Translation and Validation of the Reflux Symptom Index to Spanish

*Christian Calvo-Henríquez, †,‡Alberto Ruano-Ravina, *,§Pedro Vaamonde, ¶Kaelon Lattomus, *Alicia Sebio, *,§Raquel Fernández-Rodríguez, and *,**Carlos Martin-Martin, *†‡§**Santiago de Compostela, Spain, and ¶Tucson, Arizona

Summary: Objective. In this study, we have created and validated the Spanish translation of the Reflux Symptom Index (Sp-RSI).

Materials and Methods. RSI was translated according to the World Health Organization recommendation for the translation and adaptation of instruments. The final version of the Sp-RSI was administered twice (day 0 and day 7) to 150 participants with and without laryngopharyngeal reflux symptoms. Reliability was measured with Cronbach α, and temporal stability was measured with intraclass correlation coefficient.

Results. One hundred fifty participants were included. Internal consistency of the Sp-RSI measured with Cronbach α was 0.872. A subgroup analysis was performed according to age (<55 and >55 years) and sex (female, male), with Cronbach α values of 0.854, 0.869, 0.852, and 0.897, respectively. The intraclass correlation coefficient was 0.987. According to age (<55 and >55 years) and sex (female, male), the intraclass correlation coefficients were 0.990, 0.984, 0.987, and 0.987, respectively.

Conclusions. The Sp-RSI reported a very high internal consistency and reliability. It is, therefore, a valid, inexpensive, and reliable tool for assessing laryngopharyngeal reflux symptoms and may be used for screening among the Spanish-speaking medical community.

Key Words: Extraesophageal reflux—Laryngopharyngeal reflux—LPR—Pepsin—RSI—Reflux Symptom Index.

INTRODUCTION

Patients with gastroesophageal reflux disease may complain of typical and atypical symptoms.[1] Atypical symptoms include those caused by the backflow of gastric content into the throat, such as pepsin or acid.[2] Those symptoms are classified as laryngopharyngeal reflux (LPR),[3] which can account for 10% of patients attending otolaryngology consultations.[4]

LPR diagnosis is a challenge to otolaryngologists, and the diagnostic methodology is not well established yet. Multichannel intraluminal impedance and 24-hour dual-probe pH-metry are considered the gold standard tests.[5] Nevertheless, these tests have shown some flaws in diagnosing LPR.[6−8] Furthermore, these tests cannot be performed in all patients suspected of presenting LPR given their invasiveness and high cost. For this reason, otolaryngologists usually overdiagnose and overtreat LPR, leading to an economic burden[9] and a health risk.[10] Currently, new strategies are being developed, including pepsin detection as a diagnostic test for LPR,[11] but their role in the diagnostic protocol of LPR is still unclear.

Besides the diagnostic and therapeutic protocol of the LPR not being clear, it is widely accepted that the suspected diagnoses rely on symptoms and exploratory findings.[5]

Exploratory findings and clinical symptoms were addressed by Belafsky et al in 2002 in the Reflux Finding Score[12] and the Reflux Symptom Index (RSI).[13] The RSI is a self-administered nine-item symptom questionnaire for the assessment of symptoms related to LPR. The scale for each individual item ranges from 0 (no problem) to 5 (severe problem), with a maximum total score of 45 points.

RSI and Reflux Finding Score can be used as screening tests in patients suspected of suffering LPR. Those who fulfill the criteria may be subjected to further tests.

RSI is a valuable tool. Despite the fact that it has been called into question for its low specificity,[14,15] it has shown high sensitivity when detecting LPR.[16] In addition, it is easy to use, inexpensive, and it does not require special equipment. For all these reasons, it is widely used as a diagnostic tool in the majority of studies regarding LPR. The questionnaire has been translated into several languages, including Hebrew, Greek, Chinese, French, Filipino, Italian, and Arabic,[16−22] but not into Spanish yet. Two translations of RSI to Spanish have been published,[23,24] but they have not been validated yet.

Spanish is the second most spoken language in the world with regard to the number of native speakers. The aim of this study was to document the reliability and validity of our Spanish translation of RSI (Sp-RSI).

MATERIALS AND METHODS

RSI translation

RSI is a questionnaire designed to assist in diagnosing LPR. It includes nine related symptoms and each item can be rated

Puntúe del **0 (nada) al 5 (mucho)**, cuánto le han afectado los síntomas o problemas de la lista siguiente **durante el último mes.**

Síntoma	0	1	2	3	4	5
1. Afonía u otros cambios en la voz.						
2. Carraspeo, o necesidad de aclarar la garganta.						
3. sensación de tragar moco que cae por detrás de la nariz.						
4. Dificultad para tragar comida, líquidos o pastillas.						
5. Tos después de comer o tras estar tumbado.						
6. Dificultad para respirar o episodios de atragantamiento/ahogo.						
7. Ataques de tos.						
8. Sensación de bulto o de tener algo pegado en la garganta. Sensación de tener algo a medio tragar.						
9. Sensación de que el ácido del estomago sube a la garganta. Quemazón u opresión en el pecho.						
TOTAL						

FIGURE 1. Spanish translation of RSI (Sp-RSI).

from 0 to 5. A total score of 13 or more points represents a high suspicion of LPR.

We followed the recommendations of the World Health Organization for the translation and adaptation of instruments, the following way: (1) A first translation (**v1**) was carried out by a multidisciplinary team, which included two general otolaryngologists, a senior otolaryngology professor, two experts in swallowing and voice disorders, and an epidemiologist, all native speakers of Spanish, and a native English speaker. (2) **v1** was sent to a panel of six experts in swallowing and voice disorders throughout Spain. Two versions of the test (**v2** and **v3**) were developed until a general consensus was reached. (3) The third version (**v3**) was sent to a bilingual English-Spanish speaker general practitioner for back-translation. No changes were made. (4) Afterward, a pretest cognitive interview was carried out in 50 patients to check the comprehensibility of the translated questionnaire. A fourth version (**v4**) of the test was thus obtained (Figure 1), which was sent again to the panel of experts to obtain a general consensus.

Participants

The objective of this study was to assess the reliability of the Spanish version of the RSI scale. To this end, from December 2016 to October 2017, 150 participants with and without LPR symptoms were recruited from the otolaryngology outpatient clinic of a tertiary-level referral center in Spain.

Exclusion criteria

Patients who refused to participate in the study and patients who had started treatment before the second test were excluded, as were participants who did not have sufficient cognitive skills to complete the questionnaires. Cognitive skills were assessed by interview and from pre-existent medical records of cognitive impairment.

Lost to follow-up criteria

Participants were telephoned on the seventh day after having completed the first questionnaire. If the participants did not answer, they were called again on the eighth day. If there still was no answer, they were considered a loss to follow-up.

Variables

Demographic and clinical variables

In the first interview, clinical and demographic variables were recorded. Age, sex, profession, voice professional (yes/no), smoking habit, drinking habit, and history of reflux disease were considered.

Sp-RSI-1 and Sp-RSI-2

The results of Sp-RSI on day 0 (Sp-RSI-1) and day 7 (Sp-RSI-2) were recorded and classified into four intervals for the purpose of the study: interval 1: 0−12 points; interval 2: 13−23 points; interval 3: 24−34 points: interval 4: 35−45 points.

Statistical methods and psychometric evaluation of the instruments

Reliability

Reliability refers to consistency or dependability after repeated measurements. It can be evaluated by ascertaining consistency, temporal stability, and interobserver concordance.

Consistency. Consistency refers to the extent to which each question of the scale relates to the rest. Cronbach α is the most widely used method to measure it. Values over 0.80 suggest a strong construct validity, which gets better the closer it is to 1. Cronbach α was measured using the first RSI to prevent memory bias.

Temporal stability. This is also known as test-retest reliability; it refers to the concordance between the scores of repeated measurements from the same participant. It can be assessed by the intraclass correlation coefficient. A correlation over 0.80 is considered a high correlation value.

A repeated measurement was made after a 7-day interval to prevent memory bias, but, at the same time, ensure comparability between both tests. Both participants and interviewer did not have access to the first RSI questionnaire to prevent memory bias. Patients were asked not to follow any treatment before the second test was performed to ensure comparability between both measurements.

We used a mixed model containing random effects (participants) and fixed effects (questionnaire items), as well as the intraclass correlation coefficient and an absolute agreement model.

Statistical analysis was conducted with *SPSS* for Macintosh Version 23.0 (IBM Corp, Armonk, NY). Internal consistency was measured with Cronbach α and temporal stability by using the intraclass correlation coefficient ($P < 0.05$).

Institutional review board approval

The study was performed in accordance with the ethical standards laid down in the Declaration of Helsinki. The study protocol was approved by the Research and Ethics Committee of the Hospital Complex of Santiago de Compostela, register code 2016/395.

RESULTS

Participants

One hundred fifty participants were included. Forty-three participants were excluded because they were lost to follow-up. No participants were excluded for following treatment before the second tests were performed. Eighty-five were women and 65 were men. Mean age was 54.09 and median age 55 (range 13−97). Percentiles 25 and 75 were 40.75 and 68 years old, respectively.

RSI distribution

The average RSI result was 10.4 and the median was 7; range 38 (0−38). Percentiles 25 and 75 were 0 and 17.

There was an asymmetry of 0.913, which can be read as positive. The distribution of results by interval can be seen in Figure 2.

Consistency

Internal consistency of Sp-RSI measured with Cronbach α was 0.872 for nine items. Internal consistency was still over 0.844 after removing every item of the test (Table 1), and in none of those cases was it greater.

Consistency was measured within a subgroup analysis according to age. Patients were divided into two groups, <55 years of age, 70 patients, and ≥55 years of age, including 80 patients. The Cronbach α values were 0.854 and 0.869, respectively. After removing every item of the test, minimum
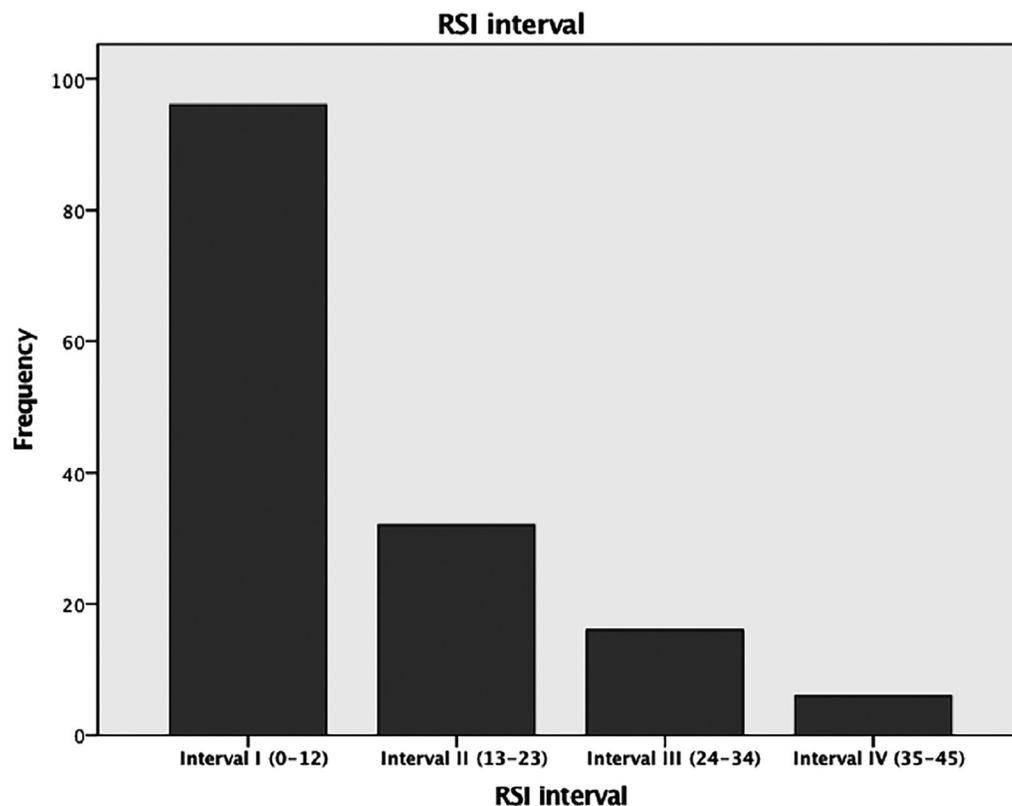

FIGURE 2. RSI interval distribution.

TABLE 1.
Consistency of the Spanish Version of the RSI Measured With Cronbach α

	Total	<55 y	≥55 y	Male	Female
Cronbach α	0.872	0.854	0.869	0.897	0.852
Cronbach α if the element has been suppressed					
RSI_1	0.862	0.841	0.861	0.890	0.840
RSI_2	0.844	0.827	0.840	0.875	0.820
RSI_3	0.858	0.847	0.850	0.882	0.841
RSI_4	0.866	0.849	0.864	0.889	0.849
RSI_5	0.862	0.847	0.858	0.892	0.838
RSI_6	0.867	0.843	0.869	0.895	0.846
RSI_7	0.847	0.823	0.848	0.875	0.826
RSI_8	0.849	0.834	0.844	0.887	0.819
RSI_9	0.860	0.836	0.859	0.886	0.840

values were 0.823 and 0.840, respectively. In no case was consistency greater (Table 1).

Regarding gender, 85 females obtained a 0.852 Cronbach α result, and for 65 males, it was 0.897.

Intraclass correlation coefficient
The intraclass correlation coefficient was 0.987 (confidence interval [CI] 95%: 982−991).

In women, it was 0.987 (CI 95%: 980−992), and, in men, it was also 0.987 (CI 95%: 979−992); there were no differences by sex.

With regard to age, ICC in <55-year-olds was 0.990 (CI 95%: 983−994), and in ≥55-year-olds, it was 0.984 (CI 95%: 975−990).

The maximum result is 1; therefore, these results categorize this test as having a very high correlation.

DISCUSSION
This is the first validated Spanish version of RSI. The psychometric evaluation of the translated test shows excellent results in consistency, reliability, and construct validity. This Sp-RSI has also shown an excellent result when studying subgroups of sex and age.

There are two previous translations of RSI into Spanish.[23,24] We considered those to our translation, but we found some issues with them. Olavarría and Cortez[23] used dialectal variations, specifically in questions 6, 8, and 9. Furthermore, they used confusing concepts and words that we detected in our pretest cognitive interview, specifically in questions 1, 3 and 7. Vázquez de la Iglesia et al[24] used only eight questions, as they suppressed the fourth question. Also, they score each question from 0 to 2, instead of 0−5. We believe some parts of the translations are not accurate, specifically questions 2, 3, and 5 (originally 6) and we found confusing concepts in questions 1, 3, and 6 (originally 7).

A high RSI value has been associated with other factors such as allergies,[25] asthma,[26] psychiatric disorders,[27] or tobacco,[28] among others. Unlike other authors, we have not excluded

those patients. Their inclusion increases the external validity of our work, and it does not affect its internal validity since we did not diagnose LPR in this study.

Furthermore, other authors have suggested modifications of the original RSI test,[15,16,23] whereas we have not changed the original version.

In our version, Cronbach α was high for the general population (0.872) and also when studying subgroups of sex and age. Internal consistency was still high after removing every item of the test and in none of those cases was it greater than 0.872. It can be read as there being no item with a higher weight in consistency than the rest. These results are similar to those published in the case of the Greek (0.865)[16] and French translations (0.85)[20]; lower than in the Filipino (0.94)[21] and Italian (0.99)[22] ones, but higher than in the Chinese (0.715)[18] and Arabic (0.72)[19] translations.

Consistency was slightly higher in the male population (0.897 male, 0.852 female), even though it was still high in females. This difference could be explained because a high percentage of the female population were menopausal females with nonspecific throat symptoms. It is known that the menopausal syndrome includes pharynx and larynx symptoms,[29,30] which can diminish consistency in this subgroup of patients.

Test-retest reliability measured with the intraclass correlation coefficient was also high, 0.987 (CI 95%: 982−991). There were no differences by sex, but there was a slight difference by age, with a better result for patients under 55 years of age, 0.990 (CI 95%: 983−994). These findings are to be expected, given that elderly patients have more difficulty in understanding, expressing, and grading their symptoms. However, it was a slight difference, and the temporal reliability measured by ICC was still very high, 0.984 (CI 95%: 975−990).

These results are similar to those published in the Chinese (0.750-0.971),[18] Filipino (0.92),[21] and Italian versions (0.99)[22]; and noticeably higher than those found in the French translation (0.78 cases, 0.71 controls),[20] Arabic translation (0.799),[19] and those in the original version by Belafsky et al (0.81).[13] Although Printza et al used different parameters in the Greek translation (Spearman-Brown; Guttman split-half reliability coefficients), they also obtained an adequate test-retest reliability.[16]

The second interview was a telephone one, and given that it might have brought about bias, we tried our best to control it. Neither the interviewer nor the participant had access to the responses of the first RSI questionnaire during the second interview.

A strong point of our study was the sample. We included a large sample of the same health area, with the same access to health services.

An apparent weakness of our work was the asymmetry of the distribution of participants. Approximately two-thirds of subjects had an RSI under 13 (Figure 2). Nevertheless, this does not debilitate the consistency of the test; it makes it stronger. Participants with low rates in RSI have less intrinsic motivation to remember the answers to the test. However, the consistency and temporal stability were high. In this study, we

did not want to overweigh higher scores of RSI, we wanted to use this scale in the whole range of results.

Our translation was done in Spain. Even though Latin American Spanish is almost the same as mainland Spain Spanish, there are slight differences between them. Two of the main authors are from Latin America, so we have tried to use words and expressions used in these countries.

In our study, we have not measured construct and criterion validity, which is a determination of the degree to which a test measures what it has been designed to. Even though construct validity has been ascertained in other papers,[16,18,21] we have designed this work to measure internal consistency and reliability exclusively.

CONCLUSIONS

RSI has shown an excellent correlation with LPR.[31] The RSI questionnaire is inexpensive, fast and easy to administer, and it is available to every otolaryngologist and primary care consultant. Our Sp-RSI reported good, highly reproducible construct validity. Therefore, we hope that Sp-RSI will be made available to all the Spanish-speaking medical community.

REFERENCES

1. Groome M, Cotton JP, Borland M, et al. Prevalence of laryngopharyngeal reflux in a population with gastroesophageal reflux. *Laryngoscope*. 2007;117:1424–1428.
2. Vaezi MF, Hicks DM, Abelson TI, et al. Laryngeal signs and symptoms and gastroesophageal reflux disease (GERD): a critical assessment of cause and effect association. *Clin Gastroenterol Hepatol*. 2003;1:333–344.
3. Koufman J, Sataloff RT, Toohill R. Laryngopharyngeal reflux: consensus conference report. *J Voice*. 1996;10:215–216.
4. Koufman JA. The otolaryngologic manifestations of gastroesophageal reflux disease (GERD): a clinical investigation of 225 patients using ambulatory 24-hour pH monitoring and an experimental investigation of the role of acid and pepsin in the development of laryngeal injury. *Laryngoscope*. 1991;101(4 suppl 53 pt 2):1–78.
5. Asaoka D, Nagahara A, Matsumoto K, et al. Current perspectives on reflux laryngitis. *Clin J Gastroenterol*. 2014;7:471–475.
6. Jacob P, Kahrilas PJ, Herzon G. Proximal esophageal pH-metry in patients with "reflux laryngitis". *Gastroenterology*. 1991;100:305–310.
7. Vaezi MF, Schroeder PL, Richter JE. Reproducibility of proximal probe pH parameters in 24-hour ambulatory esophageal pH monitoring. *Am J Gastroenterol*. 1997;92:825–829.
8. Vincent DA, Garrett JD, Radionoff SL, et al. The proximal probe in esophageal pH monitoring: development of a normative database. *J Voice*. 2000;14:247–254.
9. Francis DO, Rymer JA, Slaughter JC, et al. High economic burden of caring for patients with suspected extraesophageal reflux. *Am J Gastroenterol*. 2013;108:905–911.
10. Xie Y, Bowe B, Li T, et al. Risk of death among users of proton pump inhibitors: a longitudinal observational cohort study of United States veterans. *BMJ Open*. 2017;7:e015735.
11. Calvo-Henríquez C, Ruano-Ravina A, Vaamonde P, et al. Is pepsin a reliable marker of laryngopharyngeal reflux? A systematic review. *Otolaryngol Head Neck Surg*. 2017;157:385–391.
12. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). *Laryngoscope*. 2001;111:1313–1317.
13. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). *J Voice*. 2002;16:274–277.
14. Park KH, Choi SM, Kwon SUK, et al. Diagnosis of laryngopharyngeal reflux among globus patients. *Otolaryngol Head Neck Surg*. 2006;134:81–85.
15. Cathcart RA, Steen N, Natesh BG, et al. Non-voice-related throat symptoms: comparative analysis of laryngopharyngeal reflux and globus pharyngeus scales. *J Laryngol Otol*. 2011;125:59–64.
16. Printza A, Kyrgidis A, Oikonomidou E, et al. Assessing laryngopharyngeal reflux symptoms with the Reflux Symptom Index: validation and prevalence in the Greek population. *Otolaryngol Head Neck Surg*. 2011;145:974–980.
17. Cohen JT, Gil Z, Fliss DM. [The reflux symptom index—a clinical tool for the diagnosis of laryngopharyngeal reflux]. *Harefuah*. 2005;144:826–829 912.
18. Li J, Zhang L, Zhang C, et al. Linguistic adaptation, reliability, validation, and responsivity of the Chinese Version of Reflux Symptom Index. *J Voice*. 2016;30:104–108.
19. Farahat M, Malki KH, Mesallam TA. Development of the Arabic version of Reflux Symptom Index. *J Voice*. 2012;26:814.e15–814.e19.
20. Lechien JR, Huet K, Finck C, et al. Validity and reliability of a French version of reflux symptom index. *J Voice*. 2017;31:512.e1–512.e7.
21. Lapeña JFF, Ambrocio GMC, Carrillo RJD. Validity and reliability of the Filipino Reflux Symptom Index. *J Voice*. 2017;31:387.e11–387.e16.
22. Schindler A, Mozzanica F, Ginocchio D, et al. Reliability and clinical validity of the Italian Reflux Symptom Index. *J Voice*. 2010;24:354–358.
23. Olavarría LC, Cortez VP. Reflujo faringolaríngeo: Revisión de los nuevos métodos diagnósticos. *Rev Otorrinolaringol Cir Cabeza Cuello*. 2015; 75:61–66.
24. Vázquez de la Iglesia F, Fernández González S, Gómez Mde L. [Laryngopharyngeal reflux: correlation between symptoms and signs by means of clinical assessment questionnaires and fibroendoscopy. Is this sufficient for diagnosis?]. *Acta Otorrinolaringol Esp*. 2007;58:421–425.
25. Brauer DL, Tse KY, Lin JC, et al. The utility of the reflux symptom index for diagnosis of laryngopharyngeal reflux in an allergy patient population. *J Allergy Clin Immnuol Pract*. 2017;6:132–138 e1.
26. Hamdan AL, Jaffal H, Btaiche R, et al. Laryngopharyngeal symptoms in patients with asthma: a cross-sectional controlled study. *Clin Respir J*. 2016;10:40–47.
27. Oyer SL, Anderson LC, Halum SL. Influence of anxiety and depression on the predictive value of the Reflux Symptom Index. *Ann Otol Rhinol Laryngol*. 2009;118:687–692.
28. Branski RC, Bhattacharyya N, Shapiro J. The reliability of the assessment of endoscopic laryngeal findings associated with laryngopharyngeal reflux disease. *Laryngoscope*. 2002;112:1019–1024.
29. Furuta S, Nishimoto K, Deguchi K, et al. Relationship between abnormal sensation in the throat and menopause. *Auris Nasus Larynx*. 1996;23:69–74.
30. Hamdan A-L, Ziade G, Tabet G, et al. Vocal symptoms and acoustic findings in menopausal women in comparison to pre-menopause women with body mass index as a confounding variable. *J Menopausal Med*. 2017;23:117–123.
31. Habermann W, Schmid C, Neumann K, et al. Reflux symptom index and reflux finding score in otolaryngologic practice. *J Voice*. 2012;26: e123–e127.