INTRODUCTION

Laryngopharyngeal reflux (LPR) is defined as the reflux of gastric content into the larynx and pharynx. LPR is an extraesophageal manifestation of gastroesophageal reflux disease (GERD) and there are a large number of data on the growing prevalence of laryngopharyngeal symptoms in up to 60% in patients with GERD. Laryngopharyngeal reflux (LPR) is an extraesophageal manifestation of gastroesophageal reflux disease. Endoscopic assessment of LPR is needed for convenient diagnosis and documentation of treatment efficacy. The aim of this study was to investigate the interrater reliability of LPR among endoscopists based on endoscopic laryngeal images.

Material and Methods: Nineteen endoscopists participated in this study. Before the test, they completed an intensive education program by an otorhinolaryngologist on the reflux finding score (RFS), which is a validated laryngoscopic assessment of LPR. A total of 100 endoscopic laryngeal images were used for 3 tests of RFS. Cohen’s and Fleiss’ kappa coefficients were used to determine the degree of interrater agreement in the diagnosis of LPR.

Results: In the first test, the mean of Cohen’s kappa coefficients for LPR diagnosis between the otorhinolaryngologist and each of the 19 endoscopists was 0.3. In the second test, after additional education, the mean kappa value was 0.32. Fleiss’ kappa coefficients for diagnosis of LPR among the 19 endoscopists in the first and second tests were 0.30 and 0.26, respectively.

Conclusions: A short-term education program for endoscopists did not result in an improvement of accuracy in the diagnosis of LPR. Further studies using advanced educational programs for endoscopists are required.

Key Words: Gastroesophageal reflux; Endoscopy; Laryngopharyngeal reflux
Endoscopic findings of the larynx (Fig. 1). The aim of this study was to evaluate interrater reliability of the diagnosis of LPR by using the RFS based on endoscopic laryngeal findings not only between an otolaryngologist and endoscopists but also among nineteen endoscopists.

MATERIALS AND METHODS

1. RFS

The RFS is a validated clinical severity scale for LPR diagnosis. The 8 items of the RFS were derived from a pool of the most common laryngeal findings of patients with LPR and its scale ranges from 0 to a maximum of 26 (Table 1). Although each item on the RFS is entirely subjective, the overall finding score reliably documents improvement with antireflux therapy. The total RFS greater than 7 is considered as LPR.6

2. Study design

One otolaryngologist and 19 endoscopists, who work in university hospitals in Gangwon Province, South Korea, participated in this study. The otolaryngologist was an expert on LPR who became a reference provider. For the education of the RFS, an otolaryngologist delivered a 2-hour lecture to 19 endoscopists using educational slides including typical laryngoscopic pictures on the anatomy of the larynx and individual items of the RFS.

Endoscopic laryngeal pictures were extracted from the endoscopy database of our center by an endoscopist (L.Y.S) and selected by an otolaryngologist (K.D.K.).

Two processes were performed independently. A total of 100 laryngeal pictures with good quality were finally used for this study. Inclusion criteria of endoscopic laryngeal pictures were as follows: (1) picture focused well; (2) picture captured when vocal cords were opened; (3) picture showing larynx fully. Exclusion criteria were as follows: (1) blurred picture or picture captured too far from larynx; (2) picture captured when vocal cords were closed.

### Table 1. Reflux Finding Score

| Parameter                  | Score  |
|----------------------------|--------|
| Subglottic edema           | 0=absent |
|                           | 2=present |
| Ventricular obliteration   | 0=absent |
|                           | 2=present |
|                           | 4=complete |
| Erythema/hyperemia         | 0=absent |
|                           | 2=arytenoids only |
|                           | 4=diffuse |
| Vocal fold edema           | 0=absent |
|                           | 1=mild |
|                           | 2=moderate |
|                           | 3=severe |
|                           | 4=polyloid |
| Diffuse laryngeal edema    | 0=absent |
|                           | 1=mild |
|                           | 2=moderate |
|                           | 3=severe |
|                           | 4=obstructing |
| Posterior commissure hypertrophy | 0=absent |
|                           | 1=mild |
|                           | 2=moderate |
|                           | 3=severe |
| Granuloma/granulation tissue | 0=absent |
|                           | 2=present |
| Thick endolaryngeal mucus  | 0=absent |
|                           | 2=present |
Fig. 2. Flow diagram of study design.

Table 2. Characteristics of Nineteen Endoscopists

| Rater | Age (yr) | Experience of endoscopy | Total times of education (min) |
|-------|----------|--------------------------|-------------------------------|
| 1     | 42       | 1.000 ~ 2.000            | 30 ~ 60                       |
| 2     | 39       | 9                        | >120                          |
| 3     | 44       | 8                        | 30 ~ 60                       |
| 4     | 39       | 6                        | >3,000                        |
| 5     | 40       | 6                        | >3,000                        |
| 6     | 33       | 1                        | 500 ~ 1,000                   |
| 7     | 47       | 15                       | >3,000                        |
| 8     | 37       | 5                        | >3,000                        |
| 9     | 37       | 1                        | >3,000                        |
| 10    | 43       | 8                        | >3,000                        |
| 11    | 35       | 3                        | >3,000                        |
| 12    | 36       | 1                        | 1,000 ~ 2.000                 |
| 13    | 34       | 1                        | 1,000 ~ 2.000                 |
| 14    | 37       | 1                        | 2.000 ~ 3.000                 |
| 15    | 40       | 6                        | >3,000                        |
| 16    | 35       | 1                        | >3,000                        |
| 17    | 45       | 8                        | >3,000                        |
| 18    | 43       | 1                        | 2,000 ~ 3.000                 |
| 19    | 31       | 1                        | 2,000 ~ 3.000                 |

Fig. 2. Flow diagram of study design.
times of education before the test was varied according to the individual concern degree (Table 2).

In the first test, the average values of Cohen’s kappa coefficients for LPR diagnosis between one otorhinolaryngologist and each of 19 endoscopists was 0.3. In the second test after additional feedback education for each endoscopists, mean kappa value was slightly increased to 0.32. Many factors including age, endoscopic experience, and time of education were not associated with these scores. The kappa values of eleven endoscopists were increased in the second test (mean kappa value of 1st test ~ 2nd test from 0.139 to 0.366) after feedback education. In the second test, the mean kappa values of 16 endoscopists (84.2%) except for 3 endoscopists (rater 3, 4, and 6) who had lowest kappa value, were fair to moderate (Table 3).

In analysis of each parameter of RFS in the first and second tests, the mean kappa values of the score for ventricular obliteration were 0.56 and 0.59, respectively. The scores for 4 parameters including ventricular obliteration, vocal fold edema, posterior commissure hypertrophy, and thick endolaryngeal mucus had relatively higher kappa value than other parameters. On the other hand, the scores for other 4 parameters, including subglottic edema, erythema, diffuse laryngeal edema, and granuloma had lower kappa values (Table 4).

The Fleiss’ kappa coefficient of diagnosis of LPR among

Table 3. Interrater Reliability of Diagnosis of LPR Using RFS between Diagnosis of Otorhinolaryngologist and Endoscopists

| Rater | Agreement (%) | Cohen's kappa | Agreement (%) | Cohen's kappa |
|-------|--------------|---------------|--------------|---------------|
| 1     | 76           | 0.407 (moderate) | 62           | 0.210 (fair)  |
| 2     | 64           | 0.288 (fair)   | 34           | 0.325 (fair)  |
| 3     | 72           | 0.361 (fair)   | 68           | 0.195 (slight) |
| 4     | 70           | 0.269 (fair)   | 64           | 0.064 (slight) |
| 5     | 72           | 0.458 (moderate) | 70         | 0.376 (fair)  |
| 6     | 68           | 0.209 (fair)   | 66           | 0.101 (slight) |
| 7     | 66           | 0.349 (fair)   | 70           | 0.436 (moderate) |
| 8     | 68           | 0.353 (fair)   | 74           | 0.412 (moderate) |
| 9     | 70           | 0.288 (fair)   | 78           | 0.554 (moderate) |
| 10    | 60           | 0.064 (slight) | 72           | 0.338 (fair)  |
| 11    | 76           | 0.466 (moderate) | 72         | 0.410 (moderate) |
| 12    | 54           | 0.220 (fair)   | 66           | 0.231 (fair)  |
| 13    | 72           | 0.308 (fair)   | 82           | 0.581 (moderate) |
| 14    | 78           | 0.478 (moderate) | 66         | 0.311 (fair)  |
| 15    | 66           | 0.304 (fair)   | 72           | 0.358 (fair)  |
| 16    | 64           | 0.237 (fair)   | 58           | 0.247 (fair)  |
| 17    | 66           | 0.148 (slight) | 72           | 0.247 (fair)  |
| 18    | 74           | 0.456 (moderate) | 76         | 0.415 (moderate) |
| 19    | 52           | 0.029 (slight) | 62           | 0.302 (fair)  |

LPR, laryngopharyngeal reflux; RFS, reflux finding score.

Table 4. Interrater Reliability of Eight Parameters of Reflux Finding Score System in Second Test

| Rater | Subglottic edema | Ventricular obliteration | Erythema | Vocal fold edema | Diffuse laryngeal edema | Posterior commissure hypertrophy | Granuloma | Thick endolaryngeal mucus |
|-------|------------------|--------------------------|----------|------------------|-------------------------|----------------------------------|-----------|-------------------------|
| 1     | 0.348            | 0.406                    | 0.015    | 0.115            | 0                       | 0.120                            | 0         | 0.286                   |
| 2     | 0.211            | 0.572                    | 0.480    | 0.286            | 0.560                   | 0.263                            | 0         | 0                      |
| 3     | 0                | 0.752                    | 0.097    | 0.358            | 0.040                   | 0.470                            | 0         | 0.242                   |
| 4     | 0.328            | 0.712                    | 0.216    | 0.249            | 0.080                   | 0.556                            | 0         | 0.468                   |
| 5     | 0.228            | 0.576                    | 0.548    | 0.329            | 0.440                   | 0.307                            | 0         | 0.202                   |
| 6     | 0.177            | 0.536                    | 0        | 0.418            | 0                       | 0.256                            | 0         | 0                      |
| 7     | 0.166            | 0.205                    | 0.185    | 0.052            | 0.042                   | 0.071                            | 0         | 0                      |
| 8     | 0.400            | 0.495                    | 0.382    | 0.307            | 0.240                   | 0.216                            | 0         | 0.648                   |
| 9     | 0.194            | 0.796                    | 0.688    | 0.497            | 0.480                   | 0.194                            | 0         | 0.790                   |
| 10    | 0.565            | 0.630                    | 0.448    | 0.245            | 0                       | 0.303                            | 0         | 0.479                   |
| 11    | 0.126            | 0.754                    | 0.662    | 0.040            | 0.240                   | 0.293                            | 0         | 0                      |
| 12    | 0.275            | 0.490                    | 0.109    | 0.438            | 0.640                   | 0.490                            | 0         | 0.319                   |
| 13    | 0.257            | 0.759                    | 0.234    | 0.606            | 0.080                   | 0.216                            | 0         | 0.790                   |
| 14    | 0.091            | 0.398                    | 0.019    | 0.381            | 0.280                   | 0.042                            | 0         | 0.408                   |
| 15    | 0.275            | 0.548                    | 0.565    | 0.089            | 0.160                   | 0.107                            | 0         | 0                       |
| 16    | 0.307            | 0.678                    | 0.376    | 0.520            | 0                       | 0.129                            | 0         | 0.233                   |
| 17    | 0.004            | 0.264                    | 0.156    | 0.290            | 0.065                   | 0.160                            | 0         | 0.657                   |
| 18    | 0.397            | 0.802                    | 0.176    | 0.261            | 0.160                   | 0.427                            | 0         | 0.408                   |
| 19    | 0.228            | 0.839                    | 0.138    | 0.394            | 0.600                   | 0.168                            | 0         | 0.468                   |
Table 5. Interrater Reliability between 19 Endoscopists (Fleiss’s Kappa)

| Tests            | Subglottic edema | Ventricular obliteration | Erythema | Vocal fold edema | Diffuse laryngeal edema | Posterior commissure hypertrophy | Granuloma | Thick endolaryngeal mucus | Diagnosis of LPR |
|------------------|------------------|--------------------------|----------|------------------|--------------------------|---------------------------------|-----------|---------------------------|-----------------|
| 1st Fleiss’s Kappa | 0.2968           | 0.5454                   | 0.2142   | 0.161            | 0.2800                   | 0.2736                          | 0.3908    | 0.3291                    | 0.3055          |
| Z                | 27.45            | 50.43                    | 19.80    | 14.89            | 25.89                    | 25.30                           | 36.13     | 30.43                     | 28.25           |
| 2nd Fleiss’s Kappa | 0.4077           | 0.4702                   | 0.2150   | 0.1804           | 0.1592                   | 0.2020                          | 0.1192    | 0.3195                    | 0.2656          |
| Z                | 37.50            | 43.29                    | 19.79    | 16.59            | 14.67                    | 18.66                           | 10.99     | 29.45                     | 24.56           |

LPR, laryngopharyngeal reflux.

Table 6. Test-Retest Reliability (Correlation Coefficient)

| Rater | Subglottic edema | Ventricular obliteration | Erythema | Vocal fold edema | Diffuse laryngeal edema | Posterior commissure hypertrophy | Granuloma | Thick endolaryngeal mucus | Total score | Diagnosis of LPR |
|-------|------------------|--------------------------|----------|------------------|--------------------------|---------------------------------|-----------|---------------------------|-------------|-----------------|
| 1     | 0.630            | 0.522                    | −0.183   | 0.479            | 0.566                    | 0.720                           | 0.634     | 0.839                     | 0.559       | 0.439           |
| 2     | 0.977            | 0.934                    | 0.785    | 0.831            | 0.913                    | 0.719                           | 0.000     | 0.000                     | 0.924       | 0.900           |
| 5     | 0.570            | 0.750                    | 0.578    | 0.610            | 0.650                    | 0.661                           | 0.000     | 0.500                     | 0.709       | 0.651           |
| 7     | 0.784            | 0.854                    | 0.540    | 0.784            | 0.936                    | 0.765                           | 1.000     | 0.717                     | 0.706       | 0.341           |
| 8     | 0.756            | 0.812                    | 0.587    | 0.459            | 0.557                    | 0.666                           | 0.000     | 0.880                     | 0.688       | 0.679           |
| 9     | 1.000            | 1.000                    | 0.000    | 0.979            | 0.987                    | 1.000                           | 0.000     | 0.000                     | 0.860       | 0.747           |
| 10    | 0.847            | 0.705                    | 0.730    | 0.780            | 0.856                    | 0.726                           | 0.000     | 0.918                     | 0.837       | 0.798           |
| 11    | 0.792            | 0.953                    | 0.840    | 0.377            | 0.750                    | 0.625                           | 0.000     | 0.885                     | 0.906       | 0.792           |
| 12    | 0.000            | 0.710                    | 0.545    | 0.286            | 0.654                    | 0.652                           | 0.000     | 0.797                     | 0.688       | 0.328           |
| 13    | 0.852            | 0.857                    | 0.702    | 0.460            | 0.686                    | 0.496                           | 0.000     | 0.000                     | 0.717       | 0.704           |
| 14    | 0.748            | 0.887                    | 0.616    | 0.798            | 0.814                    | 0.730                           | 0.885     | 0.967                     | 0.854       | 0.674           |
| 15    | 1.000            | 1.000                    | 1.000    | 1.000            | 1.000                    | 1.000                           | 1.000     | 1.000                     | 0.993       | 0.952           |
| 16    | 0.838            | 0.674                    | 0.504    | 0.574            | 0.509                    | 0.282                           | 0.000     | 0.000                     | 0.856       | 0.901           |
| 17    | 0.247            | 0.756                    | 0.599    | 0.678            | 0.800                    | 0.682                           | 0.000     | 1.000                     | 0.655       | 0.608           |
| 18    | 0.836            | 0.959                    | 0.673    | 0.715            | 0.618                    | 0.695                           | −0.057    | 0.848                     | 0.851       | 0.619           |
| 19    | 0.884            | 0.913                    | 0.670    | 0.817            | 0.611                    | 0.605                           | 0.000     | 0.880                     | 0.899       | 0.692           |
| Median | 0.832          | 0.857                    | 0.670    | 0.715            | 0.750                    | 0.695                           | 0.000     | 0.848                     | 0.837       | 0.692           |
| Range | 0.000 ~ 0.522    | ~ 0.183 ~ 0.286           | 0.509    | ~ 0.057 ~ 0.282 ~ | ~ 0.057 ~ 0.000 ~ 0.559 ~           | 0.328 ~           |           |                           |             |                 |
|       | 1.000           | 1.000                    | 1.000    | 1.000            | 1.000                    | 1.000                           | 1.000     | 1.000                     | 0.993       | 0.952           |

LPR, laryngopharyngeal reflux.

19 endoscopists except for otorhinolaryngologist in the first and second tests were 0.30 and 0.26, respectively. In analysis of Fleiss’ kappa coefficient for vulnerable four parameters of RFS, kappa values of the scores for 3 parameters including subglottic edema, erythema, and vocal cord edema were increased in the second test. However, kappa values of the scores for diffuse laryngeal edema and granuloma were decreased (Table 5). The last test using the same slide of the second test without additional education was conducted 2 week after the second test. In an analysis of test-retest reliability (correlation coefficient) except for 3 endoscopists who had lowest kappa value in the 2nd test, the value of test-retest reliability coefficient was more than 0.6 in most parameters. However, the value of granuloma was lower than 0.6 (Table 6).

DISCUSSION

LPR represents one of the atypical manifestations of supranesophageal GERD characterized by morphological and functional changes in the larynx and pharynx with associated clinical symptoms.10 The most distinct changes induced by supranesophageal reflux occur in the larynx, making laryngologic examination an important tool in the diagnosis of LPR.5,11 Laryngoscopy is the first-line investigation method that is usually performed to make a diagnosis of LPR. Otorhinolaryngologists are accustomed to practice the RFS while they learn basic laryngoscopic
findings. The main research question for this study was created based on the curiosity whether endoscopists can make a diagnosis of LPR easily by means of the RFS because they always had to pass the endoscope by the larynx into the esophagus for upper endoscopy.

Some endoscopists may not be interested in laryngeal exam during upper endoscopy because they don’t have awareness of LPR significance. In reverse, LPR may stay outside of interest for endoscopists because they are unaccustomed to the convenient diagnostic tool of RFS. It should be also noted that LPR is diagnosed not only by the RFS but also by reflux symptom index or other methods by otorhinolaryngologists. It seems that laryngoscopy or endoscopy has a limited role in predicting or confirming the acid reflux pathology in the laryngeal area.5,12-15 However, considering the easy accessibility of upper endoscopic examination in Korea, endoscopists’ role for the diagnosis of LPR may have to be expanded.

To our disappointment, the result of this study showed that the interrater reliability of LPR diagnosis by using the RFS based on endoscopic laryngeal findings between an otorhinolaryngologist and an endoscopist was not so high. Individual factors, including endoscopists’ age, endoscopic experience, and time of education of RFS didn’t affect main result. Both of quality and sufficient duration are important factors in the education program. However, due to the methodological pitfalls (relatively short-term education period and pilot format study), authors could not accurately explain the influence of these factors in the overall results of this study. The Kappa value for the diagnosis of LPR had low reliability. The feedback education after the first test did not affect the Kappa value in the second test. However, in the analysis for each parameter of RFS, 4 parameters including ventricular obliteration, vocal fold edema, posterior commissure hypertrophy, and thick endolaryngeal mucus in the 2nd test had higher Kappa values than those of 1st test indicating possibility of improving diagnostic ability by more discrete education separated on each parameter. Among the vulnerable four parameters including subglottic edema, erythema, diffuse laryngeal edema, and granuloma, two parameters about subglottic edema and erythema showed improved Fleiss’s Kappa value in the second test. In the third test using the same slides of the second test, almost all parameters except for granuloma shows meaningful test-retest reliability for endoscopic diagnosis of LPR. Unfortunately, relatively a little test item about granuloma were included in our test slides and this might arise selection bias. These results show a different point of view between endoscopists and otorhinolaryngologist. The RFS to make a diagnosis of LPR which was developed by otorhinolaryngologists cannot be applied to endoscopic assessment of LPR per se for endoscopists. In addition, it would be a little bit difficult to develop an endoscopic assessment tool of LPR because the endoscopic image shows a wider range of view than laryngoscopy and doctors have different point of view between endoscopists and otorhinolaryngologists. However, considering the LPR is another manifestation of GERD, endoscopists’ role should be also emphasized.

This study has several limitations. Firstly, the basic premise is that there is no difference in the use of RFS between the endoscopic and laryngoscopic findings of the larynx. Contrary to expectations, the interrater reliability of LPR diagnosis by using the RFS based on endoscopic findings may be lower among otorhinolaryngologists. In this case, the RFS should not be applied just as it is in the endoscopic field without a separate validation study, although upper endoscopy provided clear images with higher resolution than laryngoscopy. Secondly, still images, which used for test, not video, stayed away from real clinical setting. In the study of Belašky et al.6 for validity and reliability of the RFS, each subject underwent a comprehensive history and physical examination including transnasal fiberoptic laryngoscopy with laryngeal photodocumentation. This format could be useful for the further studies. A third of all, selection bias could affect the result in the selection of still images for the test.

In conclusion, the improvement of accuracy in the diagnosis of LPR was not achieved by short-term education program for endoscopists. Further studies using advanced educational program are required for the development of diagnostic tools of LPR for endoscopists.

REFERENCES

1. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R: Global
Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101:1900-1920; quiz 1943.

2. Jaspersen D, Kulig M, Labenz J, et al. Prevalence of extra-oesophageal manifestations in gastro-oesophageal reflux disease: an analysis based on the ProGERD Study. Aliment Pharmacol Ther 2003;17:1515-1520.

3. Richter JE. Ear, nose and throat and respiratory manifestations of gastro-esophageal reflux disease: an increasing conundrum. Eur J Gastroenterol Hepatol 2004;16:857-845.

4. Richter JE. Extraoesophageal presentations of gastroesophageal reflux disease: an overview. Am J Gastroenterol 2000;95(8 Suppl):S1-S3.

5. Jonaitis L, Pribiuišienė R, Kupcinskas L, Uloza V. Laryngeal examination is superior to endoscopy in the diagnosis of the laryngopharyngeal form of gastroesophageal reflux disease. Scand J Gastroenterol 2006;41:131-137.

6. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). Laryngoscope 2001;111:1313-1317.

7. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960:20:37-46.

8. Fleiss JL, Levin B, Paik MC. The measurement of interrater agreement. Statistical methods for rates and proportions. Hoboken: John Wiley & Sons, Inc., 2004.

9. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37:360-363.

10. Shaker R, Bardan E, Gu C, Kem M, Torrico L, Toohill R. Intrapharyngeal distribution of gastric acid refluxate. Laryngoscope 2003;113:1182-1191.

11. Malagelada JR. Review article: supra-oesophageal manifestations of gastro-oesophageal reflux disease. Aliment Pharmacol Ther 2004;19 Suppl 1:45-48.

12. Poelmans J, Feenstra L, Demedts I, Rutgeerts P, Tack J. The yield of upper gastrointestinal endoscopy in patients with suspected reflux-related chronic ear, nose, and throat symptoms. Am J Gastroenterol 2004;99:1419-1426.

13. Catalano F, Terminella C, Grillo C, Biondi S, Zappalà M, Bentivegna C. Prevalence of oesophagitis in patients with persistent upper respiratory symptoms. J Laryngol Otol 2004;118:857-861.

14. Habermann W, Schmid C, Neumann K, Devaney T, Hammer HF. Reflux symptom index and reflux finding score in otolaryngologic practice. J Voice 2012;26:e126-e127.

15. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). J Voice 2002;16:274-277.