Received: 16 November 2022 | Revised: 8 February 2023 | Accepted: 19 February 2023

DOI: 10.1002/lio2.1036

ORIGINAL RESEARCH

Laryngoscope Investigative Otolaryngology

Validation and psychometric evaluation of the Swedish version of the Nasal Obstruction Symptom Evaluation scale

Ola Sunnergren PhD[1] | Amir H. Pakpour PhD[2] | Henrik Bergquist PhD[3,4] | Pernilla Sahlstrand-Johnson PhD[5] | Pär Stjärne PhD[6,7] | Anders Broström PhD[2,8,9,10]

[1]Ear, Nose & Throat Clinic, Region Jönköping County, Jönköping, Sweden

[2]Department of Nursing, School of Health and Welfare, Jönköping University, Jönköping, Sweden

[3]Department of Otorhinolaryngology, Head and Neck Surgery, Sahlgrenska University Hospital, Gothenburg, Sweden

[4]Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[5]Department of Otorhinolaryngology, Head and Neck Surgery, Skåne University Hospital, Malmö, Sweden

[6]Department of Otorhinolaryngology, Karolinska University Hospital, Stockholm, Sweden

[7]Department of Clinical Sciences, Intervention and Technology, Division of Otorhinolaryngology, Karolinska Institute, Stockholm, Sweden

[8]A.D.U.L.T., School of Health and Welfare, Jönköping University, Jönköping, Sweden

[9]Department of Clinical Neurophysiology, University Hospital Linköping, Linköping, Sweden

[10]Department of Health and Caring Sciences, Western Norway University of Applied Sciences, Bergen, Vestlandet, Norway

Correspondence
Ola Sunnergren, Ear, Nose & Throat Clinic, Region Jönköping County, Öron-, näs-och halskliniken, Länssjukhuset Ryhov, 551 85 Jönköping, Sweden.
Email: ola.sunnergren@rjl.se

Funding information
This study was in part funded by grants from Futurum- The Academy of Health and Care in Region Jönköping County, Sweden., Grant/Award Number: FUTURUM-803421; This study was in part funded by grants from the Swedish state under the ALF agreement between the Swedish government and the county councils., Grant/Award Number: ALFGBG-715771

Abstract

Objectives: The Nasal Obstruction Symptom Evaluation (NOSE) scale is a symptom-specific quality-of-life questionnaire for patients suffering from nasal obstruction. The instrument is designed specifically for patients with septal deviation and for the evaluation of the outcome of septoplasty. The aim of this study was to validate a Swedish version of the NOSE instrument for use in clinical practice and research.

Methods: A Swedish version of the NOSE was tested in a case group consisting of 125 subjects with nasal obstruction (of which 31 underwent septoplasty) and a control group consisting of 65 healthy subjects. Base line data for the case and control groups were used to evaluate face validity, known groups validity, construct validity, internal consistency and factor structure analysis. Fifty participants in both the case groups and control groups were assessed both at baseline and after 2 weeks to evaluate test–retest reliability. The participants who underwent septoplasty were assessed at baseline and after 3–6 months to evaluate responsiveness.

Results: The S-NOSE was found to be reliable, valid, and responsive. Both Cronbach's α and McDonald Omega coefficients were >0.7, and the intra class coefficient was 0.942. The S-NOSE scores were significantly correlated with nasal patency VAS in both the case group and the control group ($p < .001$ and $p = .018$, respectively). After septoplasty, the mean S-NOSE score were significantly improved ($p < .001$). Furthermore, the S-NOSE was shown to have excellent and robust psychometric properties.

Conclusion: The S-NOSE can be recommended in both clinical practice and research to evaluate the outcome of septoplasty in Swedish-speaking populations.
Level of Evidence: NA.

KEYWORDS
nasal obstruction, NOSE scale, quality of life, septoplasty, Swedish language, validation

1 | OBJECTIVES

Septoplasty of a deviated septum causing nasal obstruction is one of the most common surgical procedures performed by otorhinolaryngologists worldwide. Due to a lack of high quality studies proving the efficacy of septoplasty, and the wide variations in use across different countries, the plausibility of septoplasty as a standard treatment of the above has been under debate.[1] However, in 2019, in the so far most methodologically robust study on the outcome of septoplasty, van Egmond et al. showed that surgery is more effective than non-surgical treatment of nasal obstruction caused by a septal deviation.[2] These promising results need to be corroborated by further studies using different methodologies that reflect the current and continuously evolving clinical practice. The comparative evaluation of such studies necessitates multilingual, validated and disease-specific instruments.

A patient-reported outcome measure (PROM) collects subjective information directly from the patient regarding specific or general conditions and adds to clinical and functional outcomes.[3] One of the most commonly used instruments to evaluate the subjective outcome after septoplasty is the Nasal Obstruction Symptom Evaluation (NOSE) scale developed by Stewart et al.[4] The NOSE instrument contains five questions focusing on nasal symptoms, and the original instrument in English was found to be valid, reliable, and sensitive to change in clinical status.[4] The NOSE instrument has been translated into multiple languages, for example French, Spanish, and Arabic.[5–7] These studies have each separately contributed different psychometric evidence that NOSE works well in different clinical settings and countries, but further studies with a thorough and comprehensive psychometric evaluation of the validity and reliability of the instrument are needed to provide evidence that NOSE is a relevant PROM for use in international comparisons of septoplasty results. Therefore, there is a need for a Swedish version of the NOSE instrument for use in research, clinical practice, and health registries to measure the outcome after septoplasty in a Swedish-speaking population. The aim of this study was to translate the NOSE instrument into Swedish and accurately validate the instrument in a Swedish-speaking population.

2 | METHODS

2.1 | The NOSE instrument and its translation into Swedish

The NOSE instrument[4] contains five items: "Nasal congestion or stuffiness," "Nasal blockage or obstruction," "Trouble breathing through my nose," "Trouble sleeping," and "Unable to get enough air through my nose during exercise or exertion." The respondent is asked to grade how much of a problem each condition is on a scale from 0 to 4: 0 = Not a problem, 1 = Very mild problem, 2 = Moderate problem, 3 = Fairly bad problem, and 4 = Severe problem, giving a raw score from 0 to 20. The NOSE score is the raw score multiplied by five. A NOSE score of 0 means no problems while a score of 100 means the worst possible problems with nasal obstruction. Two separate pairs of authorized translators conducted the forward and backward translation into Swedish. The authors (Ola Sunnergren, Pär Stjärne, and Anders Broström) evaluated and merged the two versions from the forward translations into one preliminary Swedish version (S-NOSE). The S-NOSE was then translated back into English and the result was compared with the original NOSE instrument. The translation was guided by the principles of good practice for translating and culturally adapting patient-reported outcome measures,[8] and the S-NOSE was found to have both semantic and conceptual equivalence with the original NOSE instrument.

2.2 | Setting and participants

One hundred and twenty five participants with chronic nasal obstruction, referred from primary care to the Department of Otorhinolaryngology in Region Jönköping County were recruited as a case group for the study. Exclusion criteria were a clinical diagnosis of chronic rhinosinusitis, or a sino-nasal mass. Participants with seasonal allergic rhinitis were not excluded as long as they had not experienced any acute symptoms, and the timing of data collection was clearly outside the allergy season. Participants with perennial allergic rhinitis were not excluded as long as their symptoms were stable. Sixty-five healthy participants without any nasal problems, that is, a subjective perfectly functioning nasal breathing, no known sino-nasal disease, and no chronic nasal medications were recruited as a control group. The control group consisted of medical students and health care personnel. Thirty-five participants scheduled for septoplasty at one of the otorhinolaryngology departments of Region Jönköping County, Karolinska University Hospital, Sahlgrenska University Hospital, and Skåne University Hospital were recruited as a surgery group. All participants in the surgery group were also part of the case group.

2.3 | Face validity

At baseline the face validity of the S-NOSE was assessed by orally asking the case and control groups questions on whether the items in

the S-NOSE seemed relevant and accurate regarding the evaluation of nasal problems. This information was collected in face-to-face interviews either by the first author or by specially trained registered nurses at the Department of Otorhinolaryngology in Region Jönköping County.

2.4 | Known groups validity

Known groups validity, that is, the ability of the S-NOSE to distinguish between participants with and without nasal breathing problems, was assessed by comparing the S-NOSE scores of the case group and control groups at baseline (Mann–Whitney *U*-test).

2.5 | Construct validity

To assess whether the S-NOSE effectively measures the extent of nasal breathing problems the S-NOSE scores were compared with the results of a nasal patency visual analog scale (VAS). The participants were asked to grade their nasal breathing (mouth closed, using both nostrils) on a scale from 1 (total occlusion) to 10 (totally free airflow). Spearman correlation analysis of both individual item and total S-NOSE scores and nasal patency VAS scores was performed.

2.6 | Test–retest reliability

The first 65 and 50 participants in the case and control groups, respectively, were asked to complete the S-NOSE at baseline and after 2 weeks. The stability of the S-NOSE across 2 weeks was assessed using the intraclass correlation coefficient (ICC) with the two-way mixed-effect analysis of variance model with interaction for the absolute agreement between single scores. An ICC of 0.7 or greater was considered acceptable reproducibility.[9]

2.7 | Responsiveness

The 35 participants who underwent septoplasty (surgery group) were invited to complete the S-NOSE before and 6 months after surgery to assess the responsiveness of the S-NOSE. A Wilcoxon signed rank test was used to test for significant differences between the pre- and postoperative S-NOSE scores.

2.8 | Internal consistency

Cronbach's α and McDonald's Omega coefficient were computed to assess the internal consistency of the S-NOSE. Values greater than 0.70 indicate acceptable reliability. Moreover, item-total correlation corrected for overlap (criterion value: ≥0.40) was computed to measure further internal consistency of the S-NOSE.

2.9 | Factor structure analysis

The factor structure of S-NOSE was measured using confirmatory factor analysis (CFA) with full information maximum likelihood to handle missing data. Several indices were used to measure the goodness of fit of the model: Chi-square, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). An acceptable model includes CFI and TLI >0.90, RMSEA and SRMR <0.08 and a non-significant chi-square.[10] However, chi-square is sensitive to sample size and with a larger sample size, it tends to be significant. Average Variance Extracted (AVE) and Composite Reliability (CR) were computed using the CFA factor loadings. AVE >0.40 and CR >0.70 were considered acceptable.[11]

To further investigate the factor structure of the S-NOSE, the Rasch partial credit model was used. To measure item fit, the information-weighted fit statistic (infit) mean square (MnSq) and outlier-sensitive fit statistic (outfit) MnSq were used. Values between 0.5 and 1.5 were considered an acceptable fit. Moreover, the internal consistency of the S-NOSE was investigated in Rasch analysis using the item and person separation reliability >0.7 with item and person separation index >2. The unidirectionality of the S-NOSE was measured using principal component analysis (PCA) on the standardized residuals retrieved from the Rasch model. Unidimensionality was supported if the first component's eigenvalue was <2, and at least 50% of the total variance could be explained by the first component.

In addition, factorial invariance across gender for each item was assessed using the differential item functioning (DIF) in the Rasch model. The aim was to identify whether males and females interpreted any item of NOSE differently. A DIF >0.5 indicates a substantial DIF.[12]

Network analysis was conducted to understand the correlations between the items. The expected Bayesian inference criteria with the graphical least absolute shrinkage optimization (gLASSO) regularization were used to assess the network structure.[13] The network structure has two components: edge and node. Node shows variables (i.e., S-NOSE items) and edge indicates the strength of relationship between variables (or nodes) using zero-order correlations. Blue edges indicate positive correlations between nodes and a high density of the blue color in edge indicates stronger correlations between the nodes. To check how important the individual nodes were in the network (refers to centrality), four indices were used: betweenness (degree of connectivity), closeness (the distance centrality), degree (degree of centrality), and expected influence. To ensure that the network was accurate and stable, the accuracy of the edges and stability of the centrality estimates were examined using 1000 bootstrap 95% nonparametric confidence intervals (CIs).[14]

Statistical significance was defined as $p < .05$.

2.10 | Data analysis

SPSS version 27 and AMOS version 28 for Windows (IBM, Armonk, NY), JASP software version 0.14 (JASP Team, 2020) and Winstep™ software, version 4.3.0.

TABLE 1 Group characteristics and S-NOSE scores at baseline and follow-up.

	Control group ($n = 65$)	Case group ($n = 125$)	Surgery group ($n = 31$)
Age (years), Mean (±SD)	35.9 (±12.4)	33.9 (±13.0)	33.5 (±14)
Gender (male), *n* (%)	32 (49)	95 (76)	25 (81)
S-NOSE score baseline	3.5 (±5.7)	77.5 (±16.4)	79.0 (±16.1)
S-NOSE score follow-up, 2 weeks	3.5 (±5.8)[a]	72.6 (±15.9)[b]	NA
S-NOSE score follow-up, 6 months	NA	NA	26.8 (±23.7)

Note: All participants in the surgery group are also included in the case group.
Abbreviation: NA, not applicable.
[a]$n = 50$.
[b]$n = 65$.

2.11 | ETHICS STATEMENT

The study was approved by the Swedish Ethical Review Authority (2019-06015 with amendment 2021-02695). The study was conducted in accordance with the principles of the Helsinki declaration. Informed consent was obtained for all the participants in the study.

3 | RESULTS

The basal characteristics and S-NOSE scores at baseline and at follow-up of the studied populations are shown in Table 1. The main underlying causes for the chronic nasal obstruction in the case group were septal deviations $n = 102$ (81.6%), unspecified nasal obstruction $n = 17$ (13.6%), other causes $n = 6$ (4.8%). In the group that underwent surgery all participants had a septal deviation. There was no difference in mean age across the groups while the proportion of men was significantly higher in the case group compared to the control group (Fisher's exact test $p < .001$). In the group that underwent septoplasty, 31 of 35 participated in the follow-up. None of the participants in the control or case groups expressed any concern regarding the face validity of the S-NOSE items.

3.1 | Reliability

The S-NOSE was internally reliable as both Cronbach's α and the McDonald Omega coefficient were above 0.70 (Table 2).

The results of the ICCs analysis showed that all items of the S-NOSE were stable (ICCs ranged from 0.75 to 0.93). The test–retest reliability of the S-NOSE scores was also high (ICC = 0.942, 95% CI = 0.915–0.960) (Tables 2 and 3). The mean follow up time for test- retest analysis was 22 days (SD 15.2) and the median follow-up time was 16 days (range 8–117 days). Ninety-two percent of the participants completed the follow-up questionnaire ≤40 days.

TABLE 2 Psychometric properties at baseline of the S-NOSE in scale level ($n = 190$).

Psychometric testing	S-NOSE	Suggested cutoff
Internal consistency		
Cronbach's α	0.968	>0.7
McDonald Omega coefficient (ω)	0.969	>0.7
Confirmatory factor analysis		
χ^2 (*df*)	7.287 (4)	Non-significant
Comparative fit index	0.997	>0.9
Tucker–Lewis index	0.994	>0.9
Root mean square error of approximation	0.066	<0.08
Standardized root mean square residual	0.007	<0.08
Average Variance Extracted	0.854	>0.5
Composite Reliability	0.967	>0.6
Standard error of measurement	8.814	The smaller the better
Rasch partial credit model		
Item separation reliability from Rasch	0.92	>0.7
Item separation index from Rasch	3.46	>2
Person separation reliability from Rasch	0.88	>0.7
Person separation index from Rasch	2.68	>2
Test–retest reliability		
Intraclass correlation coefficient	0.942	>0.7

The item-total correlations (corrected for overlap) were all above 0.4 and ranged from 0.87 for item 4, "Trouble sleeping", to 0.95 for item 2, "Nasal blockage or obstruction" (Table 3).

All items of the S-NOSE significantly correlated with each other for the case and surgery groups (Table 4). In the control group, nine out of 15 interitem correlations were statistically significant (Table 4). Notably, four of these six non-significant correlations were related to

TABLE 3 Psychometric properties at baseline of the S-NOSE in item level ($n = 190$).

Item	Factor loading[a]	Item-total correlation	Test-retest[b]	Infit MnSq	Outfit MnSq	Difficulty	Discrimination	DIF contrast across genders[c,d]
S-NOSE1	0.90	0.90	0.902	1.10	1.22	−0.04	0.82	−0.55
S-NOSE2	0.95	0.95	0.925	0.66	0.67	−0.42	1.33	−0.13
S-NOSE3	0.97	0.94	0.934	0.68	0.70	−0.30	1.30	0.09
S-NOSE4	0.89	0.87	0.754	1.02	0.89	1.03	0.99	0.29
S-NOSE5	0.91	0.89	0.870	1.46	1.35	−0.28	0.63	0.26

Abbreviation: MnSq, mean square error.
[a]Confirmatory factor analysis, 190 participants (preoperative case and control groups).
[b]Intraclass correlation coefficient.
[c]DIF (differential item functioning) contrast >0.5 indicates substantial DIF.
[d]DIF contrast across gender categories = Difficulty for female participants - Difficulty for male participants.

TABLE 4 Spearman inter-item and item-total correlations among S-NOSE items.

Item	S-NOSE1	S-NOSE2	S-NOSE3	S-NOSE4	S-NOSE5	Corrected total
Case group, baseline ($n = 125$)						
S-NOSE1	–	0.58*	0.45*	0.47*	0.29*	0.71*
S-NOSE2		–	0.50*	0.50*	0.43*	0.77*
S-NOSE3			–	0.61*	0.48*	0.78*
S-NOSE4				–	0.35*	0.80*
S-NOSE5					–	0.66*
Corrected total						–
Control group, baseline ($n = 65$)						
S-NOSE1	–	0.42*	0.42*	−0.07	0.03	0.76*
S-NOSE2		–	0.36*	0.31*	0.02	0.63*
S-NOSE3			–	0.31*	0.16	0.64*
S-NOSE4				–	−0.04	0.23
S-NOSE5					–	0.46*
Corrected total						–
Surgery group, follow-up ($n = 31$)						
S-NOSE1	–	0.70*	0.75*	0.63*	0.60*	0.81*
S-NOSE2		–	0.75*	0.81*	0.68*	0.86*
S-NOSE3			–	0.72*	0.73*	0.90*
S-NOSE4				–	0.79*	0.89*
S-NOSE5					–	0.86*
Corrected total						–

*p-Value <.01.

item 5, "Unable to get enough air through my nose during exercise or exertion."

3.2 | Validity

There was a statistically significant difference in baseline S-NOSE scores between the case- and control groups (Mann–Whitney $p < .001$).

The results of the CFA showed that the unidimensional model was not fitted well to the data [χ^2 (df) = 4.721,[5] CFI = 0.985; TLI = 0.971; RMSEA = 0.140; and SRMR = 0.0116]. Based on modification indices of AMOS output, an error covariance between items 1 and 2 was added. Acceptable model fit was obtained based on this constraint [χ^2 (df) = 7.287,[4] CFI = 0.997; TLI = 0.994; RMSEA = 0.066; and SRMR = 0.007]. Moreover, all factor loadings were significant and ranged from 0.89 to 0.95.

Both AVE and CR were higher than the cutoff values for the NOSE (AVE = 0.85, and CR = 0.97) (Table 2).

The results of the Rasch analysis for the S-NOSE are shown in Table 2. The results of the PCA showed that the S-NOSE was unidimensional as the first contrast of the unexplained variance was 1.55.

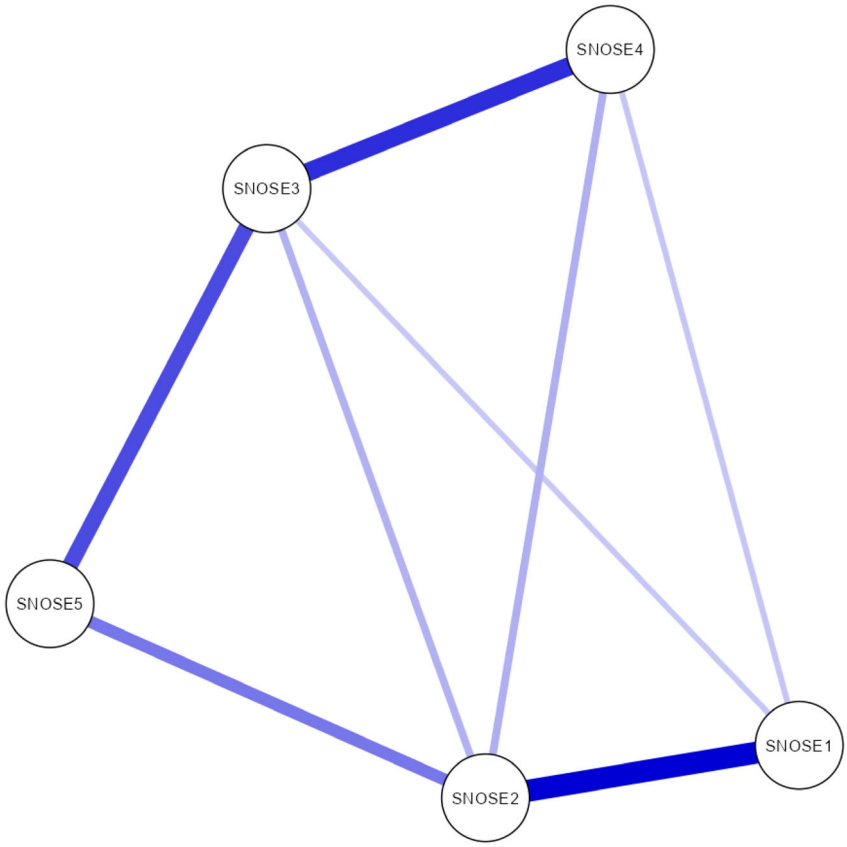

FIGURE 1 Expected Bayesian inference criteria gLASSO model based on network analysis according to the S-NOSE ($n = 190$).

Note: S-NOSE 1 - S-NOSE 5 = S-NOSE items/criteria

Moreover, the explained variance by the items was 76.8%. The most difficult item was item 4, "Trouble sleeping", while item 2, "Nasal blockage or obstruction", was the easiest item. All items had acceptable fit to the Rasch model (infit MnSq ranged from 0.68 to 1.46; Outfit MnSq ranged from 0.67 to 1.35).

DIF contrasts across gender are presented in Table 3. No items displayed substantial DIF across gender (DIF <0.50) except for item 1 with marginal DIF (DIF = -0.55).

The network model of the S-NOSE is depicted in Figure 1. Nodes S-NOSE1 (i.e., "Nasal congestion or stuffiness") and S-NOSE2 (i.e., "Nasal blockage or obstruction") had the strongest edge intensity ($r = 0.574$). Betweenness, closeness, degree, and expected influence are shown in Figure 2. The central-stability-coefficients of the S-NOSE items were 1.20, 1.20, 0.830, 0.780, and 0.710, respectively (Figures 3 and 4). The node centrality was stable and interpretable in the network.

All S-NOSE items and the total score of S-NOSE in the case and surgery groups were significantly correlated with nasal patency VAS, while in the case group only item 1 ("Nasal congestion or stuffiness") and total S-NOSE score were significantly correlated (Table 5).

3.3 | Responsiveness

Patients reported significantly lower scores for the S-NOSE 6 months after surgery (Wilcoxon rank $p < .001$) compared to baseline. The average time from baseline to follow-up after surgery was 163 days. Only one of the 31 patients that underwent surgery reported a higher S-NOSE score after surgery.

4 | DISCUSSION

All therapies in modern day medicine need to be continually scrutinized to ensure their efficacy and usefulness. In the case of septoplasty the most important outcome measure is symptom improvement after surgery. One of the strengths of the NOSE instrument is the availability in multiple languages. In this study, the Swedish version of the NOSE instrument was meticulously evaluated and validated with classical test theory (CTT) and item response theory (IRT). An important aspect of clinical practice is to understand the effectiveness of a treatment for the individual patient and not just the average improvement for the current patient group as a whole. An assessment of individual clinical changes can be made using either the CTT or the IRT.[15] The most important difference between CTT and IRT is that in a study using CTT a common estimate of measurement precision is made, which is assumed to be the same for all individuals, regardless of their characteristics and conditions. In IRT, however, the measurement precision depends on the latent value. However, it is notable that CTT and IRT each have their advantages and disadvantages.[16]

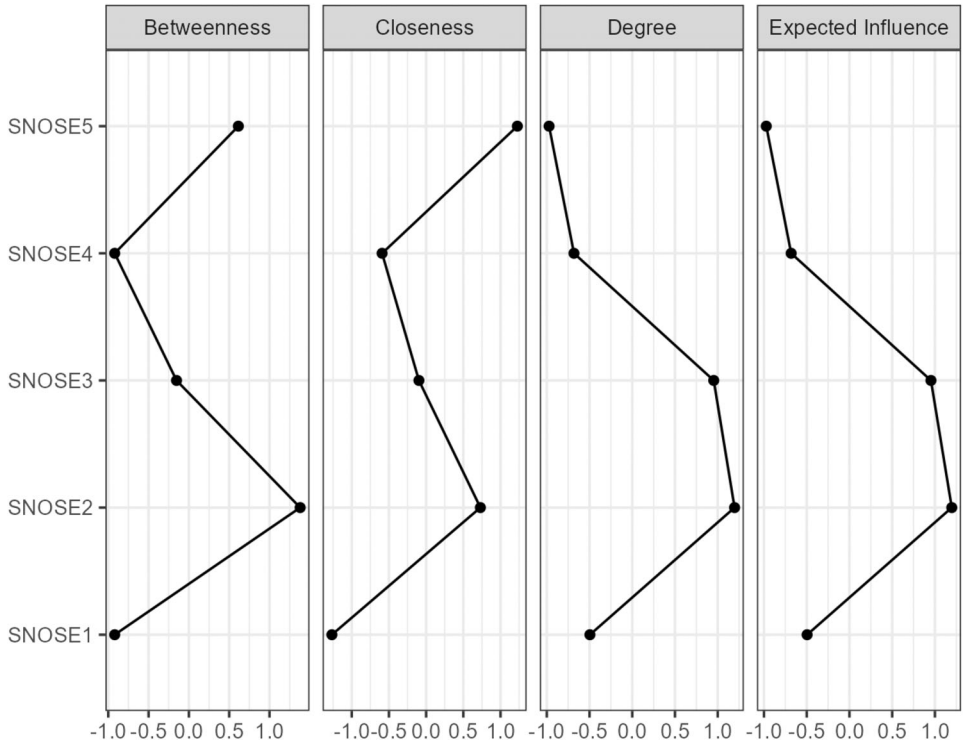

FIGURE 2 Standardized estimates of node centrality in the network ($n = 190$).

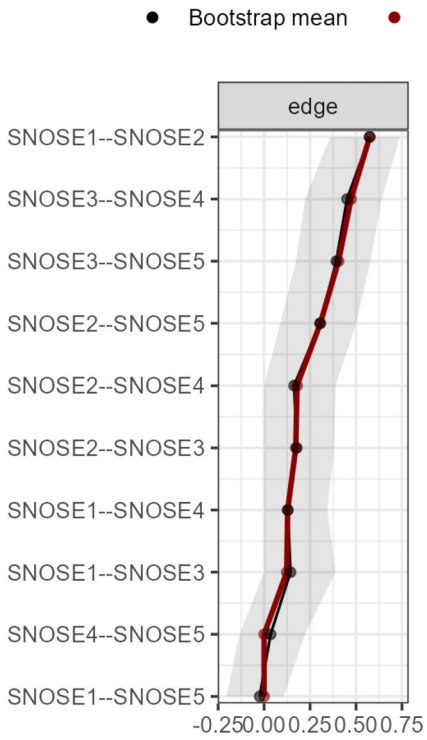

FIGURE 3 Edge stability in the network ($n = 190$).

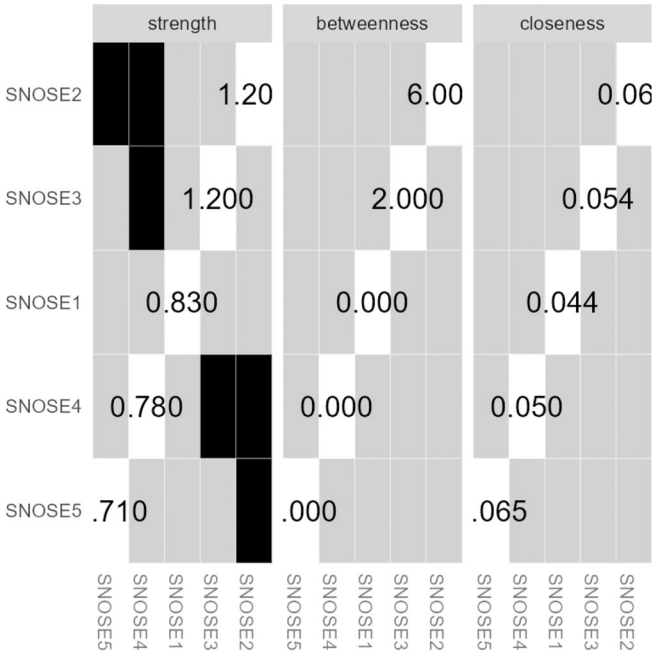

FIGURE 4 Centrality stability in the network ($n = 190$).

Therefore, both types of psychometric testing including factorial invariance across gender (i.e., to determine whether males and females interpret any item differently) were used to provide a comprehensive and modern procedure to examine the psychometric properties of NOSE. We found the S-NOSE to be reliable, valid, and responsive. The S-NOSE also had excellent and robust psychometric properties, thus showing another strength of the NOSE instrument. At construction and at every translation, including ours, the instrument has been scientifically tested with good results. This means that the NOSE instrument is one of a few scientifically robust instruments available to rhinologists to make international comparisons of outcome after septoplasty in research and clinical practice.

TABLE 5 Spearman correlation between S-NOSE and nasal patency VAS scores in different groups.

	Case group ($n = 125$)		Surgery group ($n = 31$)		Control group ($n = 65$)	
Item	ρ	p	ρ	p	ρ	p
S-NOSE1	0.40	<.001	0.63	<.001	0.39	.002
S-NOSE2	0.46	<.001	0.54	<.001	0.13	.307
S-NOSE3	0.48	<.001	0.54	<.001	0.24	.059
S-NOSE4	0.39	<.001	0.55	<.001	0.12	.364
S-NOSE5	0.34	<.001	0.61	<.001	0.02	.858
Total Score	0.54	<.001	0.61	<.001	0.30	.018

The unidimensional structure of the NOSE has been previously reported in the Dutch version.[17] The results of our CFA, Rasch and network analysis showed that S-NOSE had a unidimensional structure. Moreover, the results of the study indicated that all S-NOSE items were interpreted by male and female patients in asimilar way. As none of the S-NOSE items displayed any DIF across gender, with the exception of item 1 with mild DIF, the present study corroborates the findings of current literature that the S-NOSE is appropriate across gender.

The significant correlations between S-NOSE scores and nasal patency VAS scores showed construct validity, and the statistically significant differences in S-NOSE scores at baseline between the case and control groups showed known groups validity. These findings are in accordance with the previous French, Spanish, and Dutch NOSE validation studies.[5,6,17] It is noteworthy that the case group in our study had higher, and the control group had lower mean S-NOSE scores at baseline (77.5/3.5 SD) compared with the findings in the French, Spanish, and Dutch validation studies (70.5/8.5 SD, 58.7/15.7 SD and 63.4/6.9 SD, respectively).[5,6,17]

It was also important to ensure that the five items of the S-NOSE measured the same underlying concept of problems breathing with the nose (internal consistency). The Cronbach's α (0.968), the McDonald Omega coefficient (0.969), and the item-total correlations (all above 0.87) clearly demonstrated the internal consistency of the S-NOSE. The Cronbach's α of 0.968 in this study was higher than the corresponding Cronbach's α in the original English study (0.785), as well as in the French (0.86), Spanish (0.955), and the Dutch (0.79), validation studies.[5,6,17]

An instrument that aims to measure symptom improvement of a chronic condition after a medical intervention must be stable (test–retest reliability) if no intervention is made and responsive after a successful intervention (responsiveness). The ICC value of 0.942 showed that the test–retest reliability was excellent, and the significant difference between pre-, and post-surgery mean S-NOSE scores (79.0 and 26.8, respectively) ensures responsiveness. Interestingly, the NOSE scores after surgery in our study were comparable with the results from the original NOSE study where the baseline mean NOSE score was 67.5 and the mean score after surgery was 26.6.[18] However, the mean post-surgery NOSE score was somewhat lower in both the Spanish (14.0) and the Dutch (20.0) studies than in our study.[6,17] It can be noted that the results of surgery in the present study, as well as the other validation studies are comparable with the results from the RCT published by van Egmond et al. In their study, the preoperative mean NOSE score was 67.2 while the mean postoperative score was 32.5.[2]

As mentioned, there are several previous studies where the NOSE instrument has been translated into different languages and validated. These studies complement each other and have separately contributed psychometric evidence that the NOSE instrument is suitable for evaluating the outcome of septoplasty in the respective populations. Our study adds to the existing knowledge a thorough and comprehensive psychometric evaluation of the validity and reliability of the instrument where we use two types of psychometric testing (i.e., CCT and IRT), and also shows factorial invariance across gender (i.e., males and females do not interpret items differently). By this, we argue that this study is the most comprehensive and modern procedure examination of the psychometric properties of the NOSE instrument developed so far.

4.1 | Limitations

There are some limitations to this study. The proportion of men was larger in both the case group (76%) and the surgery group (81%) compared to the control group (49%). However, these rates are comparable to the actual rate of 74% men in Swedish septoplasties performed in the period 2012–2021.[19] As the analysis did not show any substantial gender-related differences for any of the items, the S-NOSE can be used without regard to gender in clinical practice and research.

It can be argued that the evaluation of construct validity should include objective measurements such as rhinomanometry or acoustic rhinometry. We chose not to include these since both methods have weak correlations with subjective nasal obstruction[20] and because the subjective experience of the patient is more important than numerical values of an objective measurement when it comes to the effect of septal surgery. Another alternative would be to use peak nasal inspiratory flow, but as this method is dependent on age, gender and height,[20] we found its use unsuitable given the size and composition of the sample size in this study. Instead, we chose the same strategy as van Zijl et al. in the validation of the Dutch version of the NOSE instrument,[17] namely nasal patency VAS scores. In the Dutch study, the correlations between NOSE scores and nasal patency VAS scores were significant at baseline in both the control and case groups as well as at follow-up in the group that underwent surgery. In contrast, we

found correlations at baseline and at follow up in the case and surgery group respectively but not at base line in the control group. This could be explained by the distribution of S-NOSE and VAS scores in the control group with high VAS scores and low S-NOSE scores.

All data in the case and control groups were collected at one hospital, which may cause a selection bias. However, a comparison of the mean baseline S-NOSE scores of the control group (77.5) with the surgery group (79.0), recruited at four different hospitals, indicated that the case group was representative. The reason we recruited the surgery group in four hospitals was that very few septoplasties were performed during the COVID-19 pandemic. Even with four hospitals involved we had difficulty recruiting septoplasty patients within a timeframe compatible with the other data collection in this study.

The intention was to collect data for the test–retest analysis 2 weeks after baseline. It proved difficult to get the participants to answer the questionnaire in due time. Many participants had to be reminded (some of them several times) while some participants even returned the questionnaire before 2 weeks had passed. This was not found to negatively affect the results as the test–retest reliability of the S-NOSE scores was excellent (ICC = 0.942).

5 | CONCLUSION

The present study shows that the S-NOSE is reliable, valid, responsive, and psychometrically sound. A major strength of the NOSE instrument is that it is available in multiple languages and now also in Swedish. We recommend researchers and clinicians use the S-NOSE when exploring the outcome of septoplasty in Swedish-speaking populations. Our recommendation applies not only to research but also to clinical practice.

FUNDING INFORMATION

This study was in part funded by grants from Futurum- The Academy of Health and Care in Region Jönköping County, Sweden (FUTURUM-803421), and by grants from the Swedish state under the ALF agreement between the Swedish government and the county councils (ALFGBG-715771).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Ola Sunnergren https://orcid.org/0000-0002-1192-0182

REFERENCES

1. Sommer F, Hoffmann TK. Septoplasty-a surgical or political challenge? *Lancet*. 2019;394(10195):276-278.
2. van Egmond MMHT, Rovers MM, Hannink G, Hendriks CTM, van Heerbeek N. Septoplasty with or without concurrent turbinate surgery versus non-surgical management for nasal obstruction in adults with a deviated septum: a pragmatic, randomised controlled trial. *Lancet*. 2019;394(10195):314-321.
3. Krogsgaard MR, Brodersen J, Christensen KB, et al. What is a PROM and why do we need it? *Scand J Med Sci Sports*. 2021;31(5):967-971.
4. Stewart MG, Witsell DL, Smith TL, Weaver EM, Yueh B, Hannley MT. Development and validation of the nasal obstruction symptom evaluation (NOSE) scale. *Otolaryngol Head Neck Surg*. 2004;130(2):157-163.
5. Marro M, Mondina M, Stoll D, de Gabory L. French validation of the NOSE and RhinoQOL questionnaires in the management of nasal obstruction. *Otolaryngol Head Neck Surg*. 2011;144(6):988-993.
6. Larrosa F, Roura J, Dura MJ, Guirao M, Alberti A, Alobid I. Adaptation and validation of the Spanish version of the nasal obstruction symptom evaluation (NOSE) scale. *Rhinology*. 2015;53(2):176-180.
7. Amer MA, Kabbash IA, Younes A, Elzayat S, Tomoum MO. Validation and cross-cultural adaptation of the arabic version of the nasal obstruction symptom evaluation scale. *Laryngoscope*. 2017;127(11):2455-2459.
8. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health*. 2005;8(2):94-104.
9. Yen M, Lo LH. Examining test-retest reliability: an intra-class correlation approach. *Nurs Res*. 2002;51(1):59-62.
10. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999;6:1-55.
11. Fornell C, Larcker DF. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J Market Res*. 1981;18(1):39-50.
12. Mok MM. Validation of scores from self-learning scales for primary students using true-score and Rasch measurement methods. *J Appl Meas*. 2004;5(3):258-286.
13. Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*. 2008;95(3):759-771.
14. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: network visualizations of relationships in psychometric data. *J Stat Softw*. 2012;48(4):1-18.
15. Jabrayilov EWHM, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. *Appl Psychol Measur*. 2016;40(8):559-572.
16. Prieler JA. So wrong for so long: changing our approach to change. *The Psychologist*. 2007;20:730-732.
17. van Zijl FVWJ, Timman R, Datema FR. Adaptation and validation of the Dutch version of the nasal obstruction symptom evaluation (NOSE) scale. *Eur Arch Otorhinolaryngol*. 2017;274(6):2469-2476.
18. Stewart MG, Smith TL, Weaver EM, et al. Outcomes after nasal septoplasty: results from the nasal obstruction septoplasty effectiveness (NOSE) study. *Otolaryngol Head Neck Surg*. 2004;130(3):283-290.
19. https://www.socialstyrelsen.se/statistik-och-data/statistik/statistikdatabasen/ (In Swedish) (assessed: October 2022).
20. Rimmer J, Hellings P, Lund VJ, et al. European position paper on diagnostic tools in rhinology. *Rhinology*. 2019;57(Suppl S28):1-41.

How to cite this article: Sunnergren O, Pakpour AH, Bergquist H, Sahlstrand-Johnson P, Stjärne P, Broström A. Validation and psychometric evaluation of the Swedish version of the Nasal Obstruction Symptom Evaluation scale. *Laryngoscope Investigative Otolaryngology*. 2023;8(2):357-366. doi:10.1002/lio2.1036

APPENDIX A

A.1 | SVENSK VERSION

A.1.1. | Nasal Obstruction Symptom Evaluation scale

Hur stora *problem* har du haft under *den senaste månaden* med följande? *Ringa in* det svar som stämmer bäst.

	Inga problem	Lindriga problem	Måttliga problem	Ganska stora problem	Stora problem
1. Nästäppa	0	1	2	3	4
2. Trång eller blockerad näsa	0	1	2	3	4
3. Svårt att andas genom näsan	0	1	2	3	4
4. Svårt att sova (pga. näsbesvär)	0	1	2	3	4
5. Svårt att få tillräckligt med luft genom näsan vid träning eller ansträngning	0	1	2	3	4

Summan multipliceras med 5 för att få NOSE score (0–100) _____.