

## **The Validity of Human Figure Drawings as Measures of Intellectual Maturity in Malaysian Children**

Alex Lui An Lieh  
International Islamic University Malaysia

### **Introduction**

The exclusive study of drawings in children was recorded as far back as 1886 when Ebenezer Cooke first published a paper solely on children's artwork entitled 'Our art teaching and child nature' (Cox, 1993). Even earlier in 1877, Charles Darwin had already described his observations and investigations of children's drawings in an article, "A biographical sketch of an infant" which was published in the journal *Mind*.

The human figure is one of the first topics to be drawn by many young children and also remains one of the most frequently chosen topics. According to Pikunas and Carberry (cited in Koppitz, 1968), when given the opportunity, children would spontaneously draw pictures of animals, houses, cars, boats, planes, flowers and all sorts of other things, but most often, they would draw human beings. One of the earliest study on human figure drawings was conducted by Corrado Ricci who wrote the famous monograph on 'L'arte dei Bambini' in 1887 (Cox, 1993). Ricci was said to have been stirred by graffiti drawings he saw in a doorway as he was sheltering from the rain. He was particularly struck by what seemed to be developmental differences between the lower figures, presumably drawn by younger children, and the upper ones drawn by older children.

Since then, the study of human figure drawings has spawned numerous research papers in various journals on child psychology and development. Early investigators of human figure drawings by children had already begun noticing that with a child's increasing chronological age, the number of body parts in the figure increased and the proportions of the body parts became more realistic. However, it was not until 1926 that Florence Goodenough

developed a standardised method of quantifying human figure drawings in the Draw-A-Man test as a measure of intellectual maturity.

Goodenough selected 51 details of the human figure, such as number of body parts, their proportions and the way the parts were attached to the main figure, to be credited with points. The accumulated points from the drawing would then be converted to their respective IQ scores.

Later in 1963, Harris refined the Goodenough test by adding a drawing of a woman and a drawing of the self. He also provided a nationally representative norm group that included a wider age range (i.e. ages 3 through 15) and replaced the ratio IQ by Goodenough with a deviation IQ with a mean of 100 and a standard deviation of 15. A number of researches on the revised version, which came to be known as the Goodenough-Harris Drawing Test, was shown to correlate with intelligence tests such as the Wechsler and Binet scales (cited in Naglieri, 1988) and have good inter- and intrarater reliability.

Finally in 1988, Jack A. Naglieri developed the Draw-A-Person: A Quantitative Scoring System to "meet the need for a modernized, recently normed, and objective scoring system to be applied to human figure drawings produced by children and adolescent" (Naglieri, 1988). The Naglieri scoring system was devised to reduce the influence of contemporary styles of dress and to answer the call for objectivity in scoring human figure drawings.

### **Justification of the Problem**

This research is aimed at finding out the viability of using human figure drawings as an estimate of children's intelligence and developmental level in Malaysian children.

One of the most frequently used instruments in measuring children's intelligence in Malaysia is the WISC-111. While the WISC-111 may be the most popular intelligence scale today owing to its long clinical and research history that supports interpretation by clinicians, it may not be entirely appropriate to be used in a

society whereby English is not the native language spoken. The verbal subtests are particularly problematic since proficiency in English and knowledge of the British or American culture, depending on whether the UK or US version is being used, are necessary prerequisites in responding correctly to the test items.

Even the use of an unauthorised translated version, as is often the practice, does not solve the problem of culture bias. Besides rendering the population norms invalid, some words simply could not be translated or they could be translated but lose their evaluative properties. For example, in the WISC-111 Vocabulary subtest, the word breakfast may be translated into Malay as "sarapan pagi." However, the word "sarapan" and "pagi" itself is already a giveaway that breakfast is suppose to be a "meal" taken in the "morning."

Therefore the use of nonverbal measures of intelligence as an alternative option would be very much welcomed in assessing the multicultural children's population in Malaysia. The use of the Draw-A-Person (DAP) has some advantage over the WISC-111 and other nonverbal measures of intelligence for children.

Firstly, the DAP takes a shorter time for administration and it is especially useful for children who are resistant to more traditional examination procedures. A typical WISC-111 testing duration may be anywhere from one to two hours. On the other hand, the DAP only takes about 15 minutes to administer. For this reason, the DAP is particularly helpful in testing children with very short attention span such as those with attention deficit and hyperactivity disorder or those who simply lack the motivation for any longer intelligence test.

Secondly, the DAP requires less verbal ability in comprehending and almost no verbal ability in responding to the test items. Drawing is already a natural activity of almost every child and hence not much instruction needs to be given for the child to draw. As mentioned by Kaufman (cited in Kamphaus, 1993), the verbal instructions given by the WISC-111 are simply too long, and they require too much verbal comprehension on the part of the child. This makes the test, including the supposedly nonverbal Performance subtests, more difficult to administer to limited English-proficient and hearing impaired children.

Thirdly, the DAP may be the only viable means of estimating the intelligence in children with certain disabilities and chronic conditions such as mental retardation or auditory impairment. Otherwise, it may be used as part of an intelligence test battery to provide reliable information about an individual's intellectual status.

Finally, unlike most individually administered intelligence tests, the DAP may be easily used in group settings for individual assessment or as a screening test. As an additional benefit, it may also double as a projective personality test by interpreting the drawings using other evaluative systems such as the Draw A Person: Screening Procedure for Emotional Disturbance.

### **Research Objectives**

1. To determine the validity of the DAP as a measure of children's intelligence in Malaysia.
2. To determine the reliability of the DAP as a measure of children's intelligence in Malaysia.
3. To decide if the concurrent use of the DAP and the Standard Progressive Matrices as a nonverbal intelligence test battery may subsequently increase the predictive validity of the tests in regard to academic performance.
4. To identify any cultural and gender bias that may be prevalent in the DAR
5. To identify other aspects of the drawings that may be used to increase the validity of the test in measuring intelligence.

### **Important Definitions**

*Intellectual maturity* refers to the intellectual ability of a child at a certain age as measured by relevant IQ tests.

*Academic performance* refers to the scholastic achievement of a child as reflected by his or her school examination scores.

## **Relevant Research**

### **Theoretical foundation**

An early work by Piaget and Inhelder (1969) on child development is consistent in providing the theoretical foundation in the use of human figure drawings as an indicator of intellectual maturity. Piaget and Inhelder emphasised that children's drawings progressed from what the child knows to what he or she sees.

The first stage in this progress, around two years of age, is defined as "fortuitous realism," in that the child develops what is drawn in the act of drawing it. A year or two later, the child draws recognisable elements resulting in the "tadpole" person. This is the stage of "failed realism" in that only selected elements (e.g. head and legs) are included,

Around age 5 years or so, drawings reflect "intellectual realism" in that the drawing is conceptually complete, but fails to take into account visual perspective or proportionality. It is only around age 8 or 9 that the child's drawing reflects "visual realism," accurately conveying a visual perspective and proportional dimensions. These stages, as reflected in the drawings of human figures, may therefore be assumed to indicate the maturity level of a child.

### **Validity**

The Draw-A-Person: A Quantitative System manual by Naglieri (1988) is an important source in providing a comprehensive description of the statistical properties of the test. The standardization sample of the DAP conducted by Naglieri consisted of 2,622 children between the ages of 5 and 17 years. It was a stratified sample based on the 1980 U.S. Census Bureau statistics and used the stratification variables of age, gender, race, geographic region, ethnic group, socioeconomic status, and community size.

According to the manual, the mean raw scores of the DAP drawings and mean for the Total raw score show a clear increase

with age. Naglieri (1988) found significant correlations ( $p < 0.01$ ) between age and the Man ( $r = 0.58$ ), Woman ( $r = 0.62$ ), Self ( $r = 0.60$ ) and Total ( $r = 0.64$ ) for children aged 5 through 11. This seems to indicate some evidence of construct validity.

In terms of criterion-related validity, Naglieri (1988) found significant correlation between the DAP standard scores and the Matrix Analogies Test-Short Form (MAT-SF) ( $r = 0.31$  for Grades K to 3 &  $r = 0.27$  for Grades 4 to 12,  $p < 0.01$ ), and the Multilevel Academic Survey Test (MAST) ( $r = 0.24$  for reading &  $r = 0.21$  for maths,  $p < 0.01$ ).

The MAT-SF is a test of nonverbal ability which uses abstract figural analogies of the progressive matrix type while the MAST is a test of achievement consisting of Reading and Mathematics sections. In another research, Wisniewski and Naglieri (cited in Kamphaus, 1993) found a correlation of 0.51 between the DAP Total Test and the WISC-R Full Scale for a sample of 51 children who ranged from 6 to 16 years. Naglieri (1988) also cited a couple of other researches that found the Goodenough-Harris scoring system to correlate with intelligence tests such as the Wechsler and Binet scales.

Most studies made on the validity of the DAP have not been very encouraging. Neisworth and Butlers (cited in Kamphaus, 1993) concluded that the construct validity of the DAP is arguable. Data regarding the relationship of the DAP to an academic achievement measure show low correlation between the DAP and scores for reading (0.24) and mathematics (0.21) for a sample of 1,328 nonhandicapped children. Aikman, Belter and Finch (1992) found significant but low correlations between scores of human figure drawings using the Goodenough-Harris scoring system and both Full Scale IQs and academic achievement. They concluded that human figure drawings should not be substituted for other well-established intelligence and achievement tests nor used as an additional measure when evaluating psychiatric patients.

Some researches have also suggested that the human figure drawings may be able to differentiate between children in the lower age group but not in the higher age group. Ohuche and Ohuche (1973) found that the Goodenough Draw-a-Man test was able to predict academic achievement in Sierra Leonean children in the first

three years of primary school but not in grades above that. In another study conducted by Harris and Pinder (1977), the Goodenough-Harris Drawing Test functioned as a general ability measure in the earlier years of the adolescent period but after about the age of 15, it discriminated only in the lower reaches of the ability distribution. In their research with human figure drawings among South African Black children, Richter, Griesel and Wortley (1989) cited that the Draw-a-Man test showed some relationship with scholastic achievement but it seems to be unsuitable for children over 8 years of age because it underestimates abilities.

There have also been other instances in which researches showed positive correlation between IQ scores obtained from the DAP and other measures of intelligence. For example, Abell, Horkheimer and Nguyen (1998) reported that both the Buck and the Goodenough-Harris scoring system for human figure drawings were both positively and significantly related to IQ scores on the WISC-R. The Buck system of scoring was also said to have less tendency to underestimate IQ scores and therefore hold greater promise for the intellectual assessment of adolescents with human figure drawings. Sinha (1977), however, did not find any significant correlation between the DAP and intelligence, but the DAP was found to correlate significantly with school achievement, particularly reading.

### **Reliability**

According to Kamphaus (1993), the reliability coefficients for the DAP were found to be generally quite high "taking into consideration the brevity of the measure." In terms of test-retest reliability, Naglieri (1988) cited the stability coefficients for the DAP Total score to range from a low 0.60 to a high 0.89 with a median of 0.86. The mean test-retest reliability coefficients for the Man, Woman, and Self scores are 0.70, 0.65, and 0.58, respectively.

The internal consistency coefficients for the drawings of the man, woman, and self scored separately ranges from 0.56 to 0.78 with a median of 0.70. The interrater reliability of the DAP is good, ranging from 0.93 to 0.95.

## **Gender bias**

Naglieri (1988) found a small but statistically significant overall gender difference in the DAP scores with females scoring higher than males. In spite of this, he suggested that since "the difference is only about one-fifth of a standard deviation, it has little practical significance."

Similarly, Koppitz (1968) observed some minor but consistent differences between the occurrence of what she termed as Developmental Items on the human figure drawings of boys and girls. She noted that her findings were "in accord with the observation of Goodenough, Harris and Machover, all of whom emphasize that drawings of girls in the primary grades are superior to those of boys." However, she also mentioned that there has been a consensus that the difference between genders diminishes gradually. According to her, "by age eight or nine, boys not only catch up with girls but often surpass them in the quality and details of their drawings."

## **Cultural bias**

Although Sattler (cited in Naglieri, 1988) claimed that tests using human figure drawings may not be as culturally loaded as are other tests, various researches have lately confirmed that there are elements of cultural differences in human figure drawings. Cox (1993) reported that using scoring systems as that devised by Goodenough and Harris, many researchers found differences in the human figure drawings among a range of different cultural groups. Havighurst, Gunther, and Pratt (cited in Cox, 1993), for example, found that 6- to 11- year-old American-Indian children in six different tribes had higher scores than did white children. They explained that in these Indian groups, art had a high status among the adults and the American-Indian children practised drawings more than did white children.

Cox (1993) stated some evidences that "lower-status groups which do not have a particular artistic tradition score low in their drawing performance compared with children of the wider culture."



She mentioned that in Muslim societies, "although there is a strong tradition of abstract, decorative art, representational art is perhaps relatively less important, and if one were to make predictions on this basis alone one would expect Moslem children to perform relatively low on human figure drawing tasks." However, Sinha (cited in Cox, 1993) found that immigrant Indian and Pakistani children in the UK actually scored higher than their white counterparts on their drawings of the opposite sex (although strangely not on their same-sex drawings). She explained that the general expectation of high performance in the children's school work from their immigrant parents may explain the relatively higher scores.

### **The Present Study**

#### *Research Hypotheses*

##### **Construct Validity**

1. There would be a significant positive correlation between scores obtained from the DAP and age.

##### **Concurrent Validity**

2. There would be a significant positive correlation between scores obtained from different methods of scoring the DAP.

##### **Criterion-Related Validity**

3. There would be a significant positive correlation between scores obtained from the DAP and the Raven's Standard Progressive Matrices (SPM).
4. There would be a significant positive correlation between scores obtained from the DAP and academic performance as indicated by the school midterm examination results.

##### **Discriminant Validity**

5. There would be no correlation between scores obtained from the DAP and scores of aesthetic quality of the drawings.

### **Predictive Validity**

6. There would be an increase in correlation between a combined DAP and SPM scores and academic performance (i.e., as indicated by the school midterm examination results), as compared to when independent scores were used from the DAP and the SPM to predict academic performance.

### **Internal Consistency**

7. There would be good internal consistency of scores across the 14 scoring criteria and the three figures drawn as indicated by the Cronbach alpha.

### **Interrater Reliability**

8. There would be a significant positive correlation between the scores obtained from different scorers.

### **Test-retest Reliability**

9. There would be a significant positive correlation between both the first and second test scores.

### **A nalysis of Gender and Race Differences**

10. There would be no significant difference between scores of male and female participants.
11. There would be no significant difference of scores between races.

## **Method**

### ***Participants***

Not less than 200 primary school pupils between the ages of six to twelve years old will be recruited from a national school as participants in this research. One class from each of the six standards will be selected. Since the targeted school has only between two to three classes per standard, there should be enough

variability of intellectual ability among the children within each class group. Each age group ought to be equally represented since most typical classrooms in a national school usually have somewhere between 30 and 40 students.

## **Measurements**

### *Draw-A-Person (DAP)*

The DAP is a nonverbal measure of intellectual maturity based on drawings of human figure (Naglieri, 1988). During the test, the child is asked to draw three different figures: that of a man, a woman and him- or herself. There are various methods of scoring the drawings. Based on Naglieri's Quantitative Scoring System, intelligence is evaluated by analysing fourteen different aspects of the drawings, such as specific body parts and clothing, for various criteria, including presence or absence, detail, and proportion. There are 64 scoring items for each drawing in all. A separate standard score is recorded for each drawing, and all three scores are totalled up. The raw scores are then converted into standard scores based on the norms provided in the manual. Two other popular scoring systems are the Goodenough-Harris DAP Test, which has 73 scoring items for the male figure and 71 for the female, and the Koppitz Human Figure Drawing, which scores on 30 Developmental Items.

### *Raven's Standard Progressive Matrices (SPM)*

The Raven's Standard Progressive Matrices is a nonverbal measure of intelligence containing logical patterns or designs with a missing part whereby an appropriate design must be selected that best fit that missing part. It consists of 60 problems in the form of diagrammatic puzzles exhibiting serial change in two dimensions simultaneously. It was "originally designed to cover the widest possible range of mental ability and to be equally useful with persons of all ages, whatever their education, nationality, or physical condition" (Raven, Court, & Raven, 1992).

## Procedure

The tests will be conducted in the children's respective classes during one of their school class periods, preferably during art in order to minimise disruption in their regular lessons. The tests are scheduled to commence after the UPSR (Primary School Assessment Test) in order to accommodate the standard six pupils.

The children will first be required to fill in their names, sex, class and date of birth in the response forms and answer a simple questionnaire. Questions about ambition, hobby and fondness of drawing convey a general idea of the children's personalities, while the rest of the questions are meant as rough estimates of their socioeconomic status.

They will then proceed to draw a picture of a man, a woman, and lastly of themselves, as prescribed by the DAR A Quantitative Scoring System Manual. They will be given a maximum time of 5 minutes for each drawing. Finally, they will be instructed to respond to the Raven's Standard Progressive Matrices. The entire testing session is estimated to take up approximately 1 hour.

A sub-sample of 60 children (i.e. five boys and five girls from each standard) will be randomly selected and a retest will be conducted two weeks after the initial testing in order to measure the stability of the DAP scores. Only the DAP will be administered for the retest. The children's drawings will then be scored independently by at least three graduate psychology students who will have to first pass the scoring competency tests. The drawings will be scored according to the DAR A Quantitative Scoring System, the Goodenough-Harris DAP (if available) and the Koppitz Human Figure Drawing.

The drawings will also be scored for their aesthetic qualities on a scale of 0 to 10 by at least three independent art enthusiasts.

A series of statistical test will then be performed in order to find out if there are any significant correlations between the DAP IQ scores, the Standard Progressive Matrices IQ scores, age, academic performance and other independent variables. Statistical tests will also be performed to reveal any significant difference of

IQ scores for the gender and race variables as an indication of gender and cultural bias.

### **Data Analysis Plan**

1. The Pearson product moment correlation will be used for measures of validity and reliability.
2. The Cronbach's alpha will be used for measuring internal consistency.
3. The t-tests and the F- tests will be conducted for the analyses of gender and race differences respectively.

### **References**

- Abell, S. G, Horkheimer, R., & Nguyen, S. E. (1998). Intellectual evaluations of adolescents via human figure drawings: An empirical comparison of two methods. **Journal of Clinical Psychology, 54** (6), 811-815.
- Aikman, K. G., Belter, R. W, & Finch, A. J. (1992). Human figure drawings: Validity in assessing intellectual level and academic achievement. **Journal of Clinical Psychology, 4** (10) , 114-120.
- Cox, M. V. (1993). **Children's drawings of the human figure**. East Sussex: Lawrence Erlbaum Associates.
- Hards, D. B., & Pinder, G. D. (1977). Goodenough-Hards test estimates of intellectual maturity of youths 12-17 years: Demographic and socioeconomic factors. **Vital & Health Statistics, 159** (L1), 43. Retrieved April 22, 2004, from <http://biblioline.nise.com>
- Kamphaus, R. W. (1993). **Clinical assessment of children's intelligence: A handbook for professional practice**. MA: Allyn & Bacon.
- Koppitz, E. M. (1968). **Psychological evaluation of children's human figure drawing** . MA: Allyn & Bacon.
- Naglieri, J. A. (1988). **Draw-a-PersOn: A Quantitative System**. New York: The Psychological Corporation.

- Ohuche, N. M. & Ohuche, R. O. (1973). The Draw-a-Man Test as a predictor of academic achievement. **West African Journal of Educational & Vocational Measurement**, (L1), 20-27. Retrieved April 22, 2004, from <http://bibliolinc.nisc.com>
- Piaget, J., & Inhelder, B. (1969). **The psychology of the child**. New York: Basic Books.
- Raven, J. C., Court, J. H., & Raven, J. (1992). **Manual for Raven's Progressive Matrices and Vocabulary Scales: Standard Progressive Matrices 1992 Edition**. Oxford: Oxford Psychologists Press.
- Richter, L. M, Griesel, R. D, & Wortley, M. E. (1989). The Draw-a-Man test: A 50-year perspective on drawings done by Black South African children. **South African Journal of Psychology**, 19 (1), 1-5. Retrieved April 22, 2004, from <http://biblioline.nisc.com>
- Sinha, M. (1977). **What the Draw-Yourself Test measures**. **Journal of Psychological Researches**, 21(2), 128-131. Retrieved April 22, 2004, from <http://bibliolinc.nisc.com>