ENT updates
www.entupdates.org

Clinical Research

ENT Updates 2019;9(2): 138-143
DOI: 10.32448/entupdates.596457

# Intra-rater and Inter-rater consistency of drug induced sleep endoscopy

Ahmet Erdem Kılavuz, MD[1] - Ali Alper Bayram MD[2]

*[1] Department of Otorhinolaryngology, Acibadem Healthcare Group, Maslak Hospital, Acibadem University, Istanbul, Turkey*
*ORCID ID: 0000-0001-6001-7697*
*2 Department of Otorhinolaryngology, Bahcelievler State Hospital, Istanbul, Turkey*
*ORCID ID: 0000-0002-2714-6398*

## Abstract

**Objective:** Drug induced sleep endoscopy (DISE) is a valuable tool which is used in the diagnosis of obstructive sleep apnea (OUA). The aim of this study is to evaluate inter-rater and intra-rater consistency of DISE.

**Methods:** 36 OSA patients with Apnea-hypopne index>5 included in this study. DISE was performed and recorded digitally for all patients, by the first author (OA1). VOTE scores were noted to procedure report in patients' charts. Video records of DISE were blindly evaluated six months after the last procedure, by observer 1 for the second time (OA2) and by observer 2 (OB) for the first time. DISE was evaluated by using VOTE classification. OA1 and OA2 scores were compared to determine intra-rater reliability and OA2 and OB scores were compared to determine inter-rater reliability.

**Results:** Inter-rater consistency of DISE was poor to good. Highest consistency rate was found in velum at anteroposterior configuration, while the lowest was found in the same level at lateral configuration. Intra-rater consistency of DISE was moderate to excellent. Highest consistency rate was found in epiglottis at lateral configuration, while the lowest was found in oropharynx level.

**Conclusion:** OSA is condition with possible serious complications. DISE is a tool that could change the course of treatment in OSA. The validity of DISE is quite acceptable although a golden standard classification tool could enable us to "speak the exact same language" and will surely increase the diagnostic success of DISE.

**Keywords:** Obstructive sleep apnea, drug-Induced abnormalities, endoscopy.

## Introduction

Obstructive sleep apnea (OSA) is a syndrome that consists of apnea and hypoapnea period, lower blood oxygen levels and arousals during sleep. This syndrome is known to cause fatigue and excessive daytime somnolence and could also lead to metabolic, cardiovascular and pulmonary diseases.[1]

Continuous positive airway pressure (CPAP) therapy is considered gold standard treatment for especially in moderate and severe OSA, however low compliance to this therapy leads to failure in many patients.[2] Therefore, in cases of CPAP failure and in patients with mild or moderate OSA alternative treatments such as surgery and oral appliance therapy (OAT) could be considered.[3,4]

**Correspondence:** Ahmet Erdem Kilavuz, MD
Dept. of Otorhinolaryngology, Acibadem Healthcare Group, Maslak Hospital, Acibadem University, Istanbul, Turkey
e-mail: aekilavuz@gmail.com

**Received:** 09.05.2019; **Accepted:** 22.07.2019

Online available at:
www.entupdates.org

*©2019 Continuous Education and Scientific Research Association (CESRA)*

There are numerous surgical techniques for different levels of upper airway (UA) collapse, however selecting the suitable surgery for each individual patient is a challenge in sleep surgery. The limitations of awake examination, which lacks the events that occur during sleep, negatively affect the surgical success.[5]

Drug induced sleep endoscopy (DISE), which is first described by Croft and Pringle, allows a sleep surgeon to evaluate entire UA after pharmacological induction of sleep. DISE provides valuable data regarding UA collapse level and pattern and helps selecting suitable surgical treatment for each patient.[6] Studies have demonstrated its usefulness in increasing surgical success rate, by selecting the right surgery for the right patient.[7,8] Besides its advantages, there are limitations of DISE such as its subjective nature in evaluating UA collapse. Although different classification and scoring systems were described in the literature to overcome this challenge, no standard or universally adopted approach toward classification is available yet.[5,9]

Some initial studies have demonstrated inter-observer and test-retest reliability of DISE, however there is a need of additional reports to challenge or confirm those findings and intra-rater reliability has not been assessed yet.[10-13] Therefore, this study was conducted to assess differences in DISE evaluation by different observers and compare our results with the previous studies. Secondary aim of this study was to assess intra-rater consistency of DISE, by comparing the assessment of the same surgeon in two different occasions.

## Methods

This retrospective study was approved by the local ethics committee. In total, 36 patients with an apnea–hypopnea index (AHI)>5, as determined by a previously conducted polysomnography (PSG), were included in this study. Demographic data, body mass index (BMI) and apnea-hypopnea index (AHI) were evaluated from patients' charts.

### Drug-induced sedation endoscopy

DISE was performed on all patients in a dark and silent operating room with each patient in a supine position. Sedation was achieved by intravenous administration of midazolam (bolus injection of 0.05 mg/kg) and propofol (loading dosage of 1 mg/kg and additional dosage of 20 mg in every 2 minutes). Sedation depth was confirmed with a Bispectral Index (BIS) monitor and DISE was performed under BIS levels between 50 and 70. No anticholinergic or topical anesthesia of the nose was used during the procedure. Oxygen saturation and cardiac rhythms were monitored by an anesthetic team during the procedure.

Upper airway sites containing; velum, oropharyngeal lateral walls, tongue base and epiglottis were evaluated with a flexible fiber-optic laryngoscope and findings were recorded digitally. VOTE classification was used to assess collapse pattern (lateral, anteroposterior, concentric) and collapse degree (0- no collapse or minimal vibration, 1- partial collapse, 2- total collapse).[14] DISE was performed and recorded digitally for all patients, by the first author (OA1). VOTE scores were noted to procedure report in patients' charts.

Video records of DISE were blindly evaluated six months after the last procedure, by observer 1 for the second time (OA2) and by observer 2 (OB) for the first time. OA1 and OA2 scores were compared to determine intra-rater reliability. OA2 and OB scores were compared to determine inter-rater reliability.

### Statistical Analysis

The Statistical Package for the Social Sciences for Windows version 22.0 (SPSS Inc., Chicago, IL, USA) was used to conduct the statistical tests. Descriptive statistics were expressed as mean ± standard deviation (SD), median and range, frequency, or rate. The Fleiss kappa and Cohen kappa statistical tests were used for the inter-rater and intra-rater consistency.

Results were accepted as either poor (kappa < 0.20), fair (kappa = 0.21–0.40), moderate (kappa = 0.41–0.60), good (kappa = 0.61–0.80), very good (kappa = 0.81–0.90), and excellent (kappa > 0.91). Results with a p value <0.05 were considered statistically significant.

## Results

The ages of all patients were ranged from 24 to 70 with a mean of 45.2 ± 10.8. The mean BMI was 27.5 ± 3.1 kg/m2 (ranged from 21.8 to 35.5 kg/m2). The mean AHI was 25.2 ± 11.5 (ranged from 12 to 76) **(Table 1).**

### Velum

Intra-rater consistency for velum related collapse in anteroposterior configuration was very good (consistency: 91.7%, Kappa: 0.835, p<0.05), while Inter-rater consistency for velum related collapse in anteroposterior configuration was good (consistency: 80.6%, Kappa: 0.647, p<0.05).

**Table 1.** Age, gender, BMI and AHI of the study group.

| | | Min-Max | | | Median | Mean sd./n-% | | |
|---|---|---|---|---|---|---|---|---|
| Age | | 24.0 | - | 70.0 | 45.0 | 45.2 | ± | 10.8 |
| Gender | F | | | | | 4 | | 11.1% |
| | M | | | | | 32 | | 88.9% |
| BMI | | 21.8 | - | 35.5 | 27.7 | 27.5 | ± | 3.1 |
| AHI | | 12.0 | - | 76.0 | 24.0 | 25.2 | ± | 11.5 |

Intra-rater consistency for velum related collapse in concentric configuration was good (consistency: 86.1%, Kappa: 0.667, p<0.05), while Inter-rater consistency for velum related collapse in concentric configuration was fair but statistically not significant (consistency: 63.9%, Kappa: 0.257, p>0.05).

Intra-rater consistency for velum related collapse in lateral configuration was good (consistency: 97.2%, Kappa: 0.654, p<0.05), while Inter-rater consistency for velum related collapse in lateral configuration was poor and statistically not significant (consistency: 91.7%, Kappa: -0.038, p>0.05) **(Table 2).**

## Oropharynx

Intra-rater consistency for oropharynx related collapse in lateral configuration was moderate (consistency: 63.9%, Kappa: 0.451, p<0.05), while Inter-rater consistency for oropharynx related collapse in lateral configuration was also moderate (consistency: 66.7%, Kappa: 0.497, p>0.05) **(Table 3).**

## Tongue base

Intra-rater consistency for tongue related collapse in anteroposterior configuration was moderate (consistency: 63.9%, Kappa: 0.464, p<0.05), while Inter-rater consistency for tongue related collapse in anteroposterior configuration was also moderate (consistency: 63.9%, Kappa: 0.419, p>0.05) **(Table 4).**

## Epiglottis

Intra-rater consistency for epiglottis related collapse in anteroposterior configuration was moderate (consistency: 83.3%, Kappa: 0.599, p<0.05), while Inter-rater consistency for epiglottis related collapse in anteroposterior configuration was also moderate (consistency: 80.6%, Kappa: 0.551, p<0.05).

**Table 2.** Inter-rater and intra-rater consistency of the collapse in the velum.

| | Degree of Collapse | | | | Consistency | | Kappa | p |
|---|---|---|---|---|---|---|---|---|
| | 0 | I | II | | n | % | | |
| ***Velum- Anteroposterior*** | | | | | | | | |
| Observer A1 | 13 | 1 | 22 | Observer A1-A2 | 33 | 91.7% | 0.835 | 0.000 |
| Observer A2 | 12 | 2 | 22 | | | | | |
| Observer B | 17 | 2 | 17 | Observer A2-B | 29 | 80.6% | 0.647 | 0.000 |
| ***Velum- Concentric*** | | | | | | | | |
| Observer A1 | 26 | 2 | 8 | Observer A1-A2 | 31 | 86.1% | 0.667 | 0.000 |
| Observer A2 | 26 | 0 | 10 | | | | | |
| Observer B | 21 | 3 | 12 | Observer A2-B | 23 | 63.9% | 0.257 | 0.070 |
| ***Velum- Lateral*** | | | | | | | | |
| Observer A1 | 34 | 0 | 2 | Observer A1-A2 | 35 | 97.2% | 0.654 | 0.000 |
| Observer A2 | 35 | 0 | 1 | | | | | |
| Observer B | 34 | 0 | 2 | Observer A2-B | 33 | 91.7% | -0.038 | 0.806 |

Kappa Consistency Test

**Table 3.** Inter-rater and intra-rater consistency of the collapse in the oropharynx.

| | Degree of Collapse | | | | Consistency | | Kappa | p |
|---|---|---|---|---|---|---|---|---|
| | 0 | I | II | | n | % | | |
| ***Oropharynx*** | | | | | | | | |
| Observer A1 | 11 | 12 | 13 | Observer A1-A2 | 23 | 63.9% | 0.451 | 0.000 |
| Observer A2 | 7 | 10 | 19 | | | | | |
| Observer B | 4 | 22 | 10 | Observer A2-B | 24 | 66.7% | 0.497 | 0.000 |
| Kappa Consistency Test | | | | | | | | |

**Table 4.** Inter-rater and intra-rater consistency of the collapse in the tongue.

| | Degree of Collapse | | | | Consistency | | Kappa | p |
|---|---|---|---|---|---|---|---|---|
| | 0 | I | II | | n | % | | |
| ***Oropharynx*** | | | | | | | | |
| Observer A1 | 15 | 13 | 8 | Observer A1-A2 | 23 | 63.9% | 0.464 | 0.000 |
| Observer A2 | 10 | 13 | 13 | | | | | |
| Observer B | 13 | 17 | 6 | Observer A2-B | 23 | 63.9% | 0.419 | 0.000 |
| Kappa Consistency Test | | | | | | | | |

Intra-rater consistency for epiglottis related collapse in lateral configuration was excellent (consistency: 100%, Kappa: 1.000, p<0.05), while Inter-rater consistency for epiglottis related collapse in lateral configuration poor and statistically not significant (consistency: 88.9%, Kappa: 0.158, p>0.05) **(Table 5).**

**Table 5.** Inter-rater and intra-rater consistency of the collapse in the epiglottis.

| | Degree of Collapse | | | | Consistency | | Kappa | p |
|---|---|---|---|---|---|---|---|---|
| | 0 | I | II | | n | % | | |
| ***Epiglottis- Anteroposterior*** | | | | | | | | |
| Observer A1 | 28 | 2 | 6 | Observer A1-A2 | 30 | 83.3 | 0.599 | 0.000 |
| Observer A2 | 25 | 2 | 9 | | | | | |
| Observer B | 27 | 3 | 6 | Observer A2-B | 29 | 80.6% | 0.551 | 0.000 |
| ***Epiglottis- Lateral*** | | | | | | | | |
| Observer A1 | 34 | 1 | 1 | Observer A1-A2 | 36 | 100% | 1.000 | 0.000 |
| Observer A2 | 34 | 1 | 1 | | | | | |
| Observer B | 33 | 3 | 0 | Observer A2-B | 32 | 88.9% | 0.158 | 0.213 |
| Kappa Consistency Test | | | | | | | | |

## Discussion

DISE could create an effect that may lead to a change in the treatment plan of OSA patients that makes DISE a valuable tool in patient selection and treatment decision process. VOTE classification  is one of the most used scoring system in assessment of DISE due to its practicality and simplicity.[14] However, a classification that could be accepted as golden standard for assessment of DISE is yet to be described. A meta-analysis in the literature revealed there are more than 15 classification systems currently in use for the assessment of DISE.[15] Furthermore the possibility of developing an ideal DISE scoring system is becoming less likely since new DISE classification tools are being introduced every year.[16] In the present study, we used VOTE classification to assess our DISE findings and aimed  to evaluate inter-rater and intra-rater consistency of DISE. In our study, inter-rater consistency in the velum level was poor to good with anteroposterior configuration having the highest consistency. Oropharynx and tongue base levels had both moderate inter-rater consistency, while the result in the epiglottis level was poor to moderate. Anteroposterior configuration had higher inter-rater consistency in the epiglottis level.

Intra-rater consistency rates were generally higher in our study. Intra rater consistency in the velum level was good to very good with anteroposterior configuration, again having the highest consistency. Oropharynx and tongue base levels had both moderate intra-rater consistency, while the result in the epiglottis level was moderate to excellent. Lateral configuration had higher intra-rater consistency in the epiglottis level.

Different studies evaluated DISE consistency. Even though they had slightly different study designs the results are mostly comparable. Altintas et al [13] included 55 patients into their study and compared consistency of three different observers in their study. They reported a general inter-rater consistency as poor to good, which was quite similar to our overall inter-rater consistency (also poor to good). Breakdown of inter-rater consistency in different levels were also consistent with our study as they reported inter-rater consistency in the velum, oropharynx, tongue base and epiglottis levels as poor to good, poor to fair, fair to moderate and fair to moderate, respectively.

Carrasco-Llatas et al [12] on the other hand, reported slightly higher inter-rater consistency as their general inter-rater consistency was moderate to good. They reported the highest inter-rater consistency at the oropharynx lev el. This area is followed by the soft palate, tongue base, and finally the epiglottis. On the contrary, in our study we found the highest inter-rater consistency in the velum level at AP configuration, followed by epiglottis- AP configuration, oropharynx, tongue base, velum-concentric configuration, epiglottis- lateral configuration and velum- lateral configuration. Even though the overall consistency rates were similar, the difference of the results in the specific upper airway levels could be explained by the experience of the observers, as they chose an experienced observer and a resident in training while both observers in our study had the same level of experience.

Kezirian et al [10] in an early study, reported an overall high consistency of two different observers. However this study took place before Kezirian proposed VOTE classification [14] and contained an even simpler classification: palate and hypopharynx. They reported higher consistency rates when it comes down to the presence of collapse itself rather than the degree of collapse. Moreover, they reported higher consistency rates of the structures contributing to obstruction at the hypopharynx level.

Vroegop et al [11] have also assessed intra-rater consistency alongside inter-rater consistency in both experienced and non-experienced group of observers and concluded both inter-rater and intra-rater consistency was higher in experienced versus non-experienced observers. The results of experienced group were taken into account when we compared with our results. Overall inter-rater consistency in the experienced group of observers was highest for tongue base level, followed by collapse of the palate. In this group, lowest consistency was found for hypopharyngeal level, while in our study the highest inter rater consistency was for velum level in AP configuration, followed by epiglottis in also AP configuration. Differences in both studies could be explained by the differences in the study design as they chose to include only six cases, while they employed 97 observers (90 non-experienced and 7 experienced) and finally they selected a different scoring system that was based on key elements of scoring systems recently proposed in the literature. This could be a good example to stress the importance of a universal scoring system in the assessment of DISE. Among the experienced observers, high intra-rater consistency was found for all levels but to a lesser extent for hypopharyngeal collapse, which is similar with our results, except in our study oropharynx and tongue base levels had slightly lower intra-rater consistency rates.

Gillespie et al have both used DISE index and VOTE classification systems when they evaluated validity of DISE. Overall, the intra-rater and inter-rater consistency of the DISE Index score was good, while the intra-rater and inter-rater consistency of the VOTE Index score was fair. Even though they stated no clear reasons behind this, they also proposed that the inclusion of palatine and lingual tonsils assessments in the DISE index might have made the difference.[17] In our study we used VOTE classification system and our results were comparable with those that they obtained via DISE index.

Apart from the studies above, validity of DISE was studied with different aspects. Rodriguez-Bruno et al reported a good test-retest reliability of DISE.[18] They also evaluated inter-rater and intra-rater consistency and found that the consistency of the lateral pharyngeal wall at the level of the velum was lower than for other levels of UA. Which is precisely similar in our study as we found the lowest inter-rater consistency rates at the velum in lateral configuration.

There are several limitations of our study. Primarily the sample size is not large enough. Studies with larger sample size and higher number of the observers could produce more significant results.

## Conclusion

OSA is serious condition with possible metabolic, cardiovascular and pulmonary complications. DISE is a valuable tool that could change the course of treatment for each individual OSA patient. The validity of DISE is quite acceptable although lack of a golden standard classification tool deprive us from "speaking the exact same language". A universal classification system will surely increase the success of DISE and provide a uniform training for those who seek to specialize in the subject.

## References

1. Guilleminault C, Hill MW, Simmons FB, Dement WC. Obstructive sleep apnea: electromyographic and fiberoptic studies. Exp Neurol 1978;62:48-67.
2. Sullivan C, Berthon-Jones M, Issa F, Eves L. Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares. Lancet 1981;1:862-5.
3. Grote L, Hedner J, Grunstein R, Kraiczi H. Therapy with nCPAP: incomplete elimination of sleep related breathing disorder. Eur Respir J 2000;16:921-7.
4. Ravesloot M, deVries N, Stuck BA. Treatment adherence should be taken into account when reporting treatment outcomes in obstructive sleep apnea. Laryngoscope 2014;124:344-5.
5. DeVito A, Llatas MC, Vanni A, et al. European position paper on drug-induced sedation endoscopy (DISE). Sleep Breath 2014;18:453-65.
6. Croft C, Pringle M. Sleep nasendoscopy: a technique of assessment in snoring and obstructive sleep apnoea. Clin Otolaryngol Allied Sci 1991;16:504-9.
7. Kezirian EJ. Nonresponders to pharyngeal surgery for obstructive sleep apnea: Insights from drug-induced sleep endoscopy. Laryngoscope 2011;121:1320-6.
8. Koutsourelakis I, Safiruddin F, Ravesloot M, Zakynthinos S, de Vries N. Surgery for obstructive sleep apnea: sleep endoscopy determinants of outcome. Laryngoscope 2012;122:2587-91.
9. Amos JM, Durr ML, Nardone HC, Baldassari CM, Duggins A, Ishman SL. Systematic review of drug-induced sleep endoscopy scoring systems. Otolaryngol Head and Neck Surg 2018;158:240-8.
10. Kezirian EJ, White DP, Malhotra A, Ma W, McCulloch CE, Goldberg AN. Inter rater reliability of drug-induced sleep endoscopy. Arch Otolaryngol Head Neck Surg 2010;136:393-7.
11. Vroegop AV, Vanderveken OM, Wouters K, et al. Observer variation in drug-induced sleep endoscopy: experienced versus nonexperienced ear, nose, and throat surgeons. Sleep 2013;36:947-53.
12. Carrasco-Llatas M, Zerpa-Zerpa V, Dalmau-Galofre J. Reliability of drug-induced sedation endoscopy: interobserver agreement. Sleep and Breath 2017;21:173-9.
13. Altintas A, Yegin Y, Çelik M, Kaya KH, Koç AK, Kayhan FT. Interobserver Consistency of Drug-Induced Sleep Endoscopy in Diagnosing Obstructive Sleep Apnea Using a VOTE Classification System. J Craniofac Surg 2018;29:140-3.
14. Kezirian EJ, Hohenhorst W, de Vries N. Drug-induced sleep endoscopy: the VOTE classification. Eur Arch Otorhinolaryngol 2011;268:1233-6.
15. Dijemeni E, D'Amone G, Gbati I. Drug-induced sedation endoscopy (DISE) classification systems: a systematic review and meta-analysis. Sleep Breath 2017;21:983-94.
16. Nzekwu C, Dijemeni E. Drug-induced sleep endoscopy (DISE) scoring systems: ideal DISE scoring system and comparability properties. Otolaryngol Head Neck Surg 2018;158:777.
17. Gillespie MB, Reddy RP, White DR, Discolo CM, Overdyk FJ, Nguyen SA. A trial of drug-induced sleep endoscopy in the surgical management of sleep-disordered breathing. Laryngoscope 2013;123:277-82.
18. Rodriguez-Bruno K, Goldberg AN, McCulloch CE, Kezirian EJ. Test-retest reliability of drug-induced sleep endoscopy. Otolaryngol Head Neck Surg 2009;140:646-51.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY- NC-ND3.0) Licence (http://creativecommons.org/licenses/by-nc-nd/3.0/) which permits unrestricted noncommercial use, distribution, and reproduc- tion in any medium, provided the original work is properly cited.

*Please cite this article as:* Kılavuz AE,  Bayram AA. Intra-rater and Inter-rater consistency of drug induced sleep endoscopy. ENT Updates 2019;9(2): 138–143.