DOCUMENT RESUME

ED 388 670 TM 023 109

AUTHOR Khammash, Salma B.

TITLE Omani Norms for the Arabic Picture Vocabulary Test (APVT) for Schoolchildren at the Elementary Levels.

PUB DATE 22 Apr 95

NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Arabic; Elementary Education; *Elementary School Students; Foreign Countries; Language Proficiency; *National Norms; Norm Referenced Tests; Pilot Projects; *Test Reliability; Test Use; Test Validity; *Verbal Ability

IDENTIFIERS Arabic Picture Vocabulary Test; Cross Validation; *Oman; Peabody Picture Vocabulary Test; Standardization

ABSTRACT

The Arabic Picture Vocabulary Test (APVT) was derived from the Peabody Picture Vocabulary Test in order to provide an estimate of a student's acquired verbal ability in Arabic (Modern Standard Arabic) through his or her aural vocabulary. It is an individually administered norm-referenced power test for students aged 5 to 9 plus years. Each item consists of a page of pictures together with a spoken word. The test was expected to be appropriate for the evaluation of Omani children as a measure of verbal ability. Because a pilot study involving 150 Omani students was encouraging, standardization studies were undertaken in 1993 with a representative sample of 1,600 Omani students in grades K through 4. Norms development then began with a careful check of the individual test records. Deviation-type norms and developmental-type norms were derived. Adequate reliability was found, and the concurrent validity of the test for this sample was supported. Cross-validation studies will be ongoing to establish construct validity and enable academic predictions. (Contains 5 tables and 61 references.) (SLD)

ED 388 670

OMANI NORMS FOR THE ARABIC PICTURE VOCABULARY TEST (APVT) FOR SCHOOLCHILDREN AT THE ELEMENTARY LEVELS

Salma B. Khammash

Sultan Qaboos University

Paper presented at the American Educational Research Association Annual Meeting, AERA/SIG International Studies San Francisco, April 22, 1995.

OMANI NORMS FOR THE ARABIC PICTURE VOCABULARY TEST (APVT) FOR SCHOOLCHILDREN AT THE ELEMENTARY ELEVES

by

Salma B. Khammash

Sultan Qaboos University

INTRODUCTION

The Arabic Picture Vocabulary Test (APVT) derived from the PPVT-R and developed by this writer at the University of Michigan was item validated and standardized in Jordan on a representative sample of 2900 students (1450 boys and 1450 girls) in grades K to 6 for item-validation, and K to 3 for standardization. Non-Jordanian Arab students in these grades were included in the standardization sample.

The APVT was designed to provide an estimate of a student's acquired verbal ability in Arabic (Modern Standard Arabic) through his/her aural (hearing) vocabulary. It is an individuall administered norm referenced power test, devised for students ranging in age from 5 to 9+, who have normal sight and hearing and understand Modern Standard Arabic. The Instrument contains 150 items (preceded by 7 practice items), arranged in order of increasing difficulty. A page of pictures (known as a plate) together with a spoken word constitutes a test item. Each plate contains four illustrations representing optional responses to the stimulus word associated with plate number. The four drawing on the plate, all of equal size and intensity, represent objects

familiar to the age group for whom they are intended. The child's (student) task is to identify the picture considered to illustrate best the meaning of a stimulus word presented orally by the examiner, and as the test proceeds, a point is reached where the items cease to be fully comprehended. At this time the examiner ends the testing session.

Test scores can be generally interpreted as measures of acquired verbal ability in so far as they assess through hierarchy of vocabulary difficulty, the level of semantic complexity at which the examinee (student) is able to understand words.

Features

The features of the APVT include the following:

1- Individual administration.

2- Clear, bold line drawings of high interest and broad appeal. The drawings involve a variety of objects and actions which should be familiar to children growing in modern Arab culture, for example:

- human body parts
- animals an birds
- man-made objects
- clothes and jewelry
- mathematical shapes
- human actions (gerund form)
- fruits an vegetables
- musical instruments
- plants and trees
- occupations (doctor,teacher)

3- Quick administration. Testing may take twenty (20) to thirty (30) minutes.

4- Untimed administration, to test power rather than speed.

5- No reading required of the examinee.

6- No oral or written responses are required of the examinee. When the examiner pronounces the stimulus word the examinee could simply point to the correct answer which is presented as a drawing.

7- Objective, rapid scoring which is accomplished while the test is being administered.

8- The items administered are only at and over the examinee's ability range. The examinee must answer only those items of suitable difficulty.

9- Norm-referenced interpretation. Raw scores were converted to 'stand score equivalents' and 'age equivalents' (developmental-type norms).

10- No extensive specialized preparation is need by the examiner. Test administration is simple and can be undertaken by any experienced teacher or school official who follows the instruction carefully. The examiner, however, must be proficient in Modern Standard Arabic in order to read the directions to the student (examinee), and to know the correct pronunciation of each of the stimulus words to be administered.

Purposes and Uses

1- As stated earlier, the APVT is designed to provide a quick estimate of verbal ability for children who have grown up in Modern Standard Arabic-speaking environments. In this case it is a scholastic aptitude test. It is not however, a comprehensive test of general intelligence; it measures only one important facet of general intelligence: namely vocabulary. According to Dale and Reichert "vocabulary is the best single test index of school success" (Dale and Reichert, 1957).

2- Originally, the APVT was designed to assess level of hearing vocabulary of incoming Arabic-speaking children with limited English proficiency in kindergarten and early primary grades in public schools in the United States. It was (and still is) intended to assist school personnel in assessing and making

proper grade placements, and diagnosing the learning difficulties of these students. Since the APVT provides a reasonably good measure of scholastic aptitude for these students, it could be useful as an initial screening device in scanning for bright, low ability, and/or language impaired children who may need special attention.

3- An important function of the APVT is to provide a measure of the student's aural (listening) vocabulary in Modern Standard Arabic. In this sense it is an achievement test, because it samples the student's acquired vocabulary in Arabic. It would also, be helpful in identifying underachievers, when used in conjunction with a measure of school achievement.

4- The APVT may be used for diagnostic and clinical purposes. Since the children who sit for the test are not required to read or write, the instrument especially is useful for nonreaders and other persons with written language problems. Because the responses can be gestural, it is likewise appropriate for the stutterers. Extensive verbal interaction between the examiner and the examinee (child) is not required, making the APVT less threatening than many other individual tests. Moreover, because it is individually administered, it may be used successfully with withdrawn children who will not cooperate well in taking a group test. For all the above reasons, the APVT had been in use for the past (8) years by the Center for Child Evaluation and Teaching in Kuwait.

5- In research application, the APVT could be used as either a moderator or a dependent variable for studies involving limited English proficiency students in primary grades whose home language is Arabic, as the case in the United States and other English speaking countries. Furthermore, it could provide a useful assessment of students in studies relating performance and gender, as well as performance and geographic locations.

Limitations

The specific characteristics which recommend the APVT for general use, are it convenience, shortness and simplicity. However, these characteristics also can result in casual administration and scoring. It is essential to remember that the test measures only 'verbal ability', just one aspect of the complex of cognitive and linguistic domains of intelligence. Also, the test should be viewed as a measure of the level of the person's present functioning as opposed to his/her 'generic ability'. In short, it should be considered only as one aspect of comprehensive study of the individual student.

Test Materials

The APVT materials includ the following:

1. A spiral bound test-book containing seven (7) training plates plus 150 numbered test plates, all labelled in Arabic.

2. A separately-bound Individual Test Record, also in Arabic. The test-record includes stimulus words to be used with both training and test plates, along

with instructions to the examiner. In addition, space is provided for recording student responses and for demographic information such as age, gender, grade, location and type of school, etc.

3. The APVT Manual with instructions in both English and Arabic, and technical information about the test.

Need for the APVT in the United States

The need for the development of the Arabic Picture Vocabulary Test (APVT) in the United States came in response to three factors:

- The first factor was in the form of an urgent request by public school officials to find a test in Arabic, preferably a 'vocabulary test', to help them in the proper grade placement of newly arriving Arab children. The majority of these children were coming from southern Lebanon and the occupied West Bank. Most of them did not possess the necessary documents to show their age and/or any previous schooling.

- The second factor was in the form of an urgent plea from parent of these children. They complained that their children were labelled "mentally retarded" or having severe learning disabilities and were placed in special education programs.

  After inquiry, it was found that these children were tested upon arrival in English - using such tests such as the Columbia Mental Test, the Peabody Picture Vocabulary Test (PPVT), the Stanford-Binet and/or the Wechsler Mental Test - and scored very low. As a result they placed in special education programs.

- The third factor was in the form of a court order known as the "Lao Decision of 1979". The Lao Decision handed down by the U.S. Supreme Court specified that every newly arriving student whose home language is not English should be tested in his/her own language before grade-placement in the school; particularly public schools that have Title VII bilingual programs. (The Lao Decision '79 came as a result of a lawsuit by a Laotian family of a young female student who was placed in special education wrongly against the California school system).

This writer, being the Director of Research and Evaluation at the Arabic Language Bilingual Materials Development Center (ALBMDC) at the School of Education, the University of Michigan was given the responsibility to find or develop a test in Arabic as soon as possible. That's why the Arabic Picture Vocabulary Test (APVT) was developed.

Construction of the APVT at The University of Michigan

Prior to the construction of final APVT at the Arabic Language Bilingual Materials Development Center at the University of Michigan, in 1982, item analysis for item difficulty, item discrimination and wrong answer distraction was carried out for each year-level-group separately in the item-validation sample of 1400 students (700 boys, 700 girls) in grades K to 6 in Jordan the previous year. The item analysis provided the following information for each item in the test:

1- Item difficulty, expressed as the percentage of examinees passing an item in successive age group.

2- Item discrimination, expressed for each group as the point-biserial correlation of the item with the total score.

3- Item decoy action, expressed as the percentage of examinees selecting each distracter on the plate.

4- Negatively discriminating distracter information, determined by comparing the percentage of examinees in high and low ability groups who selected each distracter.

5- Gender differences, expressed as the statistical significance between the proportion of female and male examinees answering each item correctly.

In addition to the above information, examiners' comments on the items were reviewed. Items were checked again to identify gender stereotypes or cultural illustrations. Also, and assessment was made of the kinds of words selected so that a better balance across the categories in the test could be achieved, for example: nouns, gerunds for action, descriptors, etc.

Items were selected for inclusion in the final APVT if they satisfied the following six criteria:

1- Within the 40 to 60 percent difficulty range of the item, at least one of the phi values should reach the 1% significance level.

2- Outside the 40-60%, the remaining phi values should be positive except in the case of items with a sharp difficulty rise with age, where for the youngest age group it may be negative.

3- The difficulty of an item should decrease with age uniformly.

4- The phi values of each of the three distracters for the item should be negative or zero within the 40-60% difficulty range of the item.

5- The remaining values for the distracters should also be mainly negative, except in the case of items with steep difficulty gradients where these distracters may have positive phi values for the youngest age group.

6- A distracter was also retained if it absorbed at least 10% of the total distraction.

On the basis of these criteria, a number of items had been flagged for elimination. It was decided to pool the items (from the two forms L&M used in item-validation activities) to make a single test. This was accomplished by grouping the remaining

items for each age level, covering the age range 5 to 9+ years. Where there were several items at a particular age level, the best discriminators were chosen, and where competing items were of equal quality the required number were chosen randomly. In the final version, the consolidated **APVT** contained 150 items in a single book. All the age ranges covered were presented within the test by items with difficulty levels between 40 to 60% for each year group, while the upper and lower reaches of the test were covered by items with increasingly extreme difficulty levels -- which retained significant phi values. Seven (7) practice items were also included, 4 for ages under eight years, and 3 for examinees 8 years and above. The complementary Individual Test Record was also developed by compiling the corresponding words in Arabic.

APVT STANDARDIZATION IN OMAN

The Need for the APVT in Oman

One of the goals of the Education Research Center at the College of Education and Islamic Sciences, Sultan Qaboos University, is to standardize intelligence and personality test in Oman. The **APVT** which was devised as an assessment instrument appropriate for measuring intelligence including verbal ability was seen to be suitable for Omani children in the kindergarten and elementary levels. The purpose of the test is to assist school officials in the instructional decisions,

and the proper evaluation of learning difficulties these children are experiencing in school. When asked, educators and evaluation specialists as well as teachers have indicated that such a test was nonexistent in Oman. Furthermore, no test materials were found that could be used without major modifications. This was felt to create problems in providing Omani school-children with an effective educational program geared to meet their needs, perceptions and abilities.

The APVT, with its Arab norms (Jordan), is not a comprehensive test of general intelligence, it measures only one facet of general intelligence: vocabulary. In 1937, reporting on the Revised Stanford Tests of Intelligence, Terman and Merrill wrote: "we have found the vocabulary test to be the most valuable test on the scale." More recently, the APVT was found by specialists at the Center for Child Evaluation and Teaching Kuwait, where children with learning difficulties are referred, to be the most consistent test in assessing the Kuwaiti child's verbal ability through his/her aural (hearing) vocabulary. These facts sufficiently demonstrate the crucial need for a test which measures the Omani child's vocabulary that is normed in Oman.

The APVT Pilot Study in Oman

The APVT was pilot tested in Oman in primary schools in the Muscat area in April-May of 1991. The purpose of the pilot study was to determine whether the APVT in its current version

was: 1) usable in terms of language, and 2) the test items would adequately discriminate among Omani students at each of the targeted age levels, 5 through 9+ years.

A sample of 150 Omani students (75 boys and 75 girls) in grades K to 4 from 8 schools were randomly selected for testing (non-Omani students were excluded from the sample). All 40 kindergarten students were selected from four (4) private schools, and the remaining 110 students (55 boys, 55 girls) in grades 1-4 were selected from four (4) government primary schools.

The results of the pilot study were very encouraging. These favorable results, however, were based on student performance in the Muscat area only, and did not reflect student performance in the rest of the country. Therefore, and in order to use the test properly, it became necessary to standardize the APVT on a representative student sample from schools throughout the Sultanate.

APVT Standardization

The APVT standardization activities occurred in late 1992 and the first three months of 1993. The sampling plan was based on student population data from the 1991 Documents on General Education in the Sultanate, which was furnished by the Ministry of Education and Youth.

Sample Characteristics

The stratification criteria used to ensure adequate representation in the sample included the following: (1) grade and gender representation, (2) geographic and school representation, and (3) appropriate socioeconomic representation.

A representative sample of 1600 Omani students, in grades K-4 and stratified by gender and grade, was systematically selected from (60) government and private schools in the Sultanate (1160 students in grades 1-4 from government schools, and 440 students in K-1 from private and police kindergarten schools). Non-Omani students in these grades were excluded from the sample. The distribution of the standardization sample by grade and gender appears in Table 1 below.

The 487 students in the first grade included 287 students (150 males, and 136 females) from government schools throughout the country and 200 students (100 males, and 100 females) from private schools in the Muscat area only.

The standardization plan also called for the development of norms within age levels grouped at six months intervals up to 9 years old, and every year thereafter. Ideally, we would have sampled an equal number of both genders at each age level, but this was not feasible due to the involved governmental permission process. However, the school sampling process did yield age groups sufficiently large to compute reasonably stable norms,

Table 1

APVT - Omani Standardization Sample
According to Grade and Gender
(N = 1600)

Grade	Male	Female	Total
Kindergarten	115	125	240
First Grade	250	237	487
Second Grade	143	140	283
Third Grade	157	160	317
Fourth Grade	115	158	273
Total	780	820	1600

and a fairly close balance between genders. The distribution of the students in the sample by age and gender is shown in Table 2. For the purpose of norms-development, students who age was 13 and above (N=10) were eliminated, and those students who age ranged from 11 years to 12:11 were groups together (N=100).

Schools and Geographic Locations. Government schools were selected to be representative of their geographic locations, while private schools were selected from the Muscat area. Overall, there were 48 government schools, 10 private schools and 2 police kindergarten schools, making a total of 60 schools. In all cases, except for the kindergartens and private schools, equal numbers of male and female schools were selected. Twenty

Table 2

APVT - Omani Standardization Sample
According to Age and Gender
(N = 1590)

Age Groups (yrs. & mons.)	Male	Female	Total
5:0 - 5:5	30	30	60
5:6 - 5:11	60	70	130
6:0 - 6:5	61	62	123
6:6 - 6:11	76	81	157
7:0 - 7:5	81	78	159
7:6 - 7:11	66	63	129
8:0 - 8:5	48	62	110
8:6 - 8:11	62	70	132
9:0 - 9:11	137	133	270
10:0 - 10:11	110	110	220
11:0 - 12:11	43	57	100
Total	774	816	1590

(2) schools were located in and around Muscat, these included:

- 8 government (public) elementary schools (4 male schools, and 4 female schools).
- 10 kindergarten and private elementary schools.*
- 2 police kindergarten schools.

The remaining 40 government schools were located in towns in various regions of the Sultanate. These schools and locations includes the following:

- 4 schools (2 male, 2 female) in Buraimi (Dhahira region)

* All kindergarten schools are private or police controlled, as there were no government kindergartens in Oman at the time the data were gathered.

- 4 schools (2 male, 2 female) in Ibri (Dhahira region)
- 8 schools (4 male, 4 female) in Nizwa (Dakhiliya region)
- 8 schools (4 male, 4 female) in Sohar (Batina region)
- 8 schools (4 male, 4 female) in Sur (Sharqiya region)
- 8 schools (4 male, 4 female) in Salalah (Ganubiya region)

Table 3 contains summary information on the school-location and the students included in the APVT total standardization sample.

Table 3

APVT - Omani Standardization Sample
According to Location, School and Gender
(N = 1600)

Schools and Geographic Location	Male	Female	Total
1- Muscat and adjacent areas (20 schools)*	375	385	760
2- Buraimi and Ibri (8 schools)	90	120	210
3- Nizwa (8 schools)	80	80	160
4- Sohar (8 schools)	80	80	160
5- Sur (8 schools)	80	80	160
6- Salalah (8 schools)	75	75	150
Total	780	820	1600

As indicated before, 1160 students in the sample (560 males, 600 females) in grades 1-4 were selected from government school in the country. The following table (Table 4) shows the distribution of these students according to location, grade and gender.

* Government (8), private (10), police (2)

Table 4

APVT - Omani Standardization Sample from Government Schools According to Location, Grade and Gender (N = 1160)

Geographic Location	Grade 1		Grade 2		Grade 3		Grade 4	
	M	F	M	F	M	F	M	F
Muscat	40	40	40	40	40	40	40	40
Buraimi & Ibri*	30	30	30	30	30	30	-	30
Nizwa	20	20	20	20	20	20	20	20
Sohar	20	20	20	20	20	20	20	20
Sur	20	20	20	20	20	20	20	20
Salalah**	20	7	13	10	27	30	15	28
Total	150	137	143	140	157	160	115	158
Grand Total	287		283		317		273	

Socioeconomic Representation

The standardization sample was also designed to provide a broad socioeconomic representation of Oman. Occupation of the major wage earner in the family (recorded by the examiner) was used to ensure appropriate representation in the sample. The sample contained approximately proportionate numbers of the occupations cited by the Omani Development Council, 1992 Statistical Year Book. They include:

1- Professionals (doctor, lawyer, university professor)

2- Government Ministries and related organizations (civil service employees -- including teachers in government schools).

* 30 fourth grade males were not test in a school outside Ibri (in Iraqi) due to communication problems.

** 90 students (45 males, 45 females) were not tested in Salalah due to technical problems.

3- Military service -- including police.

4- Private sector (banks, business owner, manager, clerk).

5- Skilled workers (craftsmen, goldsmith, silversmith).

6- Land owners and farmers.

7- Fisheries (fishermen and labor).

Procedures

After securing the permission of the University Officials and the Ministry of Education Officials to carry out the APVT standardization, a list of schools in several locations in the country where testing will take place was prepared by this writer. Officials in the Ministry then coordinated with the Directors in these locations making arrangements for the personal visit of the writer accompanied by a demonstrator from the Educational Research Center at Sultan Qaboos University in order to train the teachers who will carry out the testing, and to supervise on site testing. Beside training the teachers, the demonstrator helped in the selection of the students to be included in the sample. At each school, students were systematically selected within the predetermined categories of grade and gender.

In addition to the standardization sample, a number of students who were tested earlier were retested (by the demonstrators) for reliability purposes.

Omani Norms Development

Norms development began with a careful check of the 1600 individual test records. Each individual test record was checked by this writer to verify the correct determination of the basal and ceiling levels and the raw score recorded by the examiner (teacher). The corrected data were then entered into the computer for the computations of norms. Two type of Omani norms were calculated:

1. Deriving the Omani Deviation-type Norms

The following six steps were used to obtain the deviation-type norms that include the standard score equivalents, and the corresponding percentile and stanine equivalents:

1- Means and standard deviations were computed separately for each of the eleven (11) age groups tested.

2- For each of the eleven (11) age groups, the raw scores were then converted to normalized z scores based on the normal probability curve.

3- Interpolations were made to smooth the distribution of derived z scores.

4- Percentile equivalents corresponding to each z score were read from a table of the normal curve.

5- Stanines were computed from the normalized z scores by the usual formula.

6- The normalized z scores were converted to standard Score Equivalents having a mean of 100 and a standard deviation of 15.

The resulting tables of these deviation-type norms appear in Part I of the APVT Manual.

2. Deriving the Omani Developmental-Type Norms

The developmental norms known as the age equivalent scores were computed in three steps:

1- The mean raw score corresponding to a standard score of exactly 100 was computed for each age group.

2- These mean raw scores and the corresponding mid-points for age-interval were plotted on a large graph, and the curve was smoothed.

3- The range of the raw score corresponding to each month of age was read directly from the graph.

The APVT Omani age equivalents corresponding to the various raw scores appear in Part I of the APVT Manual.

It was evident from Omani standardization process that more age groups were included in the norms than in the original APVT. In order to accommodate the new age groups, it became necessary to adjust the basal items in the Individual Test Record. This in turn called for rearranging the sequence of few items in the test to fit the Omani norms.

Reliability

The reliability or consistency with which the APVT measures performance of a child or student across sets of items must be considered by those interpreting test results. Two types of reliability coefficients were reported for the APVT-Omani version: (1) internal consistency, and (2) delayed test-retest "short-term" stability.

Coefficients of Internal Consistency. Split-half correlations based on all students in the standardization sample

(grades K-4) were obtained. For each age group, the odd and even item scores were correlated, and the resulting split-half correlations were corrected for half test length by the Spearman-Brown formula. The reliability coefficients ranged from .88 -.95 for all age groups (5 to 12 years old), as compared to .90 -.93 for all age groups in the original APVT normed in Jordan (Khammash and Dunn, 1985).

Delayed Retest Reliability Coefficient. Reliability coefficients were computed for the APVT-Omani version by correlating test-retest scores. This provided test-retest "short-term" stability over a one-year interval.

The reliability sample of 160 students from schools in Muscat, Nizwa, Sohar and Sur that participated in the standardization activities (70 males, and 90 females), grades K-4 equalled 10% of the total standardization sample. The coefficients of stability ranged from .930 to .945. These results are similar to those found for the original APVT (Khammash and Dunn, 1985).

Standard Error of Measurement. Since test interpretation is to be made on the basis of standard score equivalents, it was desirable to have the standard error of measurement (SEM) of these scores. Also, since score distributions of the various age groupings did not differ from a normal distribution, the test-retest reliabilities were considered an unbiased estimate of the reliabilities of the corresponding standard score equivalents. Using these reliability estimates, and remembering that all

standard score equivalents have a standard deviation of 15, the standard error of measurement (SEM) was computed for each grade grouping using the usual formula. The standard error of measurement for the standard score equivalents at each grade level is shown in Table 5.

Table 5

Standard Error of Measurement for Standard Score Equivalents for the Reliability Sample According to Grade Level (N = 160)

Grade at APVT Testing	Reliability Coefficient	Standard Error of Measurement
Kindergarten	.935	3.82
First grade	.930	3.97
Second grade	.938	3.73
Third grade	.940	3.68
Fourth grade	.945	3.55

The observed values for the standard error of measurement (SEM) were remarkably consistent, and all below 4.0. It is therefore reasonable that 4.0 (points) can be used as a conservative estimate of the standard error of measurement of the standard score equivalents at all age levels.

What does this mean? and how does the SEM affect the individual score? It means that since the SEM=4, then the chances are about 68 in a 100 (or about 2 out of three) that the examinee's true score will differ from obtained score by no more than ±4, and 96 in a 100 that the true score will differ from obtained score by no more than ±8 points.

Validity

Normally, three types of test validity are discussed: (1) content validity, (2) construct validity, and (3) concurrent validity. Since both the content validity and the construct validity have been established for the original APVT earlier, only concurrent (criterion-related) validity of the APVT-Omani version in relation to the school-subjects taught at the primary levels in Oman will be discussed.

Concurrent Validity. Concurrent validity was measured by correlating the APVT raw score of the student with his/her subject marks at the end of the school year. Significant correlation coefficients were observed when the APVT score was correlated with marks for Arabic, Social Studies, Science and Mathematics for all age/grade levels, locations and school type (government or private) for both genders. Overall coefficients for all subjects (Arabic, Mathematics, Social Studies and Science) were .40. Higher coefficients were observed for those who were in the age groups 6:6 to 7:11 and 11:0 to 12:11 than for students in the age groups 8:0 - 9:11 (.48, and .45 as compared .32) for all subjects.

Finally, cross validation studies to establish the validity of the APVT with the Omani norms for the purpose of construct validity (such as correlations with other vocabulary subtests of other mental tests) and/or for predicting academic achievement is a continuous process that requires the accumulation of many research findings spanning several years.

BIBLIOGRAPHY

American Psychological Association. (1966b). Standards for Educational and Psychological Tests and Manuals. Washington, D.C.

American Psychological Association. (1974). Standards for Educational and Psychological Tests. Washington, D.C.

American Psychological Association. (1985). Standards for Educational and Psychological Testing. Washington, D.C.

Anastasi, A. (1982). Psychological Testing (5th ed.). New York: MacMillan.

Anderson, D.E., and Flax, M.L. (1968). A Comparison of the Peabody Picture Vocabulary Test and Wechsler Intelligence Scale for Children. Journal of Educational Research, 62, 114-116.

Anderson, T. (1965). A new focus on the bilingual child. Modern Language Journal, 49, 156-160.

Angoff, W.H. (1971). Units, scores and norms. In R.L. Thorndike (Ed.), Educational Measurement. Washington, D.C.: American Council on Education.

Ba'albaki, M. (1979). Al-Mawrid, A Modern English-Arabic Dictionary. (3rd ed.). Beirut, Lebanon: Dar Al-Ilm Lil-Malayen.

Ba'albaki, M. (1987). Al-Mawrid, A Modern English-Arabic Dictionary (4th ed.). Beirut, Lebanon: Dar Al-Ilm Lil-Malayen.

Bashaw, W.L. and Ayers, J.B. (1967). An evaluation of the norms and reliability of the PPVT for pre-school subjects. Educational and Psychological Measurement, 27, 1069-75.

Binet, A., and Simon T. (1916). The Development of Intelligence in Children. Vineland, N.J.: Training School Publication, No. 11.

Brimer, M.A. (1969). Sex differences in listening comprehension. Journal of Research and Development in Education, 3, 72-79.

Brimer, M.A. and Dunn, L.M. (1962). Manual for the English Picture Vocabulary Test (EPVT). Bristol, England: Educational Evaluation Enterprises.

Brown, F.G. (1983). Principles of Educational and Psychologica Testing (3rd ed.). New York: Holt, Rinehart and Winston.

Cartwright, G.P. (1968). A note on the use of the Peabody Picture Vocabulary Test with disadvantaged children. Journal of Educational Research, 61, p.285.

Center for Bilingual Education. (1974). Evaluation Instruments for Bilingual Education; An Annotated Bibliography.Austin, Texas: National Dissemination and Assessment Center.

Center for Bilingual Education - Northwest Regional Laboratory. (1978). Assessment Instruments in Bilingual Education. Los Angeles, CA: National Dissemination and Assessment Center.

Chase, C.I. (1978). Measurement for Educational Evaluation (2nd ed.). Reading, Mass: Addison-Wesley.

Childers, P.R. (1966). Concurrent validity of a group administered Peabody Picture Vocabulary Test. Journal of Educational Research, 60(2), 90-93.

Corwin, B.J. (1965). The influence of culture and language on performance of individual ability test. Journal of School Psychology, 3(3), 41-47.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.). Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education.

Cronbach, L.J. and Meehl, P.E. (1955). Construct validity in Psychological Tests. Psychological Bulletin, 52, 281-301.

Dale, E. and Reichert, D. (1957). Bibliography of Vocabulary Studies. Columbus, OH: Ohio State University, Center of Educational Research.

DiLorenzo, L.T. and Brady, J.J. (1968). Use of the Peabody Picture Vocabulary Test with Preschool Children. Psychological Reports, 22, 247-251.

Dunn, L.M. (1965). Peabody Picture Vocabulary Test (PPVT) Expanded Manual. circle Pines, MN.: American Guidance Service.

Dunn, L.M. and Dunn, L.M. (1981). Peabody Picture Vocabulary Test - Revised (PPVT-R) Manual. Circle Pines, MN.: American Guidance Service.

Ebel, R.L. (1979). Essentials of Educational Measurement (3rd ed.). Englewood Cliffs, N.J.: Prentice Hall.

Elliot, R.N., Jr. (1969) Comparative study of Pictorial Test of Intelligence and the Peabody Picture Vocabulary Test. Psychological Reports, 25, 528-530.

Garrett, H.E. (1967). Statistics in Psychology and Education (6th ed.). New York: McKay Co. Inc.

Glass, G.V. and Hopkins, KD. (1984). Statistical Methods in Education and Psychology (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Gronlund, N.E. (1985). Measurement and Evaluation in Teaching (5th ed.). New York: MacMillan.

Guilford, J.P. (1965). Fundamental Statistics in Psychology and Education (4th ed.). New York: McGraw-Hill.

Henryssen, S. (1971). Gathering, analyzing and using data on test items. In R.L. Thorndike (Ed.). Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education.

Hickey, T. (1972). Bilingualism and measurement of intelligence and verbal learning ability. Exceptional Children, 39, 24-28.

Johnson, D.L. and Johnson, C.A. (1971). A comparison of four intelligence tests used with culturally disadvantaged children. Psychological Reports, 28, 209-210.

Jordan Department of Statistics. (1973). A Statistical Guide to Jordan. Amman, Jordan.

Jordan National Planning Council. (1976). Jordan Five-Year Plan, 1976-1980. Amman, Jordan.

Jordan Ministry of Education. (1980). A Guide to Educational Surveys in Jordan for School-year 1980-81. Amman, Jordan: Survey and Statistics Division, (Arabic).

Khammash, S. B. and Dunn, L.M. (1985) Arabic Picture Vocabulary Test (APVT) Manual. Ann Arbor, MI: Arabic Bilingual Materials Development Center, School of Education, The University of Michigan.

Khammash, S.B. (1986). A pictorial test assessing aural-vocabulary of Arabic-speaking children at kindergarten and early primary levels. Resources in Education (RIE). Vol. 21, No. 9 (TM 860-264).

Khammash, S.B. (1990). Arabic Picture Vocabulary Test (APVT) Manual (2nd ed.). Amman, Jordan: Al-Eiman Printing Press.

Khammash, S.B. (1991). The Pilot Study of the Omani Standardization of the Arabic Picture Vocabulary Test (APVT) - A Report. Muscat, Oman: Educational Research Center, College of Education and Islamic Sciences, Sultan Qaboos University.

Khammash, S.B. (1992). Gender differences in aural-vocabulary of elementary schoolchildren in Jordan. Resources in Education (RIE). Vol. 27, No. 6.

Mehrens, W.A. and Lehmann, I.J. (1980). Standardized Test in Education (3rd ed.). New York: Holt, Reinhart and Winston.

Mehrens, W.A. and Lehmann, I.J. (1984). Measurement and Evaluation in Education and Psychology (3rd ed.). New York: Holt, Reinhart and Winston.

Mehrens, W.A. and Lehmann, I.J. (1987). Using Standardized Tests in Education (4th ed.). New York: Longman.

Mitchell, J.V. (1985). The Ninth Mental Measurements Yearbook. Lincoln: University of Nebraska Press.

Nunnally, J. (1978). Psychometric Theory (2nd ed.). New York: McGraw-Hill.

Popham, W.J. (1974). Evaluation in Education. Berkeley, CA.: McCutchan.

Popham, W.J. (1975) Educational Evaluation. Englewood Cliffs, N.J.: Prentice Hall.

Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.). Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education

Spolsky, B. (1969). Language testing - the problems of validation. Florida Foreign Language Reporter, 7(1), 100-102, 163-164.

Sultanate of Oman Development Council. (1992). Statistical Year Book. Muscat, Sultanate of Oman.

Sultanate of Oman Ministry of Education. (1991). Statistical Year Book. Muscat, Sultanate of Oman.

Terman, L.M. (1916). The Measurement of Intelligence. Boston, Mass.: Houghton Miflin.

Terman, L.M. and Merrill, M.A. (1937). Revised Stanford-Binet Tests of Intelligence. Boston, Mass.: Houghton Miflin.

Thorndike, R.L. (1971) Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education.

Thorndike, R.L. and Hagen, E. (1977). Measurement and Evaluation in Psychology and Education (4th ed.). New York: Wiley and Sons.

Traub, R.E. and Rowley, G.L. (1980). Reliability of test scores and decisions. Applied Psychological Measurement, 4(4), 517-545.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children-Revised Manual. New York: The Psychological Corporation.

Wiersma, W. and Jurs, S.G. (1990). Educational Measurement and Testing (2nd ed.). Needham Heights, Mass.: Allyn & Bacon.
