

The Score Reliability Of Draw-A-Person Intellectual Ability Test (DAP: IQ) For Rural Malawi Students

Denis S. Khasu, Ph.D., Domasi College of Education, Malawi, Africa

Thomas O. Williams, Jr., Ph.D., Virginia Tech, USA

ABSTRACT

In this brief article, the reliability of scores for the Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults (DAP: IQ; Reynolds & Hickman, 2004) was examined through several analyses with a sample of 147 children from rural Malawi, Africa using a Chichewa translation of instructions. Cronbach alpha coefficients for the 23 test items were calculated for the total sample, the six age groups represented in the sample, and gender. The interscorer reliability of test scores was also calculated. The obtained alpha coefficients for the 23 items for total sample (.81), the six age groups represented (.68 - .92), and gender (male .79, female .83) were comparable to those listed in the examiner's manual. The coefficient for interscorer reliability was .85.

Keywords: Draw-A-Person Intellectual Ability Test; DAP: IQ; Malawi; Score Reliability

INTRODUCTION

Early records of using human figure drawings (HFDs) as measures of cognitive ability have a rich history and go back to the 1800s (Barnes, 1893; Ricci, 1894). A chronological listing of the more well-known tests that purport to measure cognitive ability through HFDs, include the *Draw A Man Test* (Goodenough, 1926), *Goodenough-Harris Drawing Test* (Harris, 1963), Koppitz (1968) scoring system, *Draw A Person: A Quantitative Scoring System* (Naglieri, 1988), and *Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults* (DAP: IQ; Reynolds & Hickman, 2004). Although these tests differ in their scoring systems and types of scores produced, a feature common to these tests is that they all employ one or more HFDs to assess cognitive ability.

The most recently developed HFD test for intellectual ability is the DAP: IQ. It was designed to represent a standardized and objective scoring system from which IQ could be derived from the drawing of a single human figure. With regards to reliability, the examiner's manual reported that the DAP: IQ produced test scores that were reliable. However, evidence for score reliability concerning the DAP: IQ presented in published studies is sparse and inconsistent.

Imuta, Scarf, Pharo, & Hayne (2013) reported that the reliability of the DAP: IQ was acceptable and well established. However, only two published studies were found that specifically examined the reliability of the DAP: IQ scores: Williams et al. (2006) and Honeres and Merino (2011). Williams et al. reported an alpha coefficient of .82 for a sample of 110 college students ranging in age from 19-to-29 years from the USA. Honeres and Merino reported a mean alpha coefficient of .68 for students six-to-11 years of age from Peru with instructions translated to Spanish. In their study, the DAP: IQ did not produce scores that demonstrated a high level of reliability with a Spanish translation for young children. This was surprising because, as Imuta et al. stated in their comprehensive review, HFDs generally produced scores that are reliable.

In light of these findings, this study examined whether the DAP: IQ would produce scores that were reliable when instructions were translated into a language other than English. Specifically, this study examined the reliability of scores from the DAP: IQ with instructions translated in Chichewa from a sample of rural Malawi elementary school-

age children through several analyses. The first analysis examined the reliability of scores for the 23 test items for the total sample. The second analysis examined reliability of scores for the test items across six age groups represented in the sample. The third analysis examined the reliability of scores across gender. The fourth analysis examined interscorer reliability for a random sample of 50 participants taken from the total sample.

METHOD

Participants

Participants represented two classrooms that were composed of 147 first and second grade students from a rural school in Malawi, Africa. The ethnic groups were primarily Chewa and Yao, though multi-ethnicity was high. The class size for the two classrooms was 90 students for the first grade classroom and 57 students for the second grade classroom. It should be noted that it is not uncommon to have large number of students per class and extended age ranges in these rural schools. Fifty-two percent ($n = 77$) of the participants were male and 48 percent ($n = 70$) were female. The age for the participants ranged from five to 10 years ($M = 6.93$; $SD = 1.37$).

Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults

The DAP: IQ is a screening test designed for children through adults ranging in age from four-to-89 years of age to estimate IQ. It can be individually or group administered and scored by individuals who have had formal comprehensive training in assessment. To administer the DAP: IQ, the examinee is asked to draw a picture of him- or herself when provided with the standardized instructions in the examiner's manual. If the examinee draws a side view or only a head, the directions can be repeated and the student can draw another figure. The drawing session is not timed and it is estimated that approximately eight-to-15 minutes are required to administer and to obtain a score for the examinee.

Instruction and procedures for both individual and group instructions are provided. The English instructions for group administration (which were translated to Chichewa in this study) were:

I want you to draw a picture of yourself. Be sure to draw your whole body, not just your head, and draw how you look from the front, not from the side. Do not draw a cartoon or stick figure. Draw the very best picture of yourself that you can. Take your time and work carefully. Now go ahead. Raise your hand if you have a question or when you have completed your drawing. (Reynolds & Hickman, 2004; p. 6).

The DAP: IQ provides a common set of scoring criteria to evaluate the examinee's drawing on 23 items. The 23 items identified for scoring are the head, hair, eyes, eyelashes, eyebrows, nose, mouth, chin, ears, neck, shoulders, arms, elbows, hands, torso, waist, hips, legs, knees, ankles, feet, clothing, and accessories. Scores can range from zero to four points, depending upon the item being scored. A pictorial as well as a verbal description for each point value criterion is presented on the scoring protocol. To obtain the standardized IQ score, the 23 items scores are summed into a raw score. The raw score is then converted into a standardized IQ score ($M = 100$, $SD = 15$) using age-based norms. Scores can be converted to percentile ranks, T scores, z scores, and stanines. Age and grade equivalent conversion tables are also provided.

Reliability

Reynolds and Hickman (2004) reported alpha coefficients for 22 selected age intervals for the entire normative sample ($N = 2,295$). The median alpha coefficient was .82 and ranged from .74 to .87 for the selected age intervals. The alpha coefficients for white, African American, Hispanic, and for those participants who listed "Other" for race, ranged from .73 to .80. Mean alpha coefficients were also reported for left-handed individuals (.80) and right-handed individuals (.86). The mean coefficient alpha for males and females was reported as .80.

Published studies of score reliability for the DAP: IQ includes Williams et al. (2006), Honores and Merino (2011) and Imuta et al. (2013). Williams et al. reported alpha coefficient of .82 for a sample of 110 college students from the USA. Honores and Merino reported a mean alpha coefficient of .68 for a sample of 155 children ranging in age

from six to 11 years from Peru using a Spanish translation of instructions. Imuta et al presented a well-researched summary of DAP: IQ psychometrics indicating that the DAP: IQ usually produced scores that were reliable.

Interscorer Reliability

Reynolds and Hickman (2004) presented two studies concerning interscorer reliability in the examiner's manual. In one study 31 protocols from participants ranging from 11-to-75 years in age were employed and in a second study 148 protocols were selected from participants in six to 11 year range. After converting the raw scores to standardized IQ scores, the resulting correlation coefficients were .95 and .91 respectively. Williams et al. (2006) reported interscorer correlation coefficients of .83 for standardized IQ for a sample of 31 college students and Honores and Merino (2011) reported an interscorer coefficient of .91 for standardized IQ for a sample of 31 children from Peru six to 11 years of age.

Test-Retest Reliability

Test-retest reliability was reported by Reynolds and Hickman (2004) with a sample of 45 individuals ranging in age from 6 through 57 years who were tested twice. These individuals were retested within a 1-week time period. The resulting correlation coefficient between the two standardized IQ scores was .86.

Procedure

The instructions were translated into Chichewa. They were forward and back translated by two individuals fluent in Chichewa and English to ensure accuracy. The testing situation was presented to the students as an activity and proctored per the examiner's manual instructions. The verbatim translation used for the group instruction was:

Ndikufuna kuti aliyense adzijambule yekha. Muwonetsetse kuti mwajambula thupi lonse, osati mutu wokha. Mudzijambule m'mene inu mumaonekera chakumaso osati cha m'mbali ayi. Musajambulenso tianthu tonga ngati tindodo ayi. Mujambule chithunzi chanu chokongola m'mene mungathere. Jambulani mwachifatse ndipo mosathamanga. Panopa, mutha kuyamba. Amene ali ndi funso kapenanso wamaliza aimike dzanja lake.

Scoring instructions were followed as presented in the examiner's manual and the DAP: IQ Administration/Scoring Form. All analyses employed SPSS 22.0 (2013). Cronbach alpha coefficients were computed for the 23 test items for the total sample, the six age groups represented in the study, and for gender. The interscorer reliability was obtained by correlating the IQ scores from a random sample of 50 students' scores taken from the total sample that were scored by both researchers.

RESULTS

The Cronbach alpha coefficient for the entire sample ($n = 147$) was .81, ranging from .76 to .85 using a 95 percent confidence interval. The obtained alpha coefficients for the six age groups ranged from .68 to .92. Alpha coefficients for male and female students were .79 and .83 respectively. Table 1 shows the alpha coefficients and 95 percent confidence intervals for the total sample and the six age groups examined and for males and females. The results from the interscorer study were $r = .85$ for IQ scores from a random sample of 50 students scored by both researchers.

Table 1. Cronbach alpha coefficients from the DAP: IQ analyses.

Group	Sample Size	Alpha Coefficient Examiner's Manual	Alpha Coefficient Obtained	Alpha Coefficient 95% CI
CA = 5	15	.82	.68	.39-.87
CA = 6	57	.80	.80	.72-.87
CA = 7	28	.79	.72	.54-.85
CA = 8	25	.77	.75	.59-.87
CA = 9	13	.82	.89	.78-.96
CA = 10	9	.84	.92	.81-.98
Total Sample	147	.82	.81	.76-.85
Male	77	.80	.79	.72-.85
Female	70	.80	.83	.76-.88

Note: CA is chronological age. CI is Confidence interval.

DISCUSSION

This study evaluated the psychometric properties of the DAP: IQ by examining the reliability of test scores from a sample of Malawi elementary school children using a Chichewa translation of the group administration instructions. In summary, the DAP: IQ produced scores which demonstrated reliability coefficients that were similar to those presented in the examiner's manual (Table 1) for a sample of elementary school students from Malawi. The exception being the five year old age group. The interpretation of the test directions in Chichewa did not appear adversely affect the reliability of the scores in comparison to those listed in the examiner's manual or other published studies. Interscorer coefficients suggested that an acceptable level of agreement on test scores was found between examiners.

Potential advantages of the DAP: IQ are that it produced reliability scores psychometrically similar as those in the examiners manual with a population that is exposed to many different languages and influences. It was also relatively easy to administer and score. Observations of the children taking the test indicated that they truly enjoyed the activity appeared to be immersed in the task. Because the DAP: IQ relied on one drawing and was relatively easy to administer and score, it might prove useful for practitioners in schools who need a screening instrument for group administration. However, while the results concerning score reliability are satisfactory, there appears to be limited empirical evidence supporting the validity of the DAP: IQ for either children or adults. Therefore, additional studies examining the validity of the DAP: IQ with this population are needed for before any assertion can be made concerning the utility of DAP: IQ scores for this group.

AUTHOR BIOGRAPHIES

Denis Stanislaus Khasu received his Ph.D. in Curriculum and Instruction from Virginia Tech, USA. He is a professor of Literacy and Teacher Education at the Domasi College of Education in Malawi. Professor Denis Stanislaus Khasu, Domasi College of Education, Malawi, Africa.

Thomas O. Williams, Jr., Ph.D., received his Ph.D. in Rehabilitation and Special Education from Auburn University. He is an associate professor in the School of Education at Virginia Tech, USA. Corresponding author: Thomas O. Williams, Jr., School of Education/Teaching and Learning, War Memorial Hall, RM 314, Virginia Tech, 370 Drillfield Drive, Blacksburg, VA 24061, USA. Email: thwilli1@vt.edu.

REFERENCES

- Barnes, E. (1893). A study of children's drawings. *Pedagogical Seminary*, 2, 451-416.
- Goodenough, F. (1926). *Measurement of intelligence by drawings*. Chicago: World Book.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough Draw-A-Man Test*. New York: Harcourt Brace Jovanovich.
- Honores, L., & Merino, S., (2011). Confiabilidad de la prueba de habilidad intelectual. *International Journal of Psychological Research*, 4(2), 57-63.
- IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

- Imuta, K., Scarf, D., Pharo, H., & Hayne, H. (2013). Drawing a Close to the Use of Human Figure Drawings as a Projective Measure of Intelligence. PLoS ONE, 8(3), e58991. doi: 10.1371/journal.pone.0058991
- Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. New York: Grune & Stratton.
- Koppitz, E. M. (1984). *Psychological evaluation of human figure drawings by middle school pupils*. New York: Grune & Stratton.
- Naglieri, J. A. (1988). *Draw-A-Person: A Quantitative Scoring System*. San Antonio, TX: The Psychological Corporation.
- Reynolds, C. R., & Hickman, J. A. (2004). *Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults examiners manual*. Austin, TX: Pro-ed.
- Ricci, C. (1894). L'arte de bambini. *Pedagogical Seminary*, 3, 302-307.
- Williams, T. O., Jr., Fall, A.M, Eaves R.C., & Woods-Groves, S. (2006). The stability of scores for the Draw a Person IQ Test. *Journal of Psychoeducational Assessment*, 24(2), 137-144.

NOTES