

Persian Articulation Assessment for Children Aged 3 - 6 Years: A Validation Study

Talieh Zarifian,¹ Yahya Modarresi,² Laya Gholami Tehrani,¹ Mehdi Dastjerdi Kazemi,³ Mahyar Salavati,¹

Amir Sadeghi,⁴ and Soheila Shahshahani^{5,*}

¹Department of Speech Therapy, University of Social Welfare and Rehabilitation Sciences (USWR), Tehran, IR Iran

²Research Institute of Humanities and Cultural Studies, Tehran, IR Iran

³Research Institute of Exceptional Children, Research Institute of Education, Tehran, IR Iran

⁴Language and Literacy Research Lab, School of Teacher Education, College of Education, University of Canterbury

⁵Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences (USWR), Tehran, IR Iran

*Corresponding author: Soheila Shahshahani, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Kudakyar St, Daneshjoo Blvd, Tehran, IR Iran. Tel: +98-21-22180142; +98-9123706003, Fax: +98-2122180142, E-mail: sol_shah@yahoo.com

Received 2016 August 11; Revised 2017 June 02; Accepted 2017 July 09.

Abstract

Objectives: The present study aimed to adapt articulation assessment, a subtest of the diagnostic evaluation of articulation and phonology, and to determine its reliability and validity for Persian speaking children.

Methods: The Persian version of articulation assessment (PAA) was administered to 387 children aged between 36 - 72 months, M (SD):53.7(±10.1) by month, after the adaptation process. A methodological study including test-retest reproducibility, score-rescore consistency and evaluating validity (through content, convergent and discriminative validity) was then carried out in order to determine the psychometric properties of the instrument.

Results: Content validity ratio for Persian item's content coverage, agreement image and syllable structure were 0.86 - 1, 0.92 and 0.94, respectively. Minimum content validity index of 0.93 exceeded in terms of relevancy, simplicity and clarity of instructions. The percentage agreement for the test-retest was 91.35-100% and the score-rescore analysis was 92.95-100%. The convergent validity was reasonable. The Persian Articulation Assessment's mean scores for individuals with articulation disorders being significantly lower than those by normal children, showed discriminative validity ($t = 7.245$, $df = 34$, $P < 0.001$).

Conclusions: The Persian version of Articulation Assessment is suggestive of a reliable and valid instrument for evaluating the articulation skills in Persian speaking children.

Keywords: Articulation Assessment, Validity and Reliability, Persian, Children

1. Background

Speech sound disorder in children is one of the most common forms of communication disorders and its prevalence varies between 10% and 15% in preschool children and 6% among school-age children (1). Some professionals, particularly general practitioners, sometimes advise parents of preschool children that they will grow out of a speech disorder, and no intervention is required; but some pediatric speech and language therapists believe that intervention should be offered as early as possible because it is more cost-effective to shape a developing system (2). Since speech and language development is relatively associated with all aspects of social and educational development it is important to make accurate diagnosis and assessment of speech disorder in children, in order to prevent educational, psychosocial and communication difficulties in the future.

If in routine developmental screening, a child doesn't pass the test, the clinician will refer him/her to a speech

and language pathologist (SLP) for more detailed evaluation (3). In recent years, a large number of measurement tools has been developed to assess the articulation competence in children (4, 5). Single word data gathered through the administration of the picture naming articulation test, provide an efficient and relatively easy method to elicit sounds produced by a child (6-8). Clinicians mostly rely on results collected from standardized assessment tools that provide quantitative information to assist with the eligibility determination and the intervention planning (9-11). The appropriate selection of instruments for the outcome measurement depends on many factors including type and psychometric properties of the instrument and characteristics of the subjects among whom the instrument is intended to be used (12, 13). Additionally, it is recommended to make a comprehensive evaluation of impairment, disability, and handicap in individuals with communication disorders including children with the articulation/phonological impairment and to scaffold un-

derstanding of typical speech development (1, 4).

The sound system of the Persian language (sometimes known as Standard Persian) has some differences with English, some of them are brought in the appendix 1. Based on these differences, using foreign language screening or diagnostic tests is not suitable for evaluating phonological disorders in Persian speaking children. Despite the unanimous agreement among specialists on an approved test to be used in a comprehensive speech and language evaluation (4), scarcity of such measures in Persian language is evident for clinicians in Iran and the neighboring Persian speaking countries (14). Due to the lack of reliable and valid assessment tools for evaluating all aspects of the articulation skills of Persian speaking children, Iranian SLPs typically use informal or English-based instruments (15). In the field of evaluating articulation skills for Persian speaking children, clinicians mostly utilize only one traditional test called Phonetic Information Test (PIT) of which internal consistency and test-retest reliability were studied by Ghassisin et al (2013). Despite PIT's capabilities, it is not able to evaluate the articulation of vowels. In addition, it does not provide an opportunity for testing the stimulability of consonants (16).

On the other hand, regarding the immigration issue (11, 17), there are a few data regarding the language use, articulation and phonological skills or disorders of children who are under 6 years old and live in other countries and use Persian language at home. The necessity for valid tests in order to gather such data is obvious.

In a review for selecting an appropriate model for eliciting phonemes (consonants and vowels), articulation assessment of diagnostic evaluation articulation and phonology (DEAP) battery meets most of the criteria for a valuable instrument such as a clear definition of the test domain, evidence of validity and reliability, detailed description of test administration, detailed description of test user qualifications and quick administration. DEAP is a comprehensive, individually administered, norm-referenced battery designed to provide differential diagnoses of speech disorders in children. Five tests (two screens and three assessments) comprise the DEAP assessment process. From the diagnostic screen results can be determined if additional testing is needed and, if so, the appropriate DEAP test to administer. The articulation, phonology, and word inconsistency assessments then help clinicians differentiate between disorders of articulation; delayed phonological development and consistent phonological disorders; and inconsistent phonological disorders, respectively. If there are concerns about a possible oral motor disorder, the oral motor screen may be administered to determine if an in-depth assessment of oral motor skills is warranted. DEAP battery contains

qualitative and quantitative measures. Articulation and error pattern analysis as part of Phonology tests are qualitative measures for evaluating child's phonetic consonant inventory and error patterns. Oral motor, inconsistency and quantitative part of phonology tests are quantitative measures (2, 4).

2. Objectives

The present study reports the process of adaptation and psychometric properties of the articulation assessment of DEAP battery in Persian speaking children aged between 3 - 6 years in Tehran. This age range was chosen because it reflects the age group of most children with speech disorders (18).

3. Methods

3.1. Materials

Original test: the articulation assessment measure examines children's ability to produce individual speech sounds within words or in isolation by establishing the child's phonetic inventory. The assessment consists of two parts: the picture naming articulation and speech sound stimulability which require a child to name pictures and produce one consonant in the chains of phonemes presented in various syllable structures or in isolation.

The phonetic information test (PIT) utilized for evaluating the convergent validity. The internal consistency and test-retest reliability of the PIT were 0.79 and 0.85 respectively (16).

A voice recorder (COBY, model: MPC-7405), a laptop (Sony, model: VAIO), and the SPSS statistical software version 19 were used.

3.2. Participants

A total of 387 children (191 boys and 196 girls), aged 36 - 72 months, in 12 nurseries and kindergartens in Tehran were recruited after obtaining their parents or guardians consent following ethics approval from the medical ethics committee for the University of Social Welfare and Rehabilitation Sciences. Children were selected through a simple convenience sampling.

Only monolingual Persian speaking children with no background of speech and language impairment who could tolerate the duration of the test and attempt to imitate and follow cuing were included. The exclusion criteria were structural deficits (e.g., cleft palate), permanent hearing loss, speaking Persian as a second language at home, autism spectrum disorder and dysarthria. These were determined by the child's medical record history, a clinical

examination by an experienced speech language pathologist and reports from parents and teachers in nurseries and kindergartens. Participants were tested in a quiet place (6, 18).

Testing took place in the child's nursery or kindergarten for the total of 7 to 10 minutes depending on each participant's attention span and desire to continue. Children were required to name color pictures and were given verbal praise (e.g., Good job, Nice, Well-done, etc.), physical praise (e.g., high fives) and tangible reinforcements (e.g., stickers) for participating in the assessment.

A broad phonetic transcription was made online after the production of any words. Further, all testing procedures were audio-video recorded. When the examiner did not hear the child's production clearly, the child was asked to repeat the word. The examiner provided cues or a model for imitation if the child was unable to name a picture. As soon as the child imitated the target word, letter 'i' was inserted in front of the word to show the manner of elicitation ('i' stands for imitation).

3.3. Methods

Following permission for adapting the DEAP from the developer, Prof. Barbara Dodd, and the publisher, Pearson Inc., to ensure cross-cultural translation-adaptation, a standard procedure including multiple forward and backward translations and qualitative (via expert panel sessions for cognitive debriefing) and quantitative (measuring CVR and CVI) evaluation of the translations were followed, and necessary adaptations were implemented (19). In this article, validation processes of the articulation assessment were described. Content validity ratio (CVR) and content validity index (CVI) were measured for all individual items and the instrument as a whole (19, 20).

Translation and adaptation of the instrument followed a multistage procedure. An official permission for the questionnaire was obtained from both test developers and the publishing company. Then an English translator, a speech and language pathologist and a linguist (all Persian native speakers), translated the questionnaire independently. In order to avoid ambiguity and cultural issues, these independent translations were presented to a panel of experts to finalize the translation. To correspond with Persian linguistic properties and the Iranian culture (the context where the test was standardized), a pool of Persian items was developed consistent with the global outline of the articulation assessment of DEAP battery. The DEAP manual (21) provided criteria for generating items for the articulation test. It also provided the theoretical basis for developing the picture naming test (8, 22). All words were adopted from the Persian core vocabulary for Iranian children (23), picture-riddle dictionary

(24), children's story books, Persian dictionary for school children (25), Persian translation of McArthur-Bates communicative development inventory (26) and available Persian phonetic tests. Then generated items were attached to the primary translation. Two linguists and two speech-language pathologists evaluated the quality of translated instructions and assigned items in the preliminary Persian version of the instrument. A native English/Persian bilingual speaker with linguistic background back-translated the instrument from Persian into English. The second author rated the quality of forward-backward translation in terms of clarity, common language use and conceptual equivalence. The back translation was submitted to the Pearson Inc. to test the equivalence of back-translated version with the original version. After that the preliminary form of the instrument was reviewed by a panel of experts (four speech therapists and three linguists) who rated the relevancy of the instrument (content coverage) in terms of the items' syllable structure, items' familiarity for the targeted age group, positions of consonants and vowels in the words, image agreement and transparency of the items, and finally the instructions' relevancy, simplicity and clarity. This involved the use of a 100-point rating scale, ranging from completely undesirable (0) to completely desirable (100) and a box for comments. Backward translation resulted in no major linguistic or cultural concerns. Finally, a pilot study was conducted with 60 participants (male-female ratio = 1:1) in six various age groups $M (SD): 54.1 (\pm 11.1)$ in months. During the pilot study, three pictures seemed a bit ambiguous requiring further description to elicit the target word. These were as follows: the word /bɑd/ (meaning wind) with the prompt sentence asking, 'What makes the leaves blow?' [bærge deræxtɑro i mibære?]; the word /nej/ (meaning straw) with the prompt sentence: [ʃiro bɑ i mixori?]: 'What would you drink your milk with?', and the word /ri/ (meaning beard) with the prompt sentence: [ruje suræte mærdhɑ i dær miɑd?]: 'What grows on a man's face?'

The pilot data of the study led to the development of the final draft of the Persian articulation assessment (PAA). The sounds elicited cover all consonants of the targeted syllable in the initial and final positions along with all vowels. A stimulus list was also provided for eliciting speech sounds which were not produced at the previous stage.

In this study, reliability was assessed through test-retest reliability and score-rescore consistency (27). Test-retest analyses were reported for 52 children (13.4% with the mean age of 53.3 months) who were able to return for the re-administration of the PAA within 1 - 3 weeks after their initial test. For Score-rescore reliability (Consistency) two independent examiners who had not been involved in the PAA's development rescored audio-video recordings

of 70 children being randomly selected for the interrater (score-rescore) reliability analyses (18.8% with the mean age of 54.1 months).

Audio-video recordings were made through the assessment procedure to allow the revision of online transcription difficulties and transcription reliability measurement. For score-rescore reliability two independent examiners who had not been involved in the PAA's development, reviewed and rescored some transcription with reference to its audio-video recordings. The current study has utilized the Kappa statistics usually being used as a measure of reproducibility between repeated assessments of the same variable (28). Kappa amounts which are over 0.75 denote great reproducibility (29-31).

The evidence of construct validity was provided by a priori hypothesis patterns of association with other measures (the convergent validity with PIT) and evaluating the discriminative validity. PIT was used for evaluating convergent validity as an evidence of construct validity. For this part of study, the results of PIT and PAA were compared.

For discriminative validity a sample of 36 children aged between 3 - 6 years with and without articulation deficit (18 participants in each group, respectively whose diagnosis was approved by three experienced speech language pathologists) enrolled in this part of study. An independent t-test was utilized to analyze the difference between the two groups. Alpha level of 0.05 was considered for all statistical procedures.

4. Results

A total of 387 children (191 boys and 196 girls), aged 36 - 72 months enrolled in this study from April to September 2013. Table 1 reports the demographic data of the participants.

Table 1. Descriptive Statistics of Participants in Six Age Groups (n = 387)

Age Group, mo	No. (%)	Mean, mo \pm SD
35 - 42	60 (15.5)	39 \pm 1.8
43 - 48	82 (21.2)	46 \pm 1.5
49 - 54	60 (15.5)	51 \pm 1.6
55 - 60	68 (17.6)	57 \pm 1.7
61 - 66	62 (15.7)	64 \pm 1.7
67 - 72	56 (14.5)	69 \pm 2.2
Total	388 (100)	53.7 \pm 10.09

Abbreviation: SD, standard deviation.

After refining the preliminary form, most of the items and the whole instrument gained an appropriate CVR and

CVI (CVR \geq 0.86 and CVI \geq 0.93). Results for the reliability analyses are reported in order for the test-retest reliability and inter-rater reliability. For each type of the reliability, the percentages of agreement were reported as a measure of correlation among repeated measurements from the same participant. In these evaluations, all consonants were measured in two positions (initial and final) and all vowels in one position (middle).

We found that plosives (oral/nasal) obtained higher agreement compared to the other consonants (e.g., fricatives, affricates). Additionally, the vowels were all consistent in both test-retest and score-rescore reliability evaluation. The agreement percentage on the production of vowels in both test-retest and score-rescore evaluations was 100%.

Construct validity was evaluated in this study via convergent and discriminative validity. For evaluating the convergent validity, the PAA demonstrated significant correlation with the PIT (rkappa = 0.78, $P < 0.001$). To assess discriminative validity, the mean score of the PAA between the two groups of children with/without articulation deficit was compared by t-test. The results ($t = 7.24$, $P < 0.001$) showed the mean score of PAA for children without articulation deficit ($M = 27.5$, $SD = 0.28$) was significantly higher than for the children with articulation deficit ($M = 20.3$, $SD = 0.96$). There is a profile of articulation errors for children with articulatory deficits in appendix 5.

5. Discussion

Establishing a phonetic inventory is important because this can be used as the basis for determining the developmental appropriateness of a child's sound production (32), identifying targets for stimulability testing (33, 34), and/or selecting goals (8, 35, 36) or treatment targets (8). It is also important to determine if the content of a test is sufficient to determine a child's phonetic inventory.

The purpose of the current study was to develop a Persian version of the articulation assessment of the DEAP package - developed initially in English - and to provide evidence for reliability and validity of Persian articulation assessment (PAA) measure for young children. The PAA has been chosen because it can provide easily recognizable items with opportunities of phonetically-controlled contexts that would neither confound nor scaffold the child's ability to produce a speech sound (19, 22, 37). All occurrences of each consonant and vowel were evaluated. The PAA was assessed for its content coverage. In the PAA, all consonants occurred at least in two positions (initial and final) except /ʔ/ which occurred only twice in an initial position. As mentioned before, in Standard Persian, usually final /ʔ/ is deleted. Content relevancy and content coverage

were also examined by looking into the inclusion of consonants and vowels in phonetically-controlled contexts. Results of the present study showed that the PAA possesses a desirable CVR and CVI. PAA is a reliable and valid tool. It is relatively quick and practically easy to administer and score, its time of administration is between 10 - 15 minutes.

Similar to the English version, the test-retest agreement range for consonants in this study suggests a high level of reliability (91.35% - 100%) or a relatively low level of error due to differences in the child's performance and the clinician's judgment over two test administrations. This finding was valuable dealing with the difficulties intrinsic to this type of testing, difficulties related to variations in what concerns the child, such as attention and mood, as well as variations in what concerns the clinician, such as judgment focus and interpretation of the child's response (27, 38, 39).

In line with the English instrument, the adapted version of the PAA has obtained a high score-rescore agreement percentage (92.95% - 100%). This result suggests that the clinicians consistently assessed a child's performance on the PAA with few differences across examiners, even when a set of assessment was made from an audio-video recording of the child's performance. Given that clinicians were able to score the performance of the participant consistently across live versus taped contexts, it can be inferred that the test can be scored online or from an audio-video recording based on the demands of clinical contexts. Thus pediatricians would be able to send their patient's sounds to a SLP for online evaluation and depend on the result, refer him/her for receiving more detailed evaluation and/or intervention. Moreover, this finding is important because two of the independent examiners in the study had not been involved in the PAA's development and had received almost four hours of training on test administration which again proves the user-friendliness of the measure. Thus, the demonstration of high inter-rater agreement among examiners suggests the feasibility of the test's use by clinicians producing consistent results.

The data support the acceptable intra and inter-rater reliability (test-retest and score-rescore) and also provide evidence for supporting the validity of the PAA for the diagnostic purposes which suggests that the measure can reliably discriminate children with/without articulation disorders.

A priori hypothesis was confirmed with agreement above 0.7 between the PAA and the PIT. This reasonable coefficient between two instruments shows convergent validity. Furthermore, the PAA demonstrated reasonable power to discriminate individuals with/without articulation disorders. One of the major limitations of the present study has been the lack of data to evaluate results of the

PAA in various demographic characteristics in various populations (i.e., age, gender and socio-economic status). Further studies are required to investigate the psychometric properties of the Persian version of the articulation measure in DEAP battery.

5.1. Conclusions

In conclusion, a satisfactory level of test-retest and scoring-rescore reliability and construct validity was obtained for the Persian version of the articulation assessment of DEAP in the Persian language. Therefore, it would be appropriate to evaluate consonants and vowels usually used by 3 - 6-year-old Persian speaking children in Iran.

Supplementary Material

Supplementary material(s) is available [here](#).

Acknowledgments

The authors would like to thank professor Barbara Dodd for her kind support especially at initial steps of developing the Persian version of DEAP and the Pearson publishing Co. for granting permission to adapt the English test into Persian for research purposes. This article is part of a PhD thesis approved in University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

References

1. McLeod S. Speech pathologists' application of the ICF to children with speech impairment. *Adv Speech Language Pathol.* 2009;**6**(1):75-81. doi: [10.1080/14417040410001669516](https://doi.org/10.1080/14417040410001669516).
2. Dodd B. Differential diagnosis and treatment of children with speech disorder. John Wiley & Sons; 2013.
3. Shahshahani S, Vameghi R, Azari N, Sajedi F, Kazemnejad A. Validity and Reliability Determination of Denver Developmental Screening Test-II in 0-6 Year-Olds in Tehran. *Iran J Pediatr.* 2010;**20**(3):313-22. [PubMed: [23056723](https://pubmed.ncbi.nlm.nih.gov/23056723/)].
4. McLeod S. An holistic view of a child with unintelligible speech: Insights from the ICF and ICF-CY. *Adv Speech Language Pathol.* 2009;**8**(3):293-315. doi: [10.1080/14417040600824944](https://doi.org/10.1080/14417040600824944).
5. Abou-Elsaad T, Baz H, El-Banna M. Developing an articulation test for Arabic-speaking school-age children. *Folia Phoniatr Logop.* 2009;**61**(5):275-82. doi: [10.1159/000235650](https://doi.org/10.1159/000235650). [PubMed: [19696489](https://pubmed.ncbi.nlm.nih.gov/19696489/)].
6. Dodd B, Holm A, Hua Z, Crosbie S. Phonological development: a normative study of British English-speaking children. *Clin Linguist Phon.* 2003;**17**(8):617-43. [PubMed: [14977026](https://pubmed.ncbi.nlm.nih.gov/14977026/)].
7. Bernthal JE, Bankson N, Flipsen P. Articulation and Phonological Disorders. Allyn & Bacon, Incorporated; 2008.
8. Eisenberg SL, Hitchcock ER. Using standardized tests to inventory consonant and vowel production: a comparison of 11 tests of articulation and phonology. *Lang Speech Hear Serv Sch.* 2010;**41**(4):488-503. doi: [10.1044/0161-1461\(2009\)08-0125](https://doi.org/10.1044/0161-1461(2009)08-0125). [PubMed: [20581217](https://pubmed.ncbi.nlm.nih.gov/20581217/)].

9. Roulstone S, Peters TJ, Glogowska M, Enderby P. Predictors and outcomes of speech and language therapists' treatment decisions. *Int J Speech Lang Pathol.* 2008;**10**(3):146-55. doi: [10.1080/17549500801894362](https://doi.org/10.1080/17549500801894362). [PubMed: [20840048](https://pubmed.ncbi.nlm.nih.gov/20840048/)].
10. Friberg JC. Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teach Ther.* 2010;**26**(1):77-92. doi: [10.1177/0265659009349972](https://doi.org/10.1177/0265659009349972).
11. McLeod S, Verdon S. A review of 30 speech assessments in 19 languages other than English. *Am J Speech Lang Pathol.* 2014;**23**(4):708-23. doi: [10.1044/2014_AJSLP-13-0066](https://doi.org/10.1044/2014_AJSLP-13-0066). [PubMed: [24700105](https://pubmed.ncbi.nlm.nih.gov/24700105/)].
12. McCauley RJ, Swisher L. Psychometric review of language and articulation tests for preschool children. *J Speech Hear Disord.* 1984;**49**(1):34-42. [PubMed: [6700200](https://pubmed.ncbi.nlm.nih.gov/6700200/)].
13. McCauley RJ, Swisher L. Use and misuse of norm-referenced tests in clinical assessment: a hypothetical case. *J Speech Hear Disord.* 1984;**49**(4):338-48. [PubMed: [6389982](https://pubmed.ncbi.nlm.nih.gov/6389982/)].
14. Nilipour R. Emerging issues in speech therapy in Iran. *Folia Phoniatr Logop.* 2002;**54**(2):65-8. [PubMed: [12037418](https://pubmed.ncbi.nlm.nih.gov/12037418/)].
15. McLeod S. Resourcing speech-language pathologists to work with multilingual children. *Int J Speech Lang Pathol.* 2014;**16**(3):208-18. doi: [10.3109/17549507.2013.876666](https://doi.org/10.3109/17549507.2013.876666). [PubMed: [24833427](https://pubmed.ncbi.nlm.nih.gov/24833427/)].
16. Ghasisin L, Ahmadipour T, Mostajeran F, Moazam M, Derakhshandeh F. Evaluating the reliability and validity of phonetic information test in normal 5-6 year-old children of Isfahan city. *J Res Rehabil Sci.* 2013;**9**(2):153-60.
17. Verdon S, McLeod S, Winsler A. Linguistic diversity among Australian children in the first 5 years of life. *Speech Language Hear.* 2014;**17**(4):196-203.
18. Holm A, Crosbie S, Dodd B. Differentiating normal variability from inconsistency in children's speech: normative data. *Int J Lang Commun Disord.* 2007;**42**(4):467-86. doi: [10.1080/13682820600988967](https://doi.org/10.1080/13682820600988967). [PubMed: [17613100](https://pubmed.ncbi.nlm.nih.gov/17613100/)].
19. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol.* 1995;**50**(9):741-9. doi: [10.1037/0003-066x.50.9.741](https://doi.org/10.1037/0003-066x.50.9.741).
20. Fairchild AJ, Finney SJ. Investigating Validity Evidence for the Experiences in Close Relationships-Revised Questionnaire. *Educ Psychol Measure.* 2016;**66**(1):116-35. doi: [10.1177/0013164405278564](https://doi.org/10.1177/0013164405278564).
21. Dodd B, Zhu H, Crosbie S, Holm A, Ozanne A. Diagnostic evaluation of articulation and phonology (DEAP). Psychology Corporation; 2002.
22. Ingram D. Phonological disability in children. 2. Elsevier Publishing Company; 1977.
23. Nematzadeh S, Dadras M, Dastjerdi Kazemi M, Mansoorzadeh M. Core Vocabulary. Tehran: School; 2012.
24. Mohammadi MH. Picture-Riddle Dictionary. Tehran: Cheesta; 2004.
25. Shokri G. A Persian dictionary for school children. Tehran: Institute for Humanities and Cultural Studies; 2004.
26. YKazemi Y, Nematzadeh SH, Hajian T, Heidari M, Daneshpajouh T, Mirmoeini M. The validity and reliability coefficient of Persian translated McArthur-Bates Communicative Development Inventory. *J Res Rehabil Sci.* 2008;**4**(1).
27. Strand EA, McCauley RJ, Weigand SD, Stoeckel RE, Baas BS. A motor speech assessment for children with severe speech disorders: reliability and validity evidence. *J Speech Lang Hear Res.* 2013;**56**(2):505-20. doi: [10.1044/1092-4388\(2012\)12-0094](https://doi.org/10.1044/1092-4388(2012)12-0094). [PubMed: [23275421](https://pubmed.ncbi.nlm.nih.gov/23275421/)].
28. Rosner B. Fundamentals of Biostatistics. Boston: Cengage Learning; 2011.
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;**33**(1):159-74. [PubMed: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)].
30. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Measure.* 2016;**20**(1):37-46. doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
31. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull.* 1968;**70**(4):213-20. doi: [10.1037/h0026256](https://doi.org/10.1037/h0026256).
32. Bleile KM. Manual of articulation and phonological disorders: Infancy through adulthood. Cengage Learning; 2004.
33. Powell TW. Stimulability considerations in the phonological treatment of a child with a persistent disorder of speech-sound production. *J Commun Disord.* 1996;**29**(4):315-33. [PubMed: [8863121](https://pubmed.ncbi.nlm.nih.gov/8863121/)].
34. Miccio AW, Elbert M, Forrest K. The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies. *Am J Speech Language Pathol.* 1999;**8**(4):347-63.
35. Davis BL. In: Phonological disorders in children: Clinical decision making in assessment and intervention. Kamhi AG, Pollock KE, editors. Baltimore: Brookes; 2005. pp. 89-100. Goal and target selection for developmental speech disorders.
36. Gierut J. In: Phonological disorders in children: Clinical decision making in assessment and intervention. Kamhi AG, Pollock KE, editors. Baltimore: Brookes; 2005. pp. 101-8. Phonological intervention: The how or the what?
37. Messick S. Test validity: A matter of consequence. *Soc Indic Res.* 1998;**45**(1/3):35-44. doi: [10.1023/a:1006964925094](https://doi.org/10.1023/a:1006964925094).
38. Kent RD, Kent JF, Rosenbek JC. Maximum performance tests of speech production. *J Speech Hear Disord.* 1987;**52**(4):367-87. [PubMed: [3312817](https://pubmed.ncbi.nlm.nih.gov/3312817/)].
39. Laffin JJ, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genet Med.* 2012;**14**(11):928-36. doi: [10.1038/gim.2012.72](https://doi.org/10.1038/gim.2012.72). [PubMed: [22766611](https://pubmed.ncbi.nlm.nih.gov/22766611/)].