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 burden9 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 Score12 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...
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 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 $\alpha$ 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 $\alpha$ 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 $\alpha$ and temporal stability by using the intraclass correlation coefficient ($P < 0.05$).

**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-RI measured with Cronbach $\alpha$ 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 $\geq$55 years of age, including 80 patients. The Cronbach $\alpha$ values were 0.854 and 0.869, respectively. After removing every item of the test, minimum

![FIGURE 2. RSI interval distribution.](image-url)
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 Cortez²² used dialectal variations, specifically in questions 6, 8, and 9. Furthermore, they used confusing concepts and words that we detected in our pre-test 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,asthma,psychiatric disorders, or tobacco, 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 it was 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)¹⁷ 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

### TABLE 1.

<table>
<thead>
<tr>
<th>TABLE 1. Consistency of the Spanish Version of the RSI Measured With Cronbach α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach α</td>
</tr>
<tr>
<td>Cronbach α if the element has been suppressed</td>
</tr>
<tr>
<td>RSI_1</td>
</tr>
<tr>
<td>RSI_2</td>
</tr>
<tr>
<td>RSI_3</td>
</tr>
<tr>
<td>RSI_4</td>
</tr>
<tr>
<td>RSI_5</td>
</tr>
<tr>
<td>RSI_6</td>
</tr>
<tr>
<td>RSI_7</td>
</tr>
<tr>
<td>RSI_8</td>
</tr>
<tr>
<td>RSI_9</td>
</tr>
</tbody>
</table>
did not want to overpower 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, we have designed this work to measure internal consistency and reliability exclusively.

**CONCLUSIONS**

RSI has shown an excellent correlation with LPR. 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.

**Acknowledgments**

The authors thank the panel of experts who collaborated in this translation: Dr. M. Landa-Aranzabal, Dr. R. González-Herranz, Dr. I. García-López, Dr. L. Rodríguez-Pérez, Dr. F. Subirana-Pozo, Dr. R. Fernández-Lopez, and Dr. Aldo Rosario-Ortiz.

**REFERENCES**