Analysis of validity and reliability of intelligence assessments

Measuring intelligence has always been a pragmatic thing to do (Wilhelm & Engle, 2005). Most people find it hard to be rational in measuring their intelligence. As a common usage, intelligence is attested quantifiably by understanding the ' admissibility of a degree' (Bartholomew, 2004) In other words, there can be no way that intelligence would be perfectly quantified. Only the nearness or the relative distance can be the mode of measuring one's intelligence. David Wechsler (Bartholomew, 2004) defines intelligence as the aggregate ability of an individual to have a rational thinking, a purposive act, and an effective dealing with his environment. Some writers however define intelligence negatively. That is, the " doing away of tests." It is important to differentiate between intelligence quotient (IQ) and intelligence itself. IQ does not simply refer to the quantity of intelligence a person has, rather it refers to the relative standing of an individual with regard to his performance in an intelligence test as compared to others' who took the same test (Zastrow & Kirst-Ashman, 2007). When intelligence is measured in terms of a numerical result (i. e. I. Q. is 140), we cannot be sure that we are really talking about intelligence since there is a substitution of precision for vagueness (Govier, 2009). The original purpose of intelligence test was primarily to identify the least capable students that are finding it hard to learn from ordinary schooling (Kalat, 2010). However, it was later used to determine who among the students excel and during entrance examinations in universities.

It is seemingly difficult if not impossible to devise a test that would truly measure the innate intelligence without being bias culturally. Innate ability cannot be measured if cultural bias exists (Young, 2006). Bias refers to the

existence of nuisance factors that makes it difficult to compare the intelligence tests results across different cultural groups (Prifitera et al. 2008). There are three kinds of cultural biases in intelligence tests vis-a-viz construct bias, method bias, and differential item functioning (Lovler, Miller, & McIntire, 2010). Construct bias occurs when there are different systems of meaning from culture to culture. For instance, a good daughter in a country might have a different characteristic in another country. Method bias occurs when the procedure of the test vary from different cultures such as when in a certain country, participants are used to input data electronically while in another country, participants are used to answering manually through papers. Differential item functioning or item bias occurs when there is a great gap in the test scores of the participants in different cultures who have relatively the same abilities. For example, different regional groups might have different scores in a history test regarding their home country even if all of them are familiar with its history.

In order to eliminate cultural biases in intelligence tests, culture-free tests became a demand. As the name imply this test attempt to eliminate cultural and educational differences (Haselbauer, 2006). The most common mediums used in this test are pictures or images which assess spatial capabilities of the participants such as visualization and perception. The validity of the tests is equally important as the reliability of the test. Validity is challenged when a participant in the test cannot communicate in the language used in the test (Rhodes, 2005). Critics of IQ tests argue that the differences in IQ scores are attributable to the orientation of the test makers such as their experiences, vocabulary and language, and lifestyles (Weiten, 2008).

Construct-related validity refers to the extent which the test instrument can capture the details it is designed to measure (Sternberg, 2010).

Methods

To be able to determine the validity of intelligence tests particularly on the way it is constructed, this paper will investigate whether validity of tests vary between English speakers and non-English speakers. The kind of test that will be used is PSYGAT which focuses mainly on the verbal capability of the participant. This will be conducted after the participants have taken the Queendom test and Culture Fair IQ test. Participants will be composed of a total number of 337 from which 269 (79. 8%) are females and 68 (20. 2%) are males. The sex standard deviation for the participants with English-speaking backgrounds (ESB) will be 0. 407 while those with non-English speaking background (NESB) will be 0. 389. After getting the result of the test, it will be analyzed based on age. The age standard deviation for ESBs will be 7. 514 while for NESBs it is 7. 627. Ages range from 19-55 for ESBs and 19-62 for NESBs.

Aim

This study aims to determine if PSYGAT is internally reliable when a group of participants with English-speaking backgrounds and non-English speaking backgrounds are compared. An initial hypothesis for this study is that PSYGAT will be a reliable means of intelligence computation vis-a-viz significant convergent validity in the same way as the Queendom and Culture Fair IQ tests. However, between ESB and NESB, it is hypothesized that they will differ in terms of reliability and validity.

Results

Item Analysis

Upon checking for internal consistency of data using Cronbach's alpha, it can be assessed that results were reliable, especially as most of the generated alpha were higher than 80 percent. The following are reliability statistics with their corresponding Cronbach's alpha[1]:

Reliability of data for overall sample for all 55 questions

Reliability Statistics
Cronbach's Alpha

. 898

Reliability of data for overall sample for the 25 " best" discriminating questions

Reliability Statistics
Cronbach's Alpha

. 848

Reliability of the data for ESB group for the 25 " best" discriminating questions

Reliability Statistics
Cronbach's Alpha

. 851

Reliability of the ESB for all 55 questions

Reliability Statistics

Cronbach's Alpha

. 901

Reliability of the data for NESB group for the 25 " best" discriminating

questions

Reliability Statistics

Cronbach's Alpha

. 841

Reliability of the NESB group for all 55 questions

Reliability Statistics

Cronbach's Alpha

. 889

Validity of PSYGAT

In order to test the validity of PSYGAT, the correlation scores of the PSYGAT

(total scores) on each of the Queendom tests were obtained using SPSS.

Correlations (for all sample)

Total

Queendom Verbal Adjusted

Total

Pearson Correlation

1

. 466**

Sig. (2-tailed)

. 000

N

337

337

Queendom Verbal Adjusted

Pearson Correlation

. 466**

1

Sig. (2-tailed)

. 000

N

337

337

Queendom Culture Fair IQ

Pearson Correlation

. 098

-. 283**

Sig. (2-tailed)

. 073

. 000

N

337

337

**. Correlation is significant at the 0. 01 level (2-tailed).

Correlations for ESB Group

Total

Queendom Verbal Adjusted

Total

Pearson Correlation

1

. 433**

Sig. (2-tailed)

. 000

N

244

244

Queendom Verbal Adjusted

Pearson Correlation

. 433**

1

Sig. (2-tailed)

. 000

N

244

244

Queendom Culture Fair IQ

Pearson Correlation

. 067

-. 341**

Sig. (2-tailed)

. 295

. 000

N

244

244

**. Correlation is significant at the 0. 01 level (2-tailed).

Correlations for NESB group

Total

Queendom Verbal Adjusted

Total

Pearson Correlation

1

. 567**

Sig. (2-tailed)

. 000

N

93

93

Queendom Verbal Adjusted

Pearson Correlation

. 567**

1

Sig. (2-tailed)

. 000

N

93

93

Queendom Culture Fair IQ

Pearson Correlation

. 238*

. 011

Sig. (2-tailed)

. 022

. 918

N

93

93

**. Correlation is significant at the 0. 01 level (2-tailed).

*. Correlation is significant at the 0. 05 level (2-tailed).

Assessment of Differences between ESB and NESB groups in terms of

reliability and validity

By splitting the data according to its background (i. e., ESB and NESB group),

the following correlations were obtained:

Correlations

English

Queendom Culture Fair IQ

english speakiing background

Queendom Culture Fair IQ

Pearson Correlation

1

Sig. (2-tailed)

N

244

Queendom Verbal Adjusted

Pearson Correlation

-. 341**

Sig. (2-tailed)

. 000

N

244

non english speaking background

Queendom Culture Fair IQ

Pearson Correlation

1

Sig. (2-tailed)

N

93

Queendom Verbal Adjusted

Pearson Correlation

. 011

Sig. (2-tailed)

. 918

N

93

**. Correlation is significant at the 0. 01 level (2-tailed).

The observed z value will be obtained using the following formula:

where:

rESB= -0. 341 NESB = 244

rNESB= 0. 011 NNESB = 93

Transforming the correlation coefficient to their corresponding z values,

zESB = -0. 355 zNESB = 0. 011

Using the values above, the following observed z value was obtained:

$Z = -2.96$

To evaluate the observed z value, the obtained value must not be between -1. 96 and +1. 96 to conclude a statistical difference between the two correlation coefficients. From this, it can be assessed that there is a statistical difference between the 2 correlation coefficients, due to the fact that the observed z value does not fall between -1. 96 and +1. 96.

Discussion and Conclusions

Many researchers have argued that the measurement of intelligence through IQ tests is no doubt based on the culture to which the individual belongs to. In an attempt to assess the reliability and the construct validity of the PSYGAT — a verbal assessment developed by Psychology students in 1998 — 344 Psychology students were tasked to complete three tests which comprises of two Queendom tests (verbal adjusted and culture fair test) and the PSYGAT. The results of the PSYGAT were then analysed for internal validity. Construst validity of PSYGAT was also examined by analyzing the test scores alongside the Queendom tests. The hypothesis that the PSYGAT test would show significant internal reliability and construct validity as it was analysed alongside the other two tests was supported. Significant correlations were also found in the results for Queendom Verbal Adjusted and the PSYGAT for English speaking background group. Upon computation of observed z values, it was also concluded that there were statistical differences between the ESB and NESB correlation coefficients. Thus, it can be further assessed that PSYGAT can be used validly for verbal intelligence

assessments, although, further research is recommended for its use in culture fair factors.

One limitation though of this study is the fact that participants were all university students, specifically Psychology students. Perhaps, there is a need for further research that will involve a much diverse population, considering that participants in this study can be considered as well-educated. Despite their non-English speaking background, it can also be assumed that perhaps are also well-versed in the English language.

Another limitation is also the fact that the tests used the English language as the first language in the assessments. Such factor may pose limitations for specific ethnic groups whose first language is not English. It can be considered that future researches be conducted that will be more culture sensitive and be conducted using the first language of the specific ethnic group.

References

Bartholomew, D. J. (2004). Measuring intelligence: facts and fallacies: Cambridge University Press

Govier, T. (2009). A practical study of argument (7th ed.): Cengage Learning.

Haselbauer, N. (2006). The everything test your IQ book: discover your true intelligence: Everything Books.

Kalat, J. W. (2010). Introduction to psychology (9th ed.): Cengage Learning.

Lovler, R. L., Miller, L. A., & McIntire, S. A. (2010). Foundations of psychological testing: a practical approach (3rd ed.): SAGE.

Prifitera, A., Saklofske, D. H., & Weiss, L. G. (2008). WISC-IV clinical assessment and intervention: practical resources for the mental health professional (2nd ed.): Elsevier.

Rhodes, R. L., Ochoa, S. H. & Ortiz, S. O. (2005). Assessing culturally and linguistically diverse students: a practical guide: Guilford Press.

Weiten, W. (2008). Psychology: Themes and variations (8th ed.): Cengage Learning.

Wilhelm, O. & Engle, R. W. (2005). Handbook of understanding and measuring intelligence: SAGE.

Young, E. M. (2010). Dealing with the cultural bias in intelligence testing: culture free and culture fair iq tests. Retrieved 29 April 2011 from: .

Zastrow, C & Kirst-Ashman, K. K. (2007). Understanding human behavior and the social environment (7th ed.): Cengage Learning.