

Standardization of the Revised Arabic Language test for 4-8-year old children

Original Article

Nahla Rifaie¹, Tarek Mohamed Abdel Wahab Hamza², Yomna Hassan Elfiky³

^{1,3}Department of Otorhinolaryngology, Faculty of Medicine, Ain Shams University,

²Community Service and Environmental Development, Faculty of Arts, Fayoum University, Egypt.

ABSTRACT

Background: Clinicians use standardized assessments extensively in clinical and research settings to assess children's language skills and provide information for diagnosis, service eligibility, and intervention decisions.

Aims: The study aimed at standardization of the newly designed and formatted "Arabic Language Test (ALT)" for the age range 4-8 years.

Patients and Methods: Piloting of the test was done on 60 children with normal language development (15 from each one-year age) from 4- 8 years old. The sample of standardization was 720 Egyptian children divided into four groups each of 180 children. Group A (4-5 years), group B (5-6 years), group C (6-7 years) and group D (7-8 years). 60 children were retested after 2 weeks by the same test to measure reliability (test-retest, Split half method and Alpha Cronbach). 80 children with language disordered were tested to measure validity (Internal consistency, contrasted group validity, judgement and face validities).

Results: Highly significant scores were obtained for validity measures, as well as for reliability measures.

Conclusion: The recently revised 4-8-year-old Arabic Language Test is a reliable and effective test that can be used to assess the language development of Egyptian children and detect language deficiencies in children of the same age.

Key Words: Delayed language, language assessment, standardization.

Received: 25 August 2021, **Accepted:** 25 October 2021

Corresponding Author: Yomna Hassan Elfiky, MD, Department Otorhinolaryngology, Faculty of Medicine Ain shams University, Egypt, Tel.: 011118898052, E-mail: Yomna@med.asu.edu.eg

ISSN: 2090-0740, 2021

INTRODUCTION

Clinicians use standardized assessments extensively in clinical and research settings to assess children's language skills and provide information for diagnosis and service eligibility. Given the importance of these decisions, it is important that the evaluation has good psychometric properties^[1].

Strong psychometric assessment is an important part of developing effective and efficient interventions to support children's achievements and ambitions^[2].

Attributes include good concepts of validity and reliability^[3]. When there is a lack of data on psychometric characteristics, the problem arises from using the results of assessment to analyze the outcome data for clinical decision-making in research. The fifth edition of Clinical Evaluation of language functions (CELF)^[4], although it is not considered a comprehensive language assessment tool, it has good psychometric properties. Reliability refers to the consistency of a measure. Psychologists consider three types of consistency over time. Validity is the extent to

which the scores from a measure represent the variable they are intended to^[5]. Reliability is proved through retesting, splitting in half, and interrater methods. Whereas, the test presents content validity, concurrency validity and structural validity through the correlation between standard comprehensive scores and discriminatory analysis. In the field of Arabic, the Arabic Language Test (ALT) is the first normative reference standardized test constructed in the Middle East in 1995 by Kotby *et al.*^[6] to assess the language of Arabic-speaking children between the ages of 2 and 8 years. Many ways were chosen to assess validation of the test such as: factorial, judgmental, face, contrasted and internal consistency validities. Test retest was used to assess reliability. Since then, it has been used to assess the language of children in Arab countries. Refaei and Hassan, 2004 revised the test^[7]. Rearrangement of various items of the original test was done from the simplest to the most difficult, and a new scoring system was implemented. Reliability and validity measurements were performed again. Although the test is holistic, the phonological and pragmatic testing included few items and were not in details. Syntactic evaluation for older age was not available in the original test.

In a previous recent study^[8], a second revision was started on children aged 2-4 years. In this study we aim at revising the test for the other age ranges from 4-8 years to overcome the limitations of the test mentioned before.

PATIENTS AND METHODS:

2.1. Ethical consideration: This study was approved by the Research Ethics Committee for experimental and clinical Studies of the Faculty of Medicine Ain Shams University. A written consent has been obtained from the parents of all children included in the study. Patient privacy and confidentiality were protected. Deceptive practices were avoided during designing the research. The participants had the right to withdraw from the study at any time they wished.

2.2. Study design: The study is a mixed (qualitative and quantitative) sequential exploratory study; it is an approach that combines the qualitative phase of the test preparation and the pilot study, with the quantitative phase that analyze and examines the psychometric properties of the test. Normal children with no language disorders and patients with delayed language development attending phoniatric unit were included in the study.

2.3. Pilot study: After preparing the new design and new photos of the test format (deliberately prepared and clear, all taken by a dedicated photographer), a pilot study was conducted on 60 normal children (with no language disorders). There were 15 subjects in each group. Within the same age range as the standardized sample. Based on the results of the pilot study, to arrive at the final form of the test, many changes were made to the photos and the wording.

2.4. Standardization study: This was applied on 720 Egyptian children. Children were divided into four groups, each 180 children; the first group (group I-A) with age range 4-5 years. The second group (group I-B) ranged in age from 5-6 years. The third group (group I-C) ranged in age from 6-7 years. The fourth group (group I-D) ranged in age from 7-8 years. They were randomly selected from different public and private nurseries.

Inclusion criteria included Egyptian, Arabic native speaking children, children reported by their teachers or caregivers to have diagnosed or undiagnosed communication disorders were excluded. Average intelligence as proved by Stanford Binet Intelligence scale 5th edition^[8].

2.5. Study of reliability: A sample of 60 children (15 in each group). The children were selected randomly to be retested after 2 weeks

2.6. Study of validity: This included 80 children with language delays. These children were selected randomly in the same age range 4-5 year old (group II-A), 5-6-year-old (group II-B), 6-7 year old (group II-C) and 7-8 year old (group II-D). All of them were Egyptian children and were diagnosed as cases of delayed language development.

2.7. Test application: Testing procedures: Test was presented in Arabic Egyptian dialect. Children had the items of the test according to their age in a suitable quiet environment. 15- 20 minutes is the expected period to complete the test.

Descriptive and inferential statistical analysis was done. The data collected, tabulated and analyzed using the SPSS (Statistical Package for Social Sciences) version 20. Means and standard deviations presented the quantitative variables. Validation was done by (Internal consistency, contrasted group validity, judgment and face validity). Reliability testing with test- retest, Pearson's correlation test, P was considered significant when $P < 0.05$. Split-Half method (correlation between forms, Spearman-Brown coefficient, Gutman Split-Half coefficient and Alpha Cronbach).

RESULTS:

3.1. Statistics of the results of the four age groups:

1. Raw scores: Table (1), Table (2), Table (3) and Table (4) show the raw scores, means and standard deviations among the four groups.

Table 1: means and standard deviations of raw scores of the Arabic language test of group I-A the age range of 4-5 years:

Subtests	Range		Mean	SD
	Min	-		
Semantics	20.00	-	24.6	0.818
Receptive syntax	12.50	-	23.8	2.18
Expressive syntax	13.00	-	28.11	2.94
Pragmatics	13.00	-	23.3	2.5
Phonology	14.00	-	23.9	2.01
Total	84.00	-	123.6	7.76

Table 3: means and standard deviations of the raw scores of the Arabic language test of group I-C the age range of 6-7 years:

Subtests	Range		Mean	SD
	Min	-		
Semantics	9.00	-	12.7443	.72475
Receptive syntax	8.00	-	12.2874	1.11125
Expressive syntax	6.00	-	12.0287	1.53731
Pragmatics	5.00	-	9.3678	1.15793
Phonology	9.00	-	14.9052	.63921
Total	40.00	-	56.3506	3.20189

Table 4: means and standard deviations of the raw scores of the Arabic language test of group I-D the age range of 7-8 years:

Subtests	Range		Mean	SD
	Min	-		
Semantics	7.50	-	9.9861	.18634
Receptive syntax	7.00	-	9.4722	1.63902
Expressive syntax	7.00	-	9.8056	.60811
Pragmatics	4.00	-	9.3667	1.28550
Phonology	9.00	-	9.9028	.29473
Total	36.00	-	48.5444	2.56442

3.2. Validity testing:

1-Internal consistency.

2-Contrasted group validity.

3-Judgement validity.

4- Face validity.

1-Internal consistency

It is a measure of homogeneity of the test contents. The internal structure of the Arabic language test was examined by making correlation between:

- The total score of each item and the total score of the whole test.

Using Pearson's correlation, as shown in the following tables. Correlation coefficients for internal consistency were (statistically significant), and this proves the strong internal consistency of the test (see tables: 5, 6, 7 and 8).

Internal consistency proved that all test items are valid

Table 5: Group I-A (Age:4-5yrs): Correlations between each item of the test and the total score of the whole test.

Test items	R	P value	significance
Semantics	.385**	.000	S
Receptive syntax	.667**	.000	S
Expressive syntax	.812**	.000	S
Pragmatics	.731**	.000	S
Phonology	.473**	.000	S

Table 6: Group I-B (Age:5-6 years): Correlations between each item of the test and the total score of the whole test.

Test items	R	P value	significance
Semantics	.243**	.001	S
Receptive syntax	.382**	.000	S
Expressive syntax	.472**	.000	S
Pragmatics	.579**	.000	S
Phonology	.588**	.000	S

Table 7: Group I-C (Age:6-7years): Correlations between each item of the test and the total score of the whole test.

Test items	R	P value	significance
Semantics	.502**	.235	NS
Receptive syntax	.691**	.000	S
Expressive syntax	.825**	.000	S
Pragmatics	.616**	.000	S
Phonology	.146	.054	S

Table 8: Group I-D (Age:7-8 years): Correlations between each item of the test and the total score of the whole test.

Test items	R	P value	significance
Semantics	.118	.114	NS
Receptive syntax	.826**	.000	S
Expressive syntax	.324**	.000	S
Pragmatics	.692**	.000	S
Phonology	.183*	.014	S

2- Contrasted group validity: it was done by making correlation between children with no language delays (group I-A, group I-B, group I-C, and group I-D) and children with delayed language development (group II-A,

group II-B, group II-C and group II-D), (see tables 9, 10, 11 and 12).

Contrasted group validity proved that all items are valid.

Table 9: Comparison between the scores of the Revised- Arabic Language Test in no (Group I-A) and (Group II-A) [age range:4- 5 years]:

Subtests	Group I-A "n=180" X ± SD	Group II-A "n=20" X ± SD	T	Sig. (2-tailed)	Significance
Semantics	24.6±0.818	13.4±10.7	13.9	0.000	S
Receptive syntax	23.8 ± 2.1	9.2 ± 9	17.8	0.000	S
Expressive syntax	28.11± 2.94	3.5 ± 6.11	30.825	0.000	S
Pragmatics	23.3±2.5	2.85±4.8	31.4	0.000	S
Phonology	23.9± 2.01	8.7 ± 8.9	19.159	0.000	S
Total	100.5± 5.4	35 ± 27.9	27.4	0.000	S

1- X±SD: mean and standard deviation

2- t: t-Test of equality of means

3- p<0.05: significant

4- S: significant

5- n: number

Table 10: Comparison between the scores of the Revised- Arabic Language Test in (Group I-B) and (Group II-B) [age range:5- 6 years]:

Subtests	Group I-B "n=180" X ± SD	Group II-B "n=20" X ± SD	T	Sig. (2-tailed)	Significance
Semantics	17.8±.511	10.7±7.3	12.906	0.000	S
Receptive syntax	19.3± 1.2	4.1 ± 6.2	22.127	0.000	S
Expressive syntax	19.2±1.9	5.3±6.8	21.227	0.000	S
Pragmatics	19.3± 1.3	0.50± 0.827	63.069	0.000	S
Phonology	19.4±1.75	10.15 ± 8.2	12.913	0.000	S
Total	95.2±3.7	34.8± 26	29.083	0.000	S

1- X±SD: mean and standard deviation

2- t: t-Test of equality of means

3- p<0.05: significant

4- S: significant

5- n: number

Table 11: Comparison between the scores of the Revised- Arabic Language Test in (Group I-C) and (Group II-C) [age range:6- 7 years]:

Subtests	Group I-C "n=180" X ± SD	Group II-C "n=20" X ± SD	T	Sig. (2-tailed)	Significance
Semantics	12.7±.72	7.5±5.3	12.092	0.000	S
Receptive syntax	12.2±1.1	6.3±4.7	13.791	0.000	S
Expressive syntax	12.02±1.5	3.7±5.08	16.190	0.000	S
Pragmatics	9.3±1	2.7±4	16.845	0.000	S
Phonology	14.9±.63	8.6±6.4	12.462	0.000	S
Total	56.35±3.2	26.25±18.9	19.075	0.000	S

1- X±SD: mean and standard deviation

2- t: t-Test of equality of means

3- p<0.05: significant

4- S: significant

5- n: number

Table 12: Comparison between the scores of the Revised- Arabic Language Test in (Group I-D) and (Group II-D) [age range: 7 years - 8 years]:

Subtests	Group I-D "n=180" X ± SD	Group II-D "n=20" X ± SD	T	Sig. (2-tailed)	Significance
Concepts	9.9±.18	3.3 ±4.6	19.8	0.000	S
Receptive syntax	9.4 ± 1.6	1.4 ± 2.6	19.5	0.000	S
Expressive syntax	9.8 ± 0.6	3.3 ±5.3	15.8	0.000	S
Pragmatics	9.3±1.2	1.7±2.5	22	0.000	S
Phonology	9.9 ± 0.3	4.8 ± 4.2	16	0.000	S
Total	48.5 ± 2.56	14.5 ± 14.9	28.7	0.000	S

1- X±SD: mean and standard deviation

2- t: t-Test of equality of means

3- p<0.05: significant

4- S: significant

5- n: number

3- Judgement validity: results were some changes in the material used and patterns of questioning of some items, as well as the order of presentation of test items. The results showed that the test is fit and valid.

4- Face validity: from the superficial point of view the test appears to be valid, since it measures various domains of language including syntax, semantics, pragmatics and phonology, with all details receptively and expressively.

3.3. Tests of Reliability

(A) Reliability coefficient:

It is a quantitative expression of the reliability or consistency in the measurement of test scores.

It was measured by

1-Test re-test method (Pearson correlation).

Table 13: Pearson Correlation for testing test retest reliability analysis of the Arabic language test group I-A, I-B, I-C and I-D:

Groups "n="	Pearson Correlation	Sig. (2-tailed)	Significance
Group I-A	.952**	.000	Significant
Group I-B	.712**	.003	Significant
Group I-C	.973**	.000	Significant
Group I-D	.914**	.000	Significant

**. Correlation is significant at the 0.01 level (2-tailed).

2-Split half method:

Internal consistency for the Arabic language test for the entire sample of group (I-A, I-B, I-C and I- D) is

2- Split half method (correlation between forms, spearman-Brown coefficient).

3- Alpha Cronbach.

1- Test re-test method:

Test-retest reliability estimation based on the correlation between two (or more) administrations of the same scale with known interval. This was performed on 15 children each group. These children were tested by the Arabic language test and then re-evaluated by the same test after 2 weeks interval. The correlation between the scores of both tests was done (by Pearson correlation) and the results are shown below in (Table 13). The entered variables in the calculations are all the variables of the Arabic language Test in the first time testing and after repetition of the test for the second time.

shown in (Table 14) as correlation Between Forms and Spearman-Brown Coefficient, revealing high consistency and reliability (see table 14).

Table 14: Internal consistency of all test groups:

Groups	Correlation Between Forms	Spearman-Brown Coefficient	
		Equal Length - Unequal Length	
Group I-A	.654	.791	.791
Group I-B	.435	.606	.606
Group I-C	.429	.600	.600
Group I-D	.623	.768	.768

3- Internal consistency reliability using Cronbach's alpha coefficient; Cronbach's alpha values for the Arabic language test for the entire sample of group (I-A) aged 4-5 years, for group (I-B) aged 5- 6 years were .652, for group

(I-C) aged 6-7 years were .557 and for group (I-D) aged 7-8 years were .616 indicating high reliability of this tool as shown in (Table 15).

Table 15: Items' internal consistency coefficients for each age group:

Age (years)	Number of children	Number of items	Alpha
4-5 years	180	39	.810
5-6 years	180	26	.652
6-7 years	180	18	.557
7-8 years	180	16	.616

-Alpha Cronbach

DISCUSSION

Delayed Language development, that is a common developmental disorder among school-age children (4-8-year-old), may be sometimes mis-diagnosed. Delayed language developments mostly result from genetic, biological and cognitive weaknesses as well as environmental deprivation^[10]. Assessing the ways in which the various components of language (semantics, syntax, pragmatics and phonology) are affected by this disorder is always challenging. This study presents a completion of second revision of the Revised- Arabic Language Test (ALT), but for older children^[9]. Studies showed that nearly half of children with delayed language development can catch up to children with typical language development by the end of their preschool years but unfortunately, the other half may have persistent language delays and will be diagnosed as delayed language development^[11]. This makes it mandatory to revise/establish newly designed or updated versions of previously used tests to assess the defective points in detail among this population. Test revisions are usually done for different purposes as to update the items or materials included, to extend items to cover bigger age ranges, to add items that would give better evaluation of a certain modality or to increase size of the sample that would yield better reliability and validity of the test. Among the tests used for older children is the "The Test for Reception of grammar"^[12], it was first standardized for children aged 4-12 and then revised to extend up to 16 years old children and was also available electronically. It reported good internal consistency but fair concurrent validity. The CELF-5 was revised^[4], on 2013 where the test evaluated higher levels of language as it offered a more robust assessment of pragmatics using observations and interactive activities with current normative data reflecting recent diverse population.

Most Arabic tests used for standardized language assessment rely on the translation of foreign tests (for example, the children's Arabic token test^[13]), or the Arabicization and modification of others

(for example, the modified preschool language scale 4")^[14]. No one Arabic test was constructed de novo for the Arabic Egyptian culture as is the case for The Arabic Language Test^[6]. It was revised for children with age range 2-4-year-old in a previous recent study^[9], and it's revised for older children 4-8 year old in the current study.

The uniqueness of the currently revised ALT is that it measures all language fields more comprehensively, especially pragmatics and phonology. The items of pragmatics include pictures of stories that stimulate different pragmatic skills. While the items testing the phonology were designed to determine the presence of phonological processes as well as phonetic errors.

Its revision was meant to improve the scope of its measuring ability, to update its materials used and to get new normative data based on a bigger number of subjects. The pictures being all renewed were mounted on a book in a PDF form that when presented to the child the wording of the item is shown to the examiner. This new format of the test is easier to deal with and more pleasurable for the child tested. The raw data were converted to standard scores in the form of t-scores to obtain cut off points to every domain of the language test for each age range. Also the raw scores can be converted to language ages and language percentages.

In terms of reliability, the internal consistency survey examines the interrelationships between items of the same general structure, and the size of alpha provides strong support for the reliability of the test^[15]. In the current study, all the coefficients obtained from reliability measures are very high. Alpha Cronbach is very high between 0.5 and 0.8, while the retested Pearson coefficient is between 0.7 and 0.9. These findings indicate the accuracy of the items and the stability of the scores. The result of the validity measurement proves the homogeneity of the test content, while the highly significant measures of contrasted group denotes the ability of the test to differentiate among groups.

Although the old ALT had strong psychometric properties, the ALT revised in this study has been tested for validity and reliability through various methods, and the results showed high significance, thus rendering the new test psychometrically sound too.

CONCLUSION

It is obvious that the currently revised Arabic Language test is well qualified for good and accurate child language evaluation that helps to establish a treatment plan for those with language development delays.

ABBREVIATIONS

Arabic Language Test (ALT)

The clinical evaluation of language fundamentals (CELF)

CONFLICT OF INTEREST

There are no conflicts of interest.

REFERENCES

1. Denman D., Speyer R., Munro N., Pearce W.M., Chen Y. and Cordier R. (2017). Psychometric properties of language assessments for children aged 4-12 years: A systematic review. *Front. Psychology*, 8: 1515.
2. Dodkrell J.E., Marshall C.R. (2015). Measurement issues: assessing language skills in young children. *Child-Adolescent Health* 20,116-125.
3. Mokkink L.B., Terwee C.B., Patrick D.L., Alonso J., Stratford P.W., Knol D.L. (2010). The COSMIN study reached international consensus on taxonomy, terminology and definitions of measurement properties for health-related patient-reported outcomes. *Qual. Life Res.* 19, 539-549.
4. Wiig E.H., Semel E. and Secord W.A. (2013). Clinical evaluation of language fundamentals. 5th edition. Bloomington, MN: Pearson Psychograph.
5. Petty, R. E, Briñol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behaviour* (pp. 318–329). New York, NY: Guilford Press.
6. Kotby, M.N., Khairy A., Barakah M. Rifaie N. and El-Shoubary A. (1995). Language testing of Arabic speaking children. Proceedings of the XXIII World Congress of the International Association of Logopedics and Phoniatrics. Cairo.
7. Rifaie N. and Hassan S. (2004). The Arabic Language Test Revised. *Benha Med. J.* Vol. 21, No. 2, 205-216.
8. Bain, S.K. and Allin, J.D. (2005). Book review: Stanford –Binet intelligence scales, fifth edition. *J Psychoeducational Assess.* 23(1):87 –95. <https://doi.org/10.1177/073428290502300108>
9. Rifaie, N., Hamza, T.M.A.W. & Elfiky, Y.H (2021). Validity and reliability of the revised Arabic language test for 2–4-year-old children: cross-sectional study. *Egypt J Otolaryngol* 37, 26 . <https://doi.org/10.1186/s43163-021-00088-8>
10. Paul, R. (2020). Language disorders. *Handbook of Clinical Neurology*.174, 2020, 21-35.
11. Matte-Landry, A., Boivin, M., Tanguay-Garneau, L., Mimeau. C., Mara Brendgen, M., Vitaro, F., Tremblay, R. and Ginette Dionne, G. (2020). *Journal of Speech, Language, and Hearing Research*.63(11):3760-3774.
12. Bishop D.V.M. (2003). *Test for reception of grammar*. 2nd edition. London: Psychological Corporation.
13. Alkhamra R.A. and Aljazi A.B. (2016). Validity and reliability of the Arabic Token Test for children. *Int. J. Lang. Commun. Disord*, Vol.51, No. 2, 183-191.
14. Abouhasseba A. (2011).Standardization, translation and modification of the preschool language scale-4.Thesis submitted to Phoniatic Unit Ain Shams University. Cairo.
15. Portney L.B. and Watkins M.P. (2009). Foundations of clinical research: Applications to practice, 3rdedition (Upper Saddle River, NJ: Pearson Prentice Hall).