"Characteristics of a Sound Test and Computer-Assisted Testing: An Evaluation of the Vocational Interest Questionnaire (VIQ) in Malaysia"

By Lee Kok Wai

At Kolej Damansara Utama, Petaling Jaya, Malaysia

1.0 Introduction

The use of testing and assessment has long been considered an integral aspect in counseling. For instance, the psychometric tradition, from which psychological testing emerged, is considered to be a core foundational aspect of counseling psychology (Super, 1955, Watkins, 1983; Whiteley, 1984). In fact, assessment has become a lucrative business. It has been reported that since its first publication in 1971, the Self Directed Search has been completed by over 9 million people. The Myers Briggs Type Indicator in 1988, recorded over 1.5 million users and 1 million strong Campbell Interest Inventory Profiles were scored in the same year (Walsh, 1990). Besides standardized tests, Goldman (1990) has revealed the use of hundreds of qualitative assessment methods as reflected in the mainstream of counseling psychology, particularly those that have emanated from humanistic education, situational assessment, and rehabilitation counseling. These methods have evolved and been modified to suit relevant client or population.

Psychologists have developed a huge arsenal of tests, techniques and procedures that cover a broad range of human activity. They are constantly being used to assess accurately the moods, desires, anxieties, fears, personality, intelligence, values, potential, and other dimensions of the client (Lonner, 1985).

In a survey conducted by Watkins, Campbell & McGregor (1988), a sizable number of counseling psychologists are involved in testing and assessment activities. Currently, counselors use a variety of assessment methods, ranging from intellectual to vocational to interests. There is a diverse use of instruments in various work settings. There are cases where counselors are trained with testing knowledge to administer specific tests and assessment procedures to the needs of his or her setting while some relevant might be taught on general tests. In both cases, counselors are more readily equipped to help clients valuable

information from these tests to better understand themselves that aid in coping with real-life problems.

Graham and Lily (1984) defined assessment as "the entire process involved in collecting information about persons and using it to make important predictions or inferences. Testing is described as the use of tests (e.g. interest or personality tests), which is one means of gathering data during the assessment process. Testing is normally a part in the assessment process, but assessment can extend beyond the exclusive use of tests and instruments (Watkins, 1983).

In addition to an awareness of the role of the counselor and the importance of assessment data, it is evident that testing is a tool that help increase client's self-awareness and problem solving which occurs in the counselor / client interaction. In other words, a carefully selected test can be used as a means to gather useful and quantifiable information to help make empirical predictions about current or future behavior. With appropriate and valuable assessment data, a counselor can help the client cope with his or her issue. Today, this value is evident with the widespread use of tests. For instance, tests of ability and achievement have gained public acceptance. Testing of paper and pencil kind has served many institutionalised functions, including accountability in assessing the effectiveness of teachers and colleges and selection and placement of students in relevant programmes. While tests have their significant role in assessment, we must also keep in mind within this context that tests are done primarily for the benefit of the client. Assessment, whether by objective and standardized test or by qualitative methods, the assessor or tester should regard it as an aiding process to solve client's problems. In this respect, tests just like all tools require skillful handling and thorough knowledge of what they will and will not do. Apart from this, perhaps a more significant point is that a counselor should be able to evaluate these tools that are relevant to the client and how they can integrate and make the assessment process a meaningful part of treatment.

2.0 Current opinions about tests and some implications.

People are poor judges of other people. Our prejudices and biases, based on race, gender, age, language and education, all too often mar our views of people. Our stereotypes operate in our assessment of people. (Dahlstrom, 1993). There is too much fallibility of human judgments especially when one's judgments and evaluations are clouded by negative halos, personal deficiencies and incomplete paradigms. In a move to obtain greater objectivity, diagnosis and predictions, counselors rely on test findings that serve to reduce the likelihood of the adverse effects that come from misjudgment. Regardless of work setting, appropriate tests when carefully employed can provide significant insights to both client and counselor. This is perhaps one of the many reasons that suggest why testing is a highly purposeful activity. Therefore, tests can serve as a means to overcome the impact of these distortions only if we know enough about a test and how to review its suitability. In other words, testers should be able to evaluate, review and pick the accurate test that best serve the client's and counselor's needs.

While there are arguments supportive of standardized tests due to their cost-efficiency, practicality, accountability and objectivity, it is useful to also highlight some arguments critical of standardized tests. Miller-Jones (1989) forwarded some criticisms of standardized tests of achievement and ability from the cultural perspective chiefly, the difficulty to infer underlying cognitive processes from performances on standardized tests; the cultural and gender biases in the language and content of specific items; the complexity that requires knowledge in which thinking occurs in the given time and place as well as the role of environment and cultural practices. Issues reflected in these arguments for and against tests should be considered as we discuss the need to review the relevant criteria for selecting a suitable test.

Drummond (1996) presented a public perspective toward the use of testing and its consequences. The influence of technology in this current information age will certainly bring about higher frequency of decision-making. Constructive and informed decision-making are dependent upon valuable information.

The use of IT will have immense ramifications to the current practices and procedures used in testing. Powerful software and hardware are able to collect and download enormous data while providing real time feedback and remediation online. More clients are now able to observe and reflect on their own performance from a test bank with self directed and user friendly steps. Self-directed procedures will also allow testees to perform test interpretations. There will be greater need to review and customise tests to conform to professional bodies and meet clients' needs. Though computer-assisted and computer-adaptive testing are gaining wider following and public acceptance, counselors will need to exercise greater accountability, professionalism and care in the use of such tests.

Computer-assisted testing and assessment should be congruent with the goals of the counseling profession. Perhaps, counselors need to be reminded that computer-based test interpretation (CBTI) is merely a supportive mechanism or best used as a 'second opinion' to confirm or disconfirm hypotheses about a client (Sampson, 1990). Testing involves more than test scores that are interpreted by computer. It should take place within the context of all therapeutic alliance. In other words, we need to keep in mind that computerized testing is not magical and is less-than-perfect tool because it relies on the developer (human) to integrate and make meaningful links and applications. The necessary commitment of the human factor is relevant here. Besides, Holland (1979) highlighted some cogent implications for psychological testing. He has revealed the impact of testing on the public, its technical problems, controversies and how it impacted organizations. According to Holland, the impact of intelligence, admissions, and employment tests must be estimated by stringing together statistical studies of probable impact of a test, by a small number of legal rulings, and by anecdotal accounts of positive and negative outcomes. With this comprehensive and explicit knowledge of testing influences, the industry can stimulate the development of better tests while at the same time providing valuable information for an effective defense of sound tests and testing practices.

Because test results guide decision-making, reliability and validity of test become very important. It is evident that a test's reliability and validity are vital characteristics especially when such test has a significant role in the educational and social system. Admissions to higher education and employment in the market place have test taking done for decision

making. Imagine what a defective test can do to the industry if the test is discriminatory in nature, not reliable and the controversies and dissatisfaction it can produce?

Another adverse repercussion is how the failure to assess human potential can result in wastage of enormous human talents and capabilities. Imagine if a test is gender or culturally bias? Test information would be inaccurate, does not help improve one's self-understanding and it would not be able to screen incompetent professionals in a chosen field. Inaccurate diagnosis would certainly result in harmful prescription. Unreliable test information will not help predict human behavior. Social resources would not be efficiently harnessed and utilised. For instance, jobs would not be adequately designed to help foster job satisfaction and productivity if the diagnosis of work values done through a test did not aid in creating this environment. Defective data would definitely have negative impact on administrators, teachers, counselors, psychologists, curriculum experts, students, researchers, the public and the legislature. In short, such impact reach all groups who are consumers of test information.

Paralleling this wave of implication and concern is the need to examine some key criteria for selecting a test and identify some features of a good test. Selecting the right test is likened to choosing the right farming tool that would aid the harvesting process. These criteria should be considered prior to using any test. Often, helpful reviews would enhance testing as a method of aiding developmental problem-solving, decision-making and psychoeducational methods of helping people work with their problems (Duckworth, 1990).

3.0 Characteristics of a sound test

The following paragraphs provide some key features that contribute to a sound test. These characteristics should provide useful guidelines to both tester and testee. These characteristics help overcome ambiguity and offer a useful checklist to all test users cope with the overwhelming tests that are available in the market. By running through these characteristics test users can discriminately discern and choose a test that best apply and meet their needs. These characteristics will also help raise questions to the users when trying to explore the strengths and weaknesses / limitations of a particular test. Testing has been dominated by psychometricians and statisticians (Holland, 1979). This does not mean that other test users should be ignorant of these characteristics where most sound tests are based on. By

becoming aware of these basic ingredients of a good test, test users can fully optimise the benefits of this test in order to help address client's immediate needs effectively. Effectiveness would mean offering the counselor valuable insight into the client more rapidly than is possible through the use of counseling alone. Similar to the way a medical doctor uses x-rays and blood tests to focus on the patient's medical problems do quickly and as accurately as possible, testing has the same possibilities for the clarification of the client's personality and / or vocational interests and abilities, thereby accelerating counseling and saving the client money and valuable time (Duckworth, 1990). While not all tests can meet these criteria, they provide general or universal criteria when a review or evaluation is done. Drummond (1996) for instance, recommended a committee to review a particular test in order to gain various perceptions of the test. However the preferred way to do this may be, these criteria are necessary to minimise the risks of using an inappropriate or unreliable / poor test. Dahlstrom (1993) regarded counselors as gatekeepers of the future who can be armed with a whole new array of screening and prognostic indicators to help improve the lives of their clients. Information based poorly constructed or inadequately standardized tests can serve to mislead and distort clinical judgments based on biased samples of behavior.

Dahlstrom (1993) recommended six key features that are reflected in a well-constructed test.

1. Established validity - The robustness of measures in the test comes from the items able to stand up to real-world criteria. Such validity is usually a result of extensive test data that are obtained from the real-world.

2. Appropriate Norms - There must be appropriate norms for effective interpretation that are based on empirically determined cutting scores and correlates.

3. Objective scoring - Clear and unambiguous scoring standards. There is high interjudge reliability by spelling out these criteria in sufficient detail.

4. Immediate Recording - Important 'facts' of a person's history or behavior are captured without delay.

5. Optimal Motivation - A good test should be able to keep interest high for testees. Motivation to complete can be enhanced by the simplicity of the response format, by the comfortable style of the wording of the items, and by the wide-ranging content of test statements. A published instrument should incorporate methods of documenting a subject's test-taking compliance.

6. Standard Materials & Procedures - An established mode of administration plus the relevant materials are necessary to allow immediate recording and preserve the standard meaning of the test stimuli.

Groth-Marnat (1990) outlined certain key considerations that would help evaluate whether a test is appropriate and adequate. They are:

1. Theoretical orientation

   Counselors should research the construct that the test is suppose to measure and then examine the manner in which the test approaches this construct. Such information is normally found in the test manual. An individual analysis of test items can help a potential test user to evaluate whether or not these items appear relevant to the trait being measured.

2. Practical considerations

   This relates more to the context and manner in which the test will be used. The examinees' level of education and ability to read, comprehend and respond to the test are crucial. Long tests can result in frustration and lost of interest. If tests require interpretation and certain way of administration, then training in these areas are necessary.

3. Standardization

   Each test has norms that reflect the distribution of scores by a specific standardization sample. The basis upon which individual test scores have meaning relates directly to the similarity between that individual being tested and the standardization sample for meaningful comparisons. Test norms are important for effective comparisons. The standardization group should also be representative of the population on which the

examiner would like to use the test. Test manual should include sufficient information to determine this representativeness. Inadequate information on this can reduce the degree of confidence in the test. A good test will have specialized subgroup norms as well as broad national norms. This is particularly important when subgroups produce significantly different sets of scores from the normal standardization group. Knowledge of certain subgroup norms allows for a more appropriate and meaningful interpretation of scores. A well-constructed test should also include instructions that allow the tester to give the test in a similar structured manner as other examiners and to also maintain thus standardized administration between one testing session and the next. Any variation of administration procedures can influence the quality of responses, thereby tainting the reliability of the test.

4. Reliability

This refers to its degree of stability, consistency, predictability, and accuracy. In short, it refers to the extent to which scores obtained by a person will be the same if the person is reexamined by the same test on different occasions. There should be minimal degree of random fluctuation in the individual scores. This degree of random fluctuation signifies the degree of measurement error; especially due to some chance fluctuation.

An examiner should expect higher reliability for an intelligence test than for a test measuring a personality variable like anxiety due to influence of factors outside the test. A good test should perform adequate test construction that reduce the degree of imprecision in the test. A sound test should register a 'high' correlation that is .80 or more. As a general rule, therapists or clinicians should hope for correlations of .90 or higher in tests that will be used to make decisions about individuals particularly when theirs results are used to guide and focus appropriate interventions designed to minimise the risks of later psychopathological or psychosomatic disorders, substance abuse or criminal involvement. Comparatively, a correlation of .70 or more would be regarded as adequate for research purposes. From another angle, the purpose of reliability is to estimate the degree to which the test varies due to error. Reliability can be obtained through: test-retest (test produces consistent results upon retesting), alternate forms (the relative accuracy of a test at a given time), split half (the internal

consistency of the items) and interscorer (the degree of agreement between two examiners). A popular way to sum up the reliability methods are time to time (test-retest), form to form (alternate forms), item to item (split half), or scorer to scorer (interscorer).

5. Validity

This is the most crucial issue in test construction. Validity assesses the degree the test is to be accurate. A test that is valid should measure what it is intended to measure and should also produce information useful to the counselor. Validity must be viewed within a particular context and for a specific group of people. Although a test can be reliable without being valid, the opposite is not true. In other words, a valid test is always reliable. To put it simply, a necessary prerequisite for validity is that the test must firstly obtain an adequate level of reliability. A good test is a valid test that accurately measures the variable it is intended to measure. In constructing a test, a test designer must ensure that the construct is theoretically evaluated and described and secondly, specific test questions (operations) must be developed to measure it. Messick (1995) argued that the traditional conception of validity namely, content, criterion, and construct validities are incomplete. He proposed a new unified concept of validity that address score meaning and social values in test interpretation and test use.

For a test to be considered useful and efficient, it must be able to produce accurate results above and beyond the results that could be obtained with greater ease and less expense. A good test should produce high accuracy that is worth the time and money than information obtained through basic information such as biographical data and knowing the referral questions.

Drummond (1996) proposed several criteria for selecting a test. They are:

(a) Does the test manual describe the purpose of the test? Does the intended purpose correspond to one that the test was designed to accomplish? A good test manual states the purpose(s) of the test and a description of the behaviors it measures.

(b) A good test should be accompanied by a test booklet. This test booklet will provide test items phrased and identified to measure certain traits, objectives or behaviors.

(c) A test manual should make available validity information. If prognosis is needed, tester should be able to refer to the manual for evidence of criterion-related validity. A tester should also be able to check the authors' evidence to support the construct validity of the test as well as check the content validity if test results are used to describe competencies in a certain subject or in content fields.

(d) Reliability

The tester should be able to check the reliability information in the test manual, inspecting not only the types of reliability cited but also the magnitude of the coefficients reported. In the event of test results will be interpreted for individuals, the tester should ensure that the manual provides information on the standard error of measurement.

(e) A test should provide the necessary information to make a proper interpretation. The test author/s should provide information on how procedures to translate raw scores into derived scores and provide appropriate tables for different types of scores. The test author should also describe the techniques for interpreting the test and adequately explain the meaning of the various scales or on a specific items.

(f) Interpretive Feedback for the testee. A good test should provide immediate (if possible) feedback on screen (especially if administered on-line) and perhaps self-interpretive leaflets or booklets that testee can use to interpret and understand their results.

(g) Suitability

A good test is both suitable and practical. If is appropriate for the client in terms of vocabulary or sentence structure used. The format of the test is user friendly, easy and not cumbersome. The manner of responding to the questions is also easy. If the test is used for individuals with special needs, then the test should consider administrative factors such as place, equipment or other facilities.

(h) Unbias

Test authors should proof in their manuals of the procedures used to eliminate bias. It should be cultural or gender free.

(i) A good test should be effectively administered by trained and qualified personnel. This is particularly so if the test requires specialized training for greater accuracy and validity.

(j) Needs

A good test should meet the needs of both the examiner and the client. Needs are bounded by cost of the test, answer sheets and scoring services. In the event if external profile centre or assessment agency is needed to score and interpret the results, time is another important factor to consider. Other aspects such as usability, ease of use and confidentiality may need to be considered too.

4.0 Evaluation of Computer-Assisted Testing – An actual example used in a Test Centre in Malaysia

The following is an evaluation of a computer-assisted testing, the Vocational Interest Questionnaire (VIQ), an interest Questionnaire that helps individuals with career direction by clarifying a person's work interests and specifying what jobs seem most relevant. This Questionnaire is currently used across Australia and in Malaysia, covering secondary, tertiary and commercial settings.

This evaluation will attempt to articulate the key characteristics of the VIQ test on how congruent this test with some of the suggested characteristics of a good test. Some important features of computer-assisted testing are analysed and reviewed in light of its objectives and goals.

4.1 Test Title	:	Vocational Interest Questionnaire (VIQ)
4.1.1 Description of test	:	The VIQ helps individuals chart their career paths by clarifying their work interests and specify what jobs seem most relevant or compatible to their interests and appeal. The Internet version discussions began in October, 1997. The beta version of the VIQ was ready to trial in June 1998. Techworks (an on-line learning centre in Adelaide) administers the test on line. (Please refer to Appendix A for more details)
4.1.2 Test Centre	:	Kolej Damansara Utama, Petaling Jaya, Malaysia.
4.1.3 Test Author	:	Dr Darryl G Cross
4.1.4 Author's qualifications	:	BA (Hons), PhD, MADS, AFAIM Director, Registered Psychologist. (Please refer to Appendix B for more details)
4.1.5 Author's address	:	Darryl Cross & Associates 15 Ruthven Ave, Adelaide SA 5000, Australia. Phone : (08) 84100388 Email : dca@senet.com.au.
4.2 Description of scales	:	The VIQ has been correlated with the California Occupational Preference Scale (COPS). The result has been high and both scales were considered to be measuring similar constructs. The COPS scale has been used for some years in

a number of schools in Adelaide and was widely distributed through the Australian Council for Educational Research (ACER) in Melbourne.

4.3 Hardware requirements : CPU (Intel Pentium Processor)
Memory (Minimum 15 MB)
Monitor (Colour display)

4.3.1 Software requirements : Web Browser, Microsoft Internet Explorer, Microsoft Outlook Express and Adobe Acrobat Reader.

4.3.2 System : Pentium (Intel)

4.4 Cost : AD 25.00 for Interest Profile
(Discounted to AD 10.00 for bulk purchasing and a further discount to AD 7.50 if over 100 in one year)

4.4.1 Populations served :
- High school
- Special populations
- Young adult
- College / Junior College
- Adult

4.4.2 Potential users :
- Administrators
- Counselors
- Psychologists
- Teachers
- Students / examinees

4.4.3 Categories :
- Student
- Member of an organisation
- School
- Individual

4.4.4 Item selection / Presentation :
- Access to direction
- Type of item selection

4.4.5 Answering / Scoring :
- Backup and changing answers
- Answer registration method

4.4.6 Features available on digital	:	• Test scoring • Test score interpretation • Record-keeping (individual records) • Item analysis • Reliability information • Graphic presentations • Self-administered
4.5 Administration Feature	:	Administration Feature (Please refer to Appendix C for more details. Only refer to Interest Profile section and instructions on how to complete the VIQ)
4.5.1 Accessibility	:	VIQ Internet site http://www.findacareerpath.com A pin is required upon payment.
4.5.2 Printing of Profile / Report	:	Use Adobe Acrobat Reader or download a free copy for viewing and printing.
4.6 Questionnaire	:	220 questions displayed on the screen, on a series of windows. Testee click on appropriate responses. The VIQ does not ask about a person's liking for particular jobs. This is because people either do not know what if entails or have inappropriate information about it. Instead, the VIQ asks people to rate their liking for doing a specific tasks, skills or activities. Questions are aimed at assessing interests in career only. A copy of digitised version is not available on print. Instead, a paper and pencil format is enclosed. (Please refer to Appendix D)

4.7 Reliability and validity : Validity (concurrent)
(construct)
(Please refer to Appendix E)
Content validity (face validity has been listed under 'Testimonials'. These testimonials include quantitative and qualitative Evaluations.

4.8 Opinions about the VIQ

Using several key and technical characteristics of computer-assisted testing, a general evaluation is done on the VIQ. Certain key characteristics of a good computer-assisted test are also used to evaluate its effectiveness. The Evaluator walked through the process of scoring and received a generated VIQ profile / report. Table 1.1 refers to the technical Evaluation of the VIQ.

Table 1.1: Technical Evaluation of VIQ

Items / Technical Characteristics	Rating			
	1 (Excellent)	2 (Good)	3 (Poor	NA (Not Available)
• User Friendliness		✓		
• Test Norms (Asian-subgroup)				✓
• Information on internal consistency				✓
• Validity				
- Conceptual Validity				✓
- Incremental Validity				✓
- Construct Validity		✓		
- Concurrent Validity		✓		
- Predictive Validity				✓
• Information on culture fairness				✓
• Information on conceptual equivalence				✓
• Local (Malaysian) testimonials				✓
• Information on reliability (Reliability coefficient)				✓
• Storage capacity				✓
• Clarity of report		✓		
• Ability to change score		✓		
• Response time between input and feedback		✓		
• Ease of operation		✓		
• Ease of input / entry		✓		
• Graphic Presentations			✓	
• Printing		✓		
• Flexibility		✓		
• Link to other applications			✓	

Items / Technical Characteristics	Rating			
	1 (Excellent)	2 (Good)	3 (Poor	NA (Not Available)
• Display conventions				
- Format		✓		
- Headings		✓		
- Attractiveness			✓	
- Color			✓	
- Highlighting			✓	
- Reading level		✓		
- Help options		✓		
• Error checking				✓
• Parallel Forms		✓		
• Usefulness		✓		
• Follow through		✓		
• Bahasa Malaysia version				✓
• Counseling reference after test result				✓
• Information cross-cultural use of test				✓
• Theoretical constructs of test				✓
• Security		✓		

4.9 Profile analysis : VIQ

(Please refer to appendix F for more information on:

(a) How to use report

(b) Testee's Interest Profile

(c) Descriptions on the areas – liked most and liked least

(d) Explainations on positions

(e) Statements about areas of interest

(f) Types of jobs which relate to interest areas

(g) Action plan

More information on jobs are available at http://www.deetya.you.au/jobguideonline

5.0 Strengths of the VIQ

1. Quick, easy and economical collection of information.
2. Scoring accuracy.
3. Possible alternate formats.
4. Fast feedback to the test taker.
5. Convenience. Clients have a choice of time and date and place to perform this test. This allows a more personalized learning experience in a controlled environment.

6.0 Limitations of Evaluation

The ratings given were merely reflection of one evaluator based on his personal opinions about the VIQ. An individual's evaluation of a particular test has many limitations especially when one's judgment is bias. A panel of evaluators would be preferred. A formal experts' review of how good each item for instance, can help ascertain content validity when the VIQ is assessed by individuals with expertise in computer-assisted testing and relevant content knowledge of the subject under study. Evaluation was also done based on knowledge limited of the VIQ available on the web site. No substantial and adequate information on certain aspects under review. Information on several items were not available. There was no reliability analysis by the evaluator to assess its degree of the VIQ reliability. For example, no concurrent validity was performed and established. No study was done to see if measures or items in the VIQ correlates with 'gold standard' measures of an established and similar computer-assisted testing on interest. There was also no locally available test of the similar kind to be benchmarked. Evaluation was done fairly cursory on some key characteristics of a sound test. The features highlighted were usually found in standardized paper and pencil tests with some basic features in a typical computer-assisted testing. Regarding construct validity, no theoritical measure was done by the Evaluator to measure how meaningful the VIQ is. This is, of course, would be time consuming and is best determined by numerous investigators.

7.0 Discussion and recommendations

The evaluation suggest that a number of areas that might be raised about the digitised VIQ. Several observations about the VIQ seem worth discussing. First, the VIQ was clearly preeminent as the measure of interests. It has the strengths of most computer-assisted testing but certain vital information are necessary for the test taker that would help increase her / his confidence over the use of the VIQ. These information include test norms, culture fairness, conceptual equivalence, local testimonials and empirical reliability estimates. It is necessary to highlight important psychometric features such as reliability, validity, and the establishment of relevant norms (Lonner, 1985). This is more so when any test is 'anchored' to an originating culture and used in another culture area. To impose an etic approach, that is to employ constructs as if they mean the same in the target or nonoriginating culture may potentially open up numerous opportunities for being misuse and disservice. One consideration is perhaps to establish a conceptual equivalence or metric or equivalence. This scale measures the same behavioral property in a different culture (Lonner, 1985). In the context of Malaysia, linguistic equivalence is perhaps the most important single stumbling block to effective cross-cultural testing. The linguistic variable can range from a simple term to complex written instructions. A Bahasa version might need to be in place if the VIQ is to be considered for more extensive consumption especially in the lower and upper secondary schools. Words and terms found in the VIQ that might reflect linguistic concern are 'bidder', 'field trials', 'drugs', 'event', 'prototype', 'preservation', 'odds on entrants', 'motel', 'trucks', 'physical world', 'landmarks', and 'leisure'. Cultural differences may contribute to different-sized reliability coefficients. A local norm / subgroup norm could be established from the originating culture (where test is developed). This can be done by gathering new norms for Malaysia or by setting new cut off points for other cultural or ethnic groups if the existing norms are used. Research in this area can be valuable when different test norms of the VIQ are established for the dominant races in Malaysia, namely Malays, Chinese and Indians. Norms could also be established based on geographical boundaries such as rural and urban schools. In other words, etic measures developed within our local context can contribute to more accurate understanding thus ending divergent and convergent validation (Campbell and Friske, 1959). By developing

culturally relevant norms, one is not suggesting to restrict its usefulness and discounting its original intent of the test. A brief summing up on this point on the need to be contextual and purposeful is perhaps best explained in Walter J. Lonner's comments about counseling. "All counseling is 'Cross-cultural' in the sense that both client and counselor bring to the counseling situation their differing needs, values, expectations, and beliefs that have been shaped by cultural factors".

The CBTI nature of the VIQ is best used as a second opinion to confirm or disconfirm or to check out perceptions about a client. Assessment entails more than scores interpreted by computer. The VIQ generates a rather descriptive profile about the test-taker. A more evaluative perspective of interests might be helpful to create a balanced interpretation. The results from the VIQ should be fed into the counseling process that aid problem-solving by the counselor to make meaningful associations. Testing should not be divorced from a therapeutic alliance. A referral to a qualified counselor is needed to help clarify certain concerns and interpretations of the profile.

Another theme running through this thought is that test do not think. A test does not make decision. In most cases, a client needs someone to help provide proper guidance by exploring certain alternatives to decision-making. Besides, environments and humans are in constant interaction. Assessment of an individual is not complete without some assessment of the environment. How individuals perceive their environments and how their perceptions influence their behavior in specific contexts remains to be vital. Qualitative and quantitative data collection methods can be used to get a more complete developmental perspective. The test author of the VIQ might want to consider incorporating supplementary client-screening procedures to gather more information about the client prior to the test. Interviews and observation of behavior can help foster an active role for the client in the process of collecting and teasing meaning out of data. A computerised interview or verbal pictorial stimulus presentations can help promote holistic study of the individual rather than the isolation and precise measurement of narrowly defined discrete elements of interest. Such computer interview can generate interpretative comments and supplementary diagnostic information for interest.

Two hundred and twenty items are rather long, perhaps some graphics could be incorporated into the screens to sustain interest. 3.5 hours (average 1 minute per question) to score can be a daunting task. Sound and still-frame visuals can be featured in the test to enhanced test-taker's motivation. Based on the history of VIQ, since the last 5 years it was first designed as a paper and pencil test. Only until 1998, it was digitised. Another point to consider is that a valid paper-and-pencil version does not automatically result a computerised version being equally valid (Johnson & Mihal, 1973). This is because a computer administration may change the nature of the task and alter a subject's responses to the task, thereby resulting in questionable validity. Some information can be provided to the tester on the difference between the test validity done on these two formats and either the validity of individually computer-generated interpretations or the validity of computer-generated narrative reports.

Adaptive devices can also be considered to provide disabled clients with alternative data input options. This is worth considering if the test is to be introduced to people with special needs. Some of such devices include voice input, simplified keyboards, joysticks, pneumatic controls, head pointers, and braille keyboards. Through the use of the computer-mediated adaptive devices, the validity of the testing and assessment process can be enhanced and completed with minimal staff assistance.

8.0 Conclusions

This study reviewed several key characteristics of a good test. Some key features of computerised assessment were also highlighted. Also examined in this article was an evaluation of the VIQ employed by a local test centre. The VIQ was used as an example of a computer-assisted testing in view of the popularity of computer-assisted assessment. It's strengths and limitations were discussed. Even though there is a vast majority of tests are now automated and this trend will continue to do so, this study is aimed at reminding all present and future computer-assisted test users of some of the pertinent points and characteristics of a sound test. All test users should be able to assess these tools with confidence and a basic knowledge of the relevant characteristics is necessary. By subjecting a test to point-by-point evaluation, a

counselor can help client choose an appropriate test that best meet his or her needs. Some contemporary views issues, implications and recommendations were also discussed. Careful selection of a test is important so that clients and counselors can reap the maximum benefits from accurate test usage while minimising the negative consequences of using an inappropriate or poorly constructed test. An effective usage of test should be regarded as a combination of computer-assisted testing and human involvement in the assessment process. Test users should always reflect on the goals of counseling. Testing should not be taken out of the context of an interpersonal relationship. The counselor's touch is needed to help integrate and make the assessment process a meaningful part of the counseling process. He or she can help increase client's self-awareness and use the test result to help the client cope. Qualitative methods of assessment can complement objective and standardized methods. Although it is true that many counselors do use standardized tests and clients use self-interpretative methods, the counselor's involvement and detailed follow up and follow through will bring the process into active and intimate contact.

Computerised testing has largely captured users' attention and emerged as an good alternative to conventionally administered tests. The digitised VIQ has a number of strengths and even with the benefits, several key characteristics that contribute to a sound test should be considered. Walsh's (1990) assertion seems as relevant here. He pointed out, 'computerised testing must be considered in interaction with the counselor. It is not restricted to objective, standardized, quantifiable procedures. The role of the counselor is crucial and integral to the process of assessment. The assessor has the responsibility of applying and maintaining austute judgment. Computer testing does not have the answers. The relationship between computer-assisted testing and the goals of counseling must always be addressed. This relationship provides a focal attention on what is the appropriate test suitable for a particular need unique to the client. It is heartening to note increased attention given to etic (cross-cultural) versus emic (culture-specific) testing considerations (Lonner, 1985).

Duckworth (1990) reminded us that tests are used to teach clients important information about themselves, which can be of value for personal growth purposes. In doing this, the counselor actively involves clients in the assessment process and

acts as both teacher and facilitator during the experience. At this point, even if a test has all the necessary key attributes of a sound test, a test should be regarded as only a means to an end and the role of the test-taker can benefit most from a collaborative counseling relationship when considering computerised testing. Interactive devices and digitised qualitative methods can help increase this therapeutic alliance. In concluding, some key reminders to all test users from the counseling psychological approach to testing from Duckworth (1990):

- Testing is done for the benefit of the client and the therapist, not for the therapist's benefit only.
- As a corollary to point one, testing is done to generate information for both the client and counselor, not just for the counselor.
- The client needs to be an active participant in the testing process even to the extent, at times, of helping to select the tests to be used.
- The client is assumed to be someone who can profit from the testing process if given appropriate feedback concerning the test results. Therefore, the client is to receive an interpretation of the test results if this is at all possible. The interpretation should be done without using psychological jargon, and non-pejorative descriptions of behavior and feeling should be used.
- The world of work is an important part of people's lives; therefore, vocational tests are an important component of assessment.
- Finally, the goal of testing is empowerment of the individual, empowerment so that the individual can be more fulfilled through increased knowledge and skills.

By becoming aware of some of the key characteristics of a sound test, hence, what to evaluate, test users can profit with more knowledge, attention and ease. Test users will know exactly what a test can and cannot do. With this greater attention and sensitivity, test users can seek out an appropriate test with greater ease and confidence.