

Validation and inter-rater reliability testing of the Arabic version of speech intelligibility rating among children with cochlear implant

Yazeed A. Al-shawi, MD, Tamer A. Mesallam, MD, Nouf M. Albakheet, MD, Moath A. Alshawi, CL/PM, Rayan M. Alfallaj, MD, Turki M. Aldrees, MD, Ashwag A. Algahtani, SLP, Tabani O. Alotaibi, SLP

ABSTRACT

Objectives: To validate and assess the reliability of the new version of an Arabic speech intelligibility rating among different raters.

Methods: This cross-sectional analysis was carried out between December 2018 and January 2019. Thirty cochlear-implant (CI) children (study group) and 30 subjects (control group) were enrolled. Study candidates' speech skills were evaluated using the translated Arabic SIR by parents and original SIR by professions such as speech-language pathologists (SLPs). Inter-rater agreement, test-retest reliability, pre- and post-intervention score (responsiveness test), patient versus control score comparison (discriminant validity), and cross-validation of Arabic SIR have all been assessed.

Results: There was a good sense of agreement between the post-operative SIR parents' assessments and the professional SLPs' assessments ($r=0.920$, $p<0.001$). The mean of study subjects pre- and post-implantation score of Arabic SIR showed a statistically significant difference ($p<0.001$).

Conclusion: The Arabic SIR demonstrated excellent reliability with strong consistency. It showed its clinical ability in distinguishing healthy subjects from patients along with follow up of speech development skills over time. The Arabic SIR can be used by parents to evaluate post-CI progress of their children.

Keywords: cochlear implant, validation, inter-rater reliability, speech intelligibility rating

*Saudi Med J 2020; Vol. 41 (10): 1139-1143
doi: 10.15537/smj.2020.10.25342*

Severe-to-profound hearing loss in early childhood interferes with normal development of speech and language and had negative impact on child's education.¹ An ultimate outcome of cochlear implant

(CI) is intelligible speech acquisition and language development.² A rating scale provides a feasible method to monitor patients over a long period of time. However, a rating scale can only be used by different raters when they ascertain its inter-rater reliability. Good inter-rater reliability implies that the differences can be attributed to the behaviors of the patients being evaluated rather than the variability in scoring by the different observers.³ Inter-observer reliability is a statistical method to measure the extent of confounding errors that affect the perceived scoring, measured using intra-class correlation coefficient (ICC).⁴ High ICC values indicates high agreement among the raters.³ Speech intelligibility rating (SIR) is an adapted scale that allows both experienced listeners, including audiologists or speech-language pathologists (SLPs), and normal-hearing listeners with no prior expertise to assess speech perception before and after CI.⁵ It is comprised of 5 categories as shown in **Table 1**. Speech intelligibility is higher as the classification number increases.³ There is an increased need for a reliable validated Arabic version of SIR which will provide substantial clinical implications toward the practice of healthcare providers among Arabic-speaking populations. Moreover, this enhances the delivery of accurate assessments and monitoring the progress of CI patients. The availability of a validated Arabic SIR (that shows good reliability between professional and non-professional raters) also gives families a daily assessment tool to monitor their children's performance and helps to set home training targets for speech rehabilitation. A thorough search in the literature demonstrated no existing validated Arabic SIR.

Table 1 - Categories of speech intelligibility rating (SIR) scales.

SIR category	Speech intelligibility category
Category 5	Connected speech is intelligible to all listeners. The child is understood easily in everyday contexts
Category 4	Connected speech is intelligible to a listener who has little experience of a deaf person's speech. The listener does not need to concentrate unduly
Category 3	Connected speech is intelligible to a listener who concentrates and lip-reads within a known context
Category 2	Connected speech is unintelligible. Intelligible speech is developing in single words when context and lip-reading cues are available
Category 1	Pre-recognizable words in spoken language. The child's primary mode of everyday communication may be manual

The purpose of our study is therefore to validate and estimate the inter-observer reliability (between SPLs and parents) of Arabic SIR.

Methods. Development of an Arabic SIR. The newly designed Arabic SIR were developed. The recommendations for standard cross cultural adaptation were followed in this study.⁶ The original SIR was translated by 2 professional linguistic advisors (mother tongues are Arabic), and the Arabic SIR was evaluated by 3 SLPs who speak both languages and 10 special-needs teachers. As per their advice, minimal changes have been implemented for cultural adaptation purposes. The Arabic version was then translated back to English by another 2 linguistic advisors. Back translated version was compared to the original version of SIR and yielded high similarity, ensuring that the original meaning was preserved.

The ethical committee of College of Medicine, King Saud University, Riyadh, Saudi Arabia approved the research protocol, and informed consent was taken from subject's parents. This cross-sectional study was conducted at King Abdullah Ear Specialist Center, Riyadh, Saudi Arabia between December 2018 and January 2019. The study included 30 patients and 30 control subjects. Group of patients were enrolled from post-CI pediatric speech rehabilitation clinic, while the group of controls (with normal hearing) were enrolled from clinics of pediatrics in the same center. The study enrolled every third patient arriving at both clinics. The inclusion criteria were pediatric prelingually deaf patients who received CI surgery. The exclusion criteria were adults (aged 18 and above), any patient who has post-lingual deafness, and those with psychological or neurological diseases.

The process of validation. Every subject underwent an interview and was assessed by a SLP and one of their parents, using Arabic SIR, for inter-rater reliability. The meaningful use of speech scale (MUSS) is a test that assess the speech development in children. It has 3 domains: voice control, communication skills, and spontaneous use of language and this test was carried out for all subjects by SLPs for cross-validation with the parents' scores of their children for Arabic SIR.⁷ To

assess internal consistency of Arabic SIR, the evaluations were repeated after 14 days for test-retest reliability using intraclass correlation coefficient (ICC). Post-implantation Arabic SIR scores were compared with pre-implantation SIR scores. The patients' Arabic SIR scores were also compared with the Arabic SIR scores in the control group for discriminant validity.

Statistical analysis. The statistical analysis was performed by the Statistical Package for Social Sciences version 23.0 (Inc, IBM, Armonk, New York, USA) was used to analyze the data. The expression of date was numbers and percentages for categorical data and mean, standard deviation, and range for continuous date. For the aim of comparing 2 categorical data Chi-square tests was used. And to compare 2 different means we utilized the Fisher's exact test and t-test. The statistics of Kappa were implemented to discover any discrepancy among different tests. For determining the level of agreement between the tests in this study test-retest statistics, including the Spearman's test, inter-item and inter-class correlation, and Cronbach's alpha, were calculated. A *p* value of <0.05 was considered statistically significant.

Results. Patient group included 30 CI children with a mean age of 66.68 months (SD=37.1), while control group included 30 children with normal hearing with 80.28 months as a mean age (SD=32.57) (Table 2). The mean age of 2 groups demonstrated no statistically significant difference (*p*=0.153). Approximately half of the mothers (no:15) of children with CI held bachelor's degrees (50.0%) followed by 11 mothers with high school (36.7%) and 3 with diploma degrees (10.0%).

Table 2 - Children demographic information (N=60).

Characteristic	Cochlear implant children	Control group
Age in months, mean (SD)	66.86±37.18	80.28±32.57
Age in months, range	19 to 144	24 to 144
Gender		
Male	18 (60.0)	16(53.3)
Female	12 (40.0)	14(46.7)
The mean age of cochlear implant in months, mean (SD)	36.32±29.52	
Cochlear implant		
Unilateral	10(33.3)	
Bilateral	20 (66.7)	

Disclosure. This study was approved by the Deanship of Scientific Research, King Saud University, Riyadh, Kingdom of Saudi Arabia through Vice Deanship of Scientific Research Chairs.

20 (66.7%) of the patients were implanted bilaterally simultaneously.

Table 3 shows the agreement rates of the Arabic SIR scores between the mothers and the SLPs. There was a positive agreement between the post-operative Arabic SIR mothers' and the professional SLPs' assessments ($r=0.920$, $p<0.001$), with 20 (66.7%) in total agreement. There was variation in the agreement percentage across the SIR categories, and the lowest was in category 5 (13.3%). However, Arabic SIR showed a high Spearman's correlation coefficient for test and retest evaluation ($r=0.93$, $p<0.001$), with an excellent Cronbach's alpha level and an intra-class coefficient value of 0.958.

Table 3 - Speech intelligibility rating scales categories discrepancies between speech intelligibility rating scales and mother's assessment.

Category	No. of discrepancies
0	None
1	1 (3.3)
2	2 (6.7)
3	2 (6.7)
4	1 (3.3)
5	4 (13.3)

Values are presented as numbers and percentage (%). Kappa=0.57, $p<0.001$

In the responsiveness validity test, there were significantly higher Arabic SIR scores reported for post-CI children, at $2.97 (\pm 1.497)$, compared to their pre-operative Arabic SIR ratings of $1.07 (\pm 0.254)$, with a statistically significant level of $p<0.001$. However, discriminant validity showed significantly lower Arabic SIR scores in the patients' group, at $1.07 (\pm 0.254)$, compared to the control group, at $5.00 (\pm 0.01)$, with a statistically significant level of $p<0.001$. A positive correlation was found between the post-operative Arabic SIR and post-operative MUSS, which indicates the positive concurrent validity of the Arabic SIR (**Figure 1**).

Discussion. The purpose of this study is to create an Arabic version of SIR. Developing such a tool will help professionals as well as non-professionals, including parents and schoolteachers, to gain more insight on speech development of children with CI, which will fill the gap between those who are working and dealing with this group of children in order to achieve better outcomes.

During the process of validating Arabic SIR, it showed a strong test-retest reliability as evidenced in the results section that confirms its internal consistency and reliability. On the other hand, the inter-rater reliability of Arabic SIR has been assessed by comparing the ratings of SLPs and of the mothers of CI children following implantation. This data has been evaluated through both kappa statistics and correlation analysis. A high correlation coefficient value of 0.9 has been reported and there was also a high agreement between the mothers and SLPs regarding the Arabic SIR ratings,

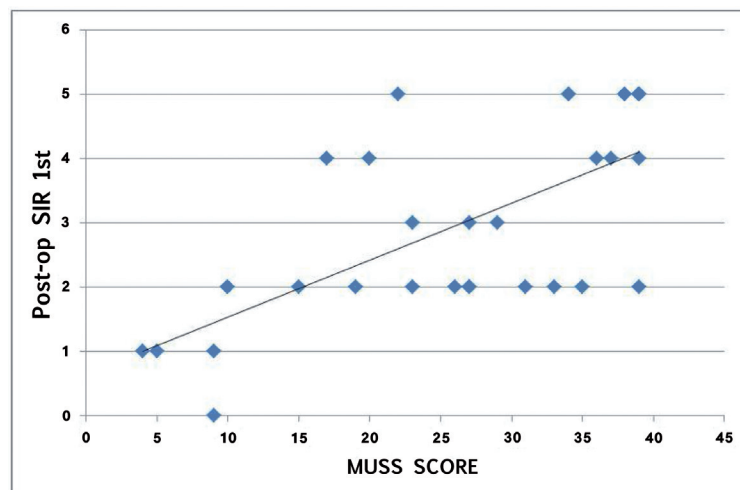


Figure 1 - Concurrent validity between speech intelligibility rating (SIR) and meaningful use of speech scale (MUSS). Pearson correlation=0.588, $p<0.001$

as indicated by the kappa value of nearly 0.6. However, there were a few discrepancies reported for category 5, with a tendency toward higher scores being reported by the mothers compared to those reported by SLPs. This discrepancy might be explained by the greater familiarity of mothers with their children's speech, which can make them overestimate this item of the scale and consider their child's speech to be more intelligible. Variation in SIR scores between patients could be attributed to multiple factors such as duration since surgery, differences in rehabilitation, and personal cognitive differences between candidates. Similar findings of inter-rater reliability were reported by Allen et al⁷ in their study of developing SIR with a high correlation coefficient and agreement between raters for most of the SIR categories. High inter-rater agreement has also been reported in other related studies that investigated the reliability of the speech intelligibility rating in hearing-impaired and CI children.^{3,8,9} These findings confirm the validity of Arabic SIR scale being used in the assessment of speech of CI children not only by professionals but also by others who may lack the experience in listening to children with impaired hearing. Moreover, current study results proved discriminative validity of the Arabic SIR demonstrated by significant difference in SIR ratings between pre and post implant and between CI patients and control subjects. ($p < 0.001$). Meaningful use of speech scale is a widely used test for evaluating speech development of children with CI.¹⁰ Significant positive correlation between MUSS and Arabic SIR was demonstrated which indicates the concurrent validity.

A validated Arabic SIR helps to improve accuracy of assessment of CI patients among Arabic speaking healthcare providers. Moreover, this test can be used by parents, school and special education teachers, and non-professional who are in direct contact of CI patients to monitor their progress. Such a tool will encourage parents to participate more in rehabilitation of their implanted children. The study findings demonstrated that Arabic SIR is valid and can be used by professional and nonprofessional native Arabic speakers, who are in direct contact with CI-fitted children, to ensure improved quality in post CI rehabilitation monitoring. A validated Arabic category of auditory performance CAPII has been previously developed by the group.¹¹

The limitation of this study is the relatively small sample size. There is no other validated Arabic SIR test in the literature. Considering the increased numbers of CI surgeries performed worldwide on children with hearing loss, the need for the development of validated versions of SIR in different languages would be an essential requirement.

In conclusion, the Arabic SIR appears to be a reliable and valid tool for the assessment of spoken language development in Arabic-speaking children with hearing impairment post CI. Developing an Arabic SIR allows for the opportunity to assess a child's speech abilities not only in the clinical setup but in everyday situations.

Received 20th June 2020. Accepted 16th August 2020.

From the Otolaryngology Department (Al-shawi Y), Prince Sultan Military Medical City; from the ENT Department (Mesallam), College of Medicine, King Saud University; from the Department of Otolaryngology Head & Neck Surgery (Alfallaj), King Saud bin Abdulaziz University Hospital; from the Department of Otolaryngology Head & Neck Surgery (Albakheet), King Saud bin Abdulaziz University for Health Science; from Trusted Partners (Alshawi M); from the Speech and Aural Rehabilitation Unit (Algahtani), King Abdullah Ear Specialist Center; and from the Communication & Swallowing Disorders Unit (Alotaibi), ENT Department, King Abdulaziz University Hospital; and from the Department of Otolaryngology-Head and Neck Surgery (Aldrees), College of Medicine, Prince Sattam Bin Abdulaziz University, Riyadh Kingdom of Saudi Arabia.

Address correspondence and reprints request to: Dr. Yazeed A. Al-shawi, Otolaryngology Department, Prince Sultan Military Medical City, Riyadh, Kingdom of Saudi Arabia. E-mail: Dr.alsbawi@windouslive.com
ORCID ID: <https://orcid.org/0000-0002-8497-8326>

Acknowledgment. The authors are grateful to the Deanship of Scientific Research, King Saud University, Riyadh, Kingdom of Saudi Arabia for funding through Vice Deanship of Scientific Research Chairs. This research was recommended by the Saudi Otorhinolaryngology Society. Also, the authors gratefully acknowledge Scribendi (www.scribendi.com) for English language editing.

References

1. Bhamjee A, Roux TL, Schlemmer K, Perold J, Cass N, Schroeder K, et al. Parent-perceived challenges related to the pediatric cochlear implantation process and support services received in South Africa. *Int J Pediatr Otorhinolaryngol* 2019; 126: 109635.
2. Bruijnzeel H, Ziylan F, Stegeman I, Topsakal V, Grolman W. A systematic review to define the speech and language benefit of early (<12 months) pediatric cochlear implantation. *Audiol Neurotol* 2016; 21: 113-126.
3. Wilkinson AS, Brinton JC. Speech intelligibility rating of cochlear implanted children: inter-rater reliability. *Cochlear Implants Int* 2003; 4: 22-30.
4. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-428.
5. Calmels M-N, Saliba I, Wanna G, Cochard N, Fillaux J, Deguine O, et al. Speech perception and speech intelligibility in children after cochlear implantation. *Int J Pediatr Otorhinolaryngol* 2004; 68: 347-351.
6. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract* 2010; 17: 268-724.
7. Allen C, Nikolopoulos TP, Dyar D, O'donoghue GM. Reliability of a rating scale for measuring speech intelligibility after pediatric cochlear implantation. *Otol Neurotol* 2001; 22: 631-633.

8. Doyle J. Reliability of audiologists' ratings of the intelligibility of hearing-impaired children's speech. *Ear Hear* 1987; 8: 170-174.
9. Samar VJ, Metz DE. Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *J Speech Lang Hear Res* 1988; 31: 307-316.
10. Robbins AM, Osberger MJ (1991) Meaningful use of speech scale (muss). American Speech-Language and Hearing Association, Atlanta. [Updated 1991]. Available from URL: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Robbins+AM%2C+Osberger+MJ+%281991%29+Meaningful+use+of+speech+scale+%28muss%29.+American+Speech-Lan+guage+and+Hearing+Association%2C+Atlanta&btnG=
11. Al-Shawi Y, Mesallam TA, Alfallaj R, Aldrees T, Albakheet N, Alshawi M, et al. Inter-rater reliability and validity of the Arabic version of Categories of Auditory Performance-II (CAP-II) among children with cochlear implant. *Otol Neurotol* 2020; 41: e597-e602.
12. Liu R, Jiao Q, Ji F, Wang Q, Zhao H, Li J, et al. The use of the MUSS and the SIR scale in late-implanted prelingually deafened adolescents and adults as a subjective evaluation. *Acta Otolaryngol* 2019; 140: 94-98.

Appendix 1

التصنيف	معياري ووضوح الكلام
١	الحديث المتواصل ليس مفهوماً، الطفل في المرحلة ما قبل الكلامية للغة المنطوقة، طريقة التواصل الأساسية للطفل في الحياة اليومية من الممكن ان تكون بالإشارة
٢	الحديث المتواصل غير مفهوم، بينما حديثه المفهوم يكون على شكل كلمات مفردة في حال اتضح له سياق الحديث و أعطى تلميحات عن طريق قراءة الشفاه
٣	حديث الطفل المتواصل مفهوم عندما يستمع له المستمع بتركيز ويقرأ الشفاه
٤	حديث الطفل المتواصل مفهوم عندما يكون عند المستمع خبرة قليلة في فهم حديث الصم
٥	حديث الطفل المتواصل مفهوم لجميع المستمعين حيث يفهم الطفل بسهولة في سياق الحديث اليومي.