



Research Paper:

Investigating the Inter-rater Agreement of Farsispeaking Raters on the Grade, Roughness, Breathiness, Asthenia, and Strain Auditory Perceptual Scale

Amirmohammd Danesh¹^(D), Farhad Torabinezhad¹^{*}^(D), Arezoo Saffarian²^(D), Jamileh Abolghasemi³^(D), Mohammad Kamali⁴^(D), Mohammad-sadigh Mahmudzadeh¹^(D)

- 1. Department of Speech and Language Pathology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran.
- 2. Rehabilitation Research Center, Department of Speech and Language Pathology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran.

3. Department of Biostatistics, School of Health, Iran University of Medical Sciences, Tehran, Iran.

4. Department of Rehabilitation Management, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran.



*This work has been published under CC BY-NC-SA 4.0 license.

Article info: Received: 05 Oct 2022 Accepted: 27 Nov 2022 Available Online: 26 Dec 2022

Funding

This article is extracted from the master's thesis of the first author in the Department of Speech Therapy, School of Rehabilitation Sciences, Iran University of Medical Sciences.

Conflict of interest

The authors declare no conflict of interest.

Use your device to scan and read the article online

Cite this article as Danesh A, Torabinezhad F, Saffarian A, Abolghasemi J, Kamali M, Mahmudzadeh M. Investigating the Inter-rater Agreement of Farsi-speaking Raters on the Grade, Roughness, Breathiness, Asthenia, and Strain Auditory Perceptual Scale. Function and Disability Journal. 2022; 5:E63. http://dx.doi.org/10.32598/fdj.5.63

doj http://dx.doi.org/10.32598/fdj.5.63

* Corresponding Author:

Farhad Torabinezhad, Assistant Professor.

Address: Department of Speech and Language Pathology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran. Tel: +98 (912) 7632809

E-mail: torabinezhad.f@iums.ac.ir

ABSTRACT

Background: Auditory perceptual assessment is one of the important evaluations for voice assessment. Among the available auditory-perceptual assessments, the grade, roughness, breathiness, asthenia, strain (GRBAS) auditory perceptual scale has attracted the attention of many researchers and therapists in Iran and other parts of the world. The GRBAS scale is a perceptual voice assessment scale that subjectively assesses voice quality; however, the audio samples in the original GRBAS scale that are used as training tools for raters are Japanese. There are various segmental and suprasegmental differences between the Japanese and Persian languages since these factors can affect the accuracy of the perceptual evaluation.

Objectives: This research aims to investigate the inter-rater agreement among the Persian raters who only had access to the Japanese samples in the main profile.

Methods: In this study, 8 speech and language pathologists were selected as raters with more than 5 years of clinical experience in evaluating and treating voice disorders. Several 137 audio samples, containing the prolongation of vowel /a/ and reading the standard text "Grandfather Passage" were provided to the participating raters. The raters were asked to score the audio samples based on the GRBAS auditory perceptual scale. The results were statistically analyzed via the Cohen kappa coefficient.

Results: The highest agreement in the R parameter was related to rater number 5 (kappa=0.585) while the lowest value was related to parameter S, related to rater number 1 (kappa=-0.018).

Conclusion: According to the results, the clinical experience of raters using the Japanese samples cannot lead to an increase in the agreement and ability of Persian-speaking raters in evaluating Persian samples.

Keywords: Voice, Voice disorder, Auditory-perceptional assessment



What is "already known" in this topic:

GRBAS scale is Auditor-perceptual voice assessment tool. Using this scale is simple and fast. However, some parameters such as language of raters can affect the results obtained. These effects can be different according to the characteristics of each language.

What this article adds:

The characteristics of Persian language can affect the judgment of raters using the GRBAS auditory perception scale.

Introduction

oice is an auditory perception term that describes an audible sound that is made by the larynx and represents parameters, such as pitch, loudness, and quality [1]. An abnormal voice attracts the attention of the listeners, does not meet the profes-

sional and social needs of the person, and is not appropriate for the age, gender, and social status of the speaker [1]. Although we believe that voice disorders exist when the quality, pitch, and loudness of a person's voice are different from other people who are similar to them in terms of age, gender, and cultural and social class, there is still no fixed criterion for identifying a normal voice. We considered a wide range of sounds as normal [1]. A comprehensive assessment of voice is the first step to effectively treating voice disorders [2]. To evaluate different aspects of voice production, 4 approaches are used among therapists as follows [3]:

1) Aauditory-perceptual assessment, 2) acoustic assessment, 3) aerodynamic assessment, and 4) imaging.

Among experienced voice therapists, perceptual assessments are one of the most widely used clinical assessments [4, 5]. Some voice therapists and researchers have progressed a step further and considered perceptual evaluations as the gold standard for classifying voice disorders [6, 7]. Among the existing perceptual assessment scales, several scales exist that, in addition to many studies based on them, have many clinical applications (compared to other available evaluation tools) [8-10]. In the meantime, the grade, roughness, breathiness, asthenia, strain (GRBAS) auditory perceptual scale has attracted the attention of many researchers and therapists. In addition to being a valid scale for assessing voice quality [11, 12], this scale is also used by speech therapy students as an educational tool [12].

The GRBAS scale is a perceptual assessment tool that assesses the overall dysphonia grade, roughness, breathiness, asthenia, and strain. The GRBAS scale was designed by Hirano in 1981 in Japanese and the form of a profile [13]. GRBAS is a 5-parameter scale that attributes qualitative values to the quality of a person's voice. Despite the simplicity and quick implementation method [5, 14], and the good correlation that this scale has with acoustic parameters [15], many factors, such as listener experience [12, 16, 17], cultural and social factors [9, 18, 19], the severity of voice damage [19], and the type of speech task [20] can affect a person's auditory judgment. [21]. A group of experts believes that each person's perception of voice goes back to their past linguistic experiences [22]. Also, another group of researchers believes that the effect of familiarity with a language can affect the better identification of the characteristics of the voice [23, 24]. Yiu et al. investigated the role of cultural and linguistic differences in the perception of sound quality. In this research, 40 speech and language pathologists from Australia and Hong Kong were asked to evaluate the quality of breathiness and harshness of voice. Accordingly, there is evidence of the influence of language and culture on the perception of some features of a sound [25]. Vaz Freitas et al. investigated inter-reliability and intra-reliability based on the GRBAS scale. The results showed a 95% intra-rater correlation and a higher correlation between the 3 parameters of the overall intensity of the disorder (G), roughness (R), and breathiness of the voice (B), and a lower correlation between the 2 parameters of weakness (A) and effort and struggle in the voice (S). Also, weak inter-rater reliability was observed between 40% of raters [26]. Chaves et al. investigated the effect of the mother tongue on the perceptual evaluation of Canadian and Brazilian speech and language pathologists. In this research, 46 samples of continuous speech (extracted from 35 women and 11 men) and 46 samples of vowel stretching (extracted from 37 women and 9 men) based on the GRBAS scale were evaluatedby two



groups of speech and language pathologists from both countries. In this research, moderate to weak values were reported for inter-rater reliability [27].

All the aforementioned studies in this section emphasize the importance of considering linguistic features as an influential factor in perceptual evaluations. Language differences can cause differences in the perception of different features of a voice. In some languages, the difference in linguistic features is small; therefore, it does not have a significant impact on perceptual evaluation. On the contrary, in some languages, these differences are significant and can cause great differences in evaluation. This difference can be such that a voice is considered normal in one country and abnormal in another country. Hence, apart from considering the research results, attention should be paid to the difference between the two languages.

Therefore, the purpose of this study is to investigate the agreement rate of Persian language raters on the GRBAS auditory-perceptual scale who only had access to the Japanese samples in the main profile.

Materials and Methods

To check the inter-rater agreement, 137 voice samples were collected from patients referred to the speech therapy clinic of the Iran University of Medical Sciences. These samples were prepared by the BBC Pro Sound

Table 1. The inter-rate	r agreement based on	the kappa coefficient
-------------------------	----------------------	-----------------------

recorder and BBS-MU-435 unidirectional-Hong Kong microphone. These samples contained reading a standard passage (grandfather passage) and prolongation of the vowel /a/, which is necessary to perform a perceptual assessment based on the GRBAS auditory-perceptual scale. For ease of access, these samples were uploaded on the internet along with the scoring from 8 raters, 7 of whom were speech-language pathologists and one of the raters was a laryngology specialist. All raters who had more than 5 years of experience in the auditory-perceptual voice assessment were selected to perform the assessment. Audio samples were played once for them on the internet and then were asked to score each of the GRBAS parameters based on it. After scoring, the sample was out of reach of the evaluator and it was not possible to change it. The data were analyzed via the SPSS software, version 25.

Results

To check the agreement rate between the raters, we used the Cohen kappa statistics. The results are summarized in Table 1.

The highest agreement in the G parameter is related to rater number 4 (kappa=0.546) and the lowest value is related to rater number 2 (kappa=0.073), the highest agreement in the R parameter is related to rater number 5 (kappa=0.585), and the lowest is related to rater number 3 (kappa=0.073), the highest agreement in parameter B

Kappa Coefficient	G		R		В		А		S	
Raters*	Minimum	Maximum								
Rater 1	0.163	0.401	0.132	0.271	0.138	0.418	0.049	0.144	0.018-	0.189
Rater 2	0.073	0.339	0.096	0.380	0.079	0.240	0.254	0.418	0.284	0.379
Rater 3	0.230	0.343	0.073	0.207	0.164	0.333	0.158	0.317	0.209	0.319
Rater 4	0.302	0.546	0.288	0.377	0.223	0.590	0.153	0.509	0.212	0.473
Rater 5	0.424	0.505	0.379	0.585	0.278	0.495	0.245	0.351	0.296	0.532
Rater 6	0.335	0.482	0.333	0.341	0.299	0.404	0.324	0.382	0.333	0.442
Rater 7 and 8**	0.3	66	0.3	20	0.2	223	0.2	24	0.2	.96

Notes: * To calculate the inter-rater agreement, firstly, the agreement of rater number 1 with rater number 2 and later is calculated, and in examining the inter-rater agreement number 2, the agreement of this rater with rater number 3 and later is calculated.

^{**} To calculate the inter-rater agreement, firstly, the agreement of rater number 1 with rater number 2 is calculated, and in the review of the inter-rater agreement number 2, the agreement of this rater with rater number 3 is calculated, in the case of rater number 7, only rater number 8 remains, therefore, the minimum and maximum value is not relevant.



is related to rater number 4 (kappa=0.509) and its lowest value corresponds to rater number 2 (kappa=0.079). The highest agreement in parameter A is related to rater number 4 (kappa=0.509) and its lowest value is related to rater number 1 (kappa=0.049). The highest agreement in parameter S is related to rater number 5 (kappa=0.532) and the lowest value is attributed to rater number 1 (kappa=-0.018).

Discussion

This study aimed to investigate the agreement between raters on the GRBAS scale among Persian language raters who only had access to Japanese samples for training. The results of this research showed a weak to moderate inter-rater agreement. These results indicated that considering that the Persian language raters only have access to the Japanese samples included in the main sample of the GRBAS auditory perception scale for training, there was a weak agreement between them in determining the scores of the GRBAS auditory perception scale in the Persian samples. Similar to the results obtained in this research, in the study of Vaz Freitas et al. poor inter-rater reliability was reported among half of the participants in the research [26]. Also, in the research of Chaves et al., which investigated the effect of language in perceptual evaluation, moderate to weak values were reported for inter-rater reliability. This is following the results obtained in this research [27].

On the other hand, the moderate to a weak inter-rater agreement is proof of the impact of language on perceptual evaluation. This finding is similar to what Yiu et al. stated about the role of language and culture differences in the perception of sound quality [25].

Conclusion

A moderate to weak agreement was observed between Persian-speaking raters for scoring Persian samples based on the GRBAS listening perception scale. Therefore, considering the important role of language, differences in loudness, leaning, and the degree of breathiness of the voice that may be considered normal or abnormal in different cultures, in perceptual raters, the access of Persian language raters to reliable Persian samples can play an important role in increasing the inter-raters' agreement.

Ethical Considerations

Compliance with ethical guidelines

The Ethics Committee of Iran University of Medical Sciences approved this study (Ethics Code: IR.IUMS. REC.1400.913). All participants in this study signed an informed consent form.

Funding

This article is extracted from the master's thesis of Amirmohammad Danesh entitled "Persian version of GRBAS scale" presented at the Department of Speech Therapy, School of Rehabilitation Sciences, Iran University of Medical Sciences.

Authors' contributions

Conceptualization and supervision: Amirmohammad Danesh, Farhad Torabinezhad and Arezoo Saffarian; Methodology: Jamileh Abolghasemi and Mohammad Kamali; Investigation, writing-review & editing: All authors; Writing-original draft: Amirmohammad Danesh and Mohammadsadigh Mahmudzadeh; Funding acquisition and Resources: Farhad Torabinezhad.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

I would like to express my gratitude to all the professors who have contributed to the results

References

- Aronson AE, Bless D. Clinical voice disorders. New York: Thieme; 2011. [Link]
- [2] Sataloff RT, Lalwani AK, Fried MP, Tabaee A, Benninger MS, Hartnick CJ. Sataloff's comprehensive textbook of otolaryngology head and neck surgery: Laryngology. New Delhi: Jaypee Brothers Medical Publishers; 2016. [Link]
- [3] Mehta DD, Hillman RE. Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. Curr Opin Otolaryngol Head Neck Surg. 2008; 16(3):211-5. [DOI:10.1097/ MOO.0b013e3282fe96ce] [PMID] [PMCID]
- [4] Behrman A. Common practices of voice therapists in the evaluation of patients. J Voice. 2005; 19(3):454-69. [DOI:10.1016/j. jvoice.2004.08.004] [PMID]



- [5] Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA. The reliability of three auditory perceptual scales for dysphonia. Eur Arch Otorhinolaryngol. 2004; 261(8):429-34. [DOI:10.1007/s00405-003-0707-7] [PMID]
- [6] Kent RD. Hearing and believing: Some limits to the auditoryperceptual assessment of speech and voice disorders. Am J Speech Lang Pathol. 1996; 5(3):7-23. [DOI:10.1044/1058-0360.0503.07]
- [7] Ma EP, Yiu EM. Multiparametric evaluation of dysphonic severity. J Voice. 2006; 20(3):380-90. [DOI:10.1016/j.jvoice.2005.04.007]
 [PMID]
- [8] Carding P, Carlson E, Epstein R, Mathieson L, Shewell C. Formal perceptual evaluation of voice quality in the United Kingdom. Logopedics Phoniatrics Vocology. 2000; 25(3):133-8. [DOI:10.1080/14015430050175860] [PMID]
- [9] Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, et al. Committee on Phoniatrics of the European Laryngological Society (ELS). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). Eur Arch Otorhinolaryngol. 2001; 258(2):77-82. [DOI:10.1007/s004050000299] [PMID]
- [10] Zraick RI, Kempster GB, Connor NP, Thibeault S, Klaben BK, Bursac Z, et al. Establishing validity of the consensus auditory-perceptual evaluation of voice (CAPE-V). Am J Speech Lang Pathol. 2011; 20(1):14-22 [DOI:10.1044/1058-0360(2010/09-0105)]
 [PMID]
- [11] De Bodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. J Voice. 1997; 11(1):74-80. [DOI:10.1016/S0892-1997(97)80026-4] [PMID]
- [12] Bassich CJ, Ludlow CL. The use of perceptual methods by new clinicians for assessing voice quality. J Speech Hear Disord. 1986; 51(2):125-33. [DOI:10.1044/jshd.5102.125] [PMID]
- [13] Hirano M. Clinical examination of voice. London: Springer; 1981. [Link]
- [14] Silva RS, Simões-Zenari M, Nemr NK. Impacto de treinamento auditivo na avaliação perceptivo-auditiva da voz realizada por estudantes de Fonoaudiologia. J Soc Bras Fonoaudiol. 2012; 24(1):19-25. [DOI:10.1590/S2179-64912012000100005] [PMID]
- [15] Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. J Voice. 2004; 18(3):299-304. [DOI:10.1016/j.jvoice.2003.12.004] [PMID]
- [16] Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. J Speech Hear Res. 1992; 35(3):512-20. [DOI:10.1044/jshr.3503.512] [PMID]
- [17] Lu FL, Matteson S. Speech tasks and interrater reliability in perceptual voice evaluation. J Voice. 2014; 28(6):725-32. [DOI:10.1016/j. jvoice.2014.01.018] [PMID]
- [18] Hammarberg B. Perceptual and acoustic analysis of dysphonia. Huddinge: Huddinge University Hospital; 1986. [Link]
- [19] Yu P, Ouaknine M, Revis J, Giovanni A. Objective voice analysis for dysphonic patients: A multiparametric protocol including acoustic and aerodynamic measurements. J Voice. 2001; 15(4):529-42. [DOI:10.1016/S0892-1997(01)00053-4] [PMID]

- [20] Gerratt BR, Kreiman J, Garellek M. Comparing measures of voice quality from sustained phonation and continuous speech. J Speech Lang Hear Res. 2016; 59(5):994-1001. [DOI:10.1044/2016_JSL-HR-S-15-0307] [PMID] [PMCID]
- [21] Perrachione TK, Furbeck KT, Thurston EJ. Acoustic and linguistic factors affecting perceptual dissimilarity judgments of voices. J Acoust Soc Am. 2019; 146(5):3384-99. [DOI:10.1121/1.5126697] [PMID] [PMCID]
- [22] Fleming D, Giordano BL, Caldara R, Belin P. A language-familiarity effect for speaker discrimination without comprehension. Proc Natl Acad Sci U S A. 2014; 111(38):13795-8. [DOI:10.1073/ pnas.1401383111] [PMID] [PMCID]
- [23] Goggin JP, Thompson CP, Strube G, Simental LR. The role of language familiarity in voice identification. Mem Cognit. 1991; 19(5):448-58.[DOI:10.3758/BF03199567] [PMID]
- [24] Levi SV. Methodological considerations for interpreting the language familiarity effect in talker processing. Wiley Interdiscip Rev Cogn Sci. 2019; 10(2):e1483. [DOI:10.1002/wcs.1483] [PMID]
- [25] Yiu EM, Murdoch B, Hird K, Lau P, Ho EM. Cultural and language differences in voice quality perception: A preliminary investigation using synthesized signals. Folia Phoniatr Logop. 2008; 60(3):107-19. [DOI:10.1159/000119746] [PMID]
- [26] Freitas SV, Pestana PM, Almeida V, Ferreira A. Audio-perceptual evaluation of Portuguese voice disorders—An inter-and intrajudge reliability study. J Voice. 2014; 28(2):210-5. [DOI:10.1016/j. jvoice.2013.08.001] [PMID]
- [27] Chaves CR, Campbell M, Gama AC. The influence of native language on auditory-perceptual evaluation of vocal samples completed by brazilian and Canadian SLPs. J Voice. 2017; 31(2):258-e1-5. [DOI:10.1016/j.jvoice.2016.05.021] [PMID]





مقاله پژوهشی

بررسی توافق بین ارزیابهای فارسی زبان برروی مقیاس ادار کی شنیداری GRBAS

اميرمحمد دانش 💿، •فرهاد ترابينژاد 💿، آرزو صفاريان 💿، جميله ابوالقاسمي 💿، محمد كمالي 🐌، محمد صديق محمودزاده 🌀

۱ .گروه گفتار درمانی، دانشکده علوم توانبخشی، دانشگاه علوم پزشکی ایران، تهران، ایر ان. ۲. مرکز تحقیقات توانبخشی، گروه آسیب شناسی گفتار و زبان، دانشکده علوم توانبخشی، دانشگاه علوم پزشکی ایران، تهران، ایر ان. ۳. گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران، تهران، ایر ان. ۴. گروه مدیریت توانبخشی، دانشکده علوم توانبخشی، دانشگاه علوم پزشکی ایران، تهران، ایران.

> تاریخ دریافت: ۱۳ مهر ۱۴۰۱ تاریخ پذیرش: ۰۶ آذر ۱۴۰۱ تاریخ انتشار: ۵۰ دی ۱۴۰۱

> > كليدواژهها:

صدا، اختلال صوت،

ارزیابی صوت، ارزیابی

ادراکی شنیداری،

مقياس GRBAS

حِکيد

مقدمه ارزیابی ادراکی شنیداری یکی از ارزیابی های مهم برای ارزیابی صدا است. از میان ارزیابیهای ادارکی-شنیداری موجود، مقیاس ادراکی شنیداری GRBAS، توجه بسیاری از محققان و درمانگران را در ایران و سایر نقاط جهان به خود جلب کرده است. مقیاس GRBAS کیفیت صدا را به صورت ادراکی ارزیابی میکند. با این حال، نمونههای صوتی موجود در مقیاس اصلی GRBAS که به عنوان ابزار آموزشی برای ارزیابها استفاده می شود، ژاپنی هستند. تفاوتهای بین زبانها می توانند بر دقت ارزیابی ادراکی تأثیر بگذارند. بر این اسی، این تحقیق با هدف بررسی توافق بین ارزیابها در بین ارزیابهای فارسی زبانی که فقط به نمونههای ژاپنی در نمایه اصلی دسترسی داشتند، انجام شد تا از این طریق بتوانیم اهمیت زبان در این نوع ارزیابی را بررسی کنیم.

مواد و روشها در این مطالعه ۸ آسیب شناس گفتار و زبان با بیش از ۵ سال سابقه بالینی در ارزیابی و درمان اختلالات صدا به عنوان ارزیاب انتخاب شدند. تعداد ۱۳۷ نمونه صوتی شامل کشیده گویی واکه /a/ و خواندن متن استاندارد "پدربزرگ" در اختیار ارزیاب های شرکت کننده قرار گرفت. از ارزیاب ها خواسته شد تا نمونه های صوتی را بر اساس مقیاس ادراکی شنوایی GRBAS امتیاز دهند. نتایج با استفاده از ضریب کاپا کوهن مورد تجزیه و تحلیل آماری قرار گرفت.

یافتهها بیشترین توافق مربوط به پارامتر R و ارزیاب شماره ۵ با (kappa=۰/۵۸۵) و پایینترین میزان توافق مربوط به پارامتر S و ارزیاب شماره ۱ با (kappa=۰/۰۱۸) بود.

نتیجه گیری زبان فارسی و تفاوت آن با زبان ژاپنی میتواند روی قضاوت ارزیابها اثر بگذارد.



Cite this article as Danesh A, Torabinezhad F, Saffarian A, Abolghasemi J, Kamali M, Mahmudzadeh M. Investigating the Inter-rater Agreement of Farsi-speaking Raters on the Grade, Roughness, Breathiness, Asthenia, and Strain Auditory Perceptual Scale. Function and Disability Journal. 2022; 5:E63. http://dx.doi.org/10.32598/fdj.5.63

doi http://dx.doi.org/10.32598/fdj.5.63

» نویسنده مسئول:

دکتر فرهاد ترابینژاد **نشانی**: تهران، دانشگاه علوم پزشکی ایران، دانشکده علوم توانبخشی،گروه گفتار درمانی. **تلفن: ۱۹۳۲۸۰۹ (۹۱۲) ۹۸**+ **رایانامه:** torabinezhad.f@iums.ac.ir