handbook of selection and assessment, p475-492.

Mauer, K.F., 2000. *Psychological test use in South Africa.* [Online] Available from http://www.pai.org.za [Downloaded; 2004-03-12]

McCrae, R.R. & Costa, P.T. Jr., 1987. *Validation of the five-factor model of personality across instrum ents and observers.* Journal of Personality and Social Psychology, 52, 81-90.

Muchinsky, P.M., Kriek, H.J., Schreuder, A.M.G., 1998. *Personnel Psychology*. International Thomson Publishing.

Neuman, G & Baydoun, R., 1998. *Computerization of paper-and-pencil tests: Wehen are they equivalent?* Applied Psychological Measurement, 22, 71-83.

OPQ32 Manual and User's Guide (1999), SHL Group plc.

Park HR, 1999. *Graduates in the eyes of employers.* Park HR and The Guardian.

Pfenninger, D.T., Klion, R.E., Wenzel, M.U., 2003. *Integrated Web Assessment Solutions.* Paper presented to the International Military Testing Association.

Price, R. E. & Patterson, F. (2003). *Online application forms: Psychological impact on applicants and applications for recruiters.* Selection and Development Review, Vol. 19, p12.

Reed Executive plc., 2002. *Paper prejudice hits job seekers.* [Online] Available from http://cws.huginonline.com [Downloaded 2004-04-10].

Rigdon, E., 2004. *Structural Equation Modeling: An overview.* [Online] Available from www.gsu.edu [Accessed: 2004-02-28].

Roth, P.L., Bevier, C.A., Bobko, P., Switzer III, F.S. & Tyler, P, 2001. *Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis.* Personnel Psychology, 54, 297-330.

Society for Industrial and Organizational Psychology Inc., 2003. *Principles for the validation and use of personnel selection procedures, 4th ed.*

Thorndike, R.L., 1971. *Concepts of culture fairness.* Journal of Educational Measurement, 8, 64.

Wroe, N., 2000. *SHL survey suggests online selection processes increase positive perception of companies.* SHL News release, Bolder, Colorado.

Zieky, M., 2002. *Ensuring the fairness of licensing tests.* CLEAR Exam Review, Vol xii, 1, p20.