

Universitätsklinik und Poliklinik für Hals – Nasen – Ohrenheilkunde, Kopf– und  
Halschirurgie der

Martin-Luther-Universität Halle-Wittenberg

Direktor: Prof. Dr. med habil. Stefan Plontke

**Beurteilung der Interraterreliabilität bei der Diagnostik  
von schlafbezogenen Atmungsstörungen im Rahmen  
der medikamentös–induzierten–Schlafendoskopie**

**Anwendung und Vergleich von verschiedenen  
Klassifikationssystemen**

Dissertation

zur Erlangung des akademischen Grades

Doktor der Medizin (Dr. med.)

vorgelegt

der Medizinischen Fakultät

der Martin-Luther-Universität Halle-Wittenberg

von

Clara Heidkamp

geboren am 19.08.1986 in Bergisch Gladbach

Betreuer: PD Dr. med. habil. Michael Herzog

Gutachter:

1. PD Dr. med. habil. Michael Herzog, Cottbus

2. Prof. Dr. med. habil. Matthias Girndt

3. Prof. Dr. med. habil. Steffen Maune, Köln

02.05.2017

06.11.2017

Für meine Eltern  
und meine verstorbene Großmutter Nenna

## Referat

Die vorliegende Studie wurde als prospektive Evaluation der Interraterreliabilität bei der Beurteilung von Kollapsmustern im Rahmen der medikamentös-induzierten Schlafendoskopie (MISE) durchgeführt. Die MISE stellt ein häufig angewandtes Verfahren zur Topodiagnostik der schlafbezogenen Atmungsstörungen dar. Die Einschätzung der Kollapsneigungen der oberen Atemwege ist dabei abhängig von den subjektiven Kriterien des Untersuchers, verbindliche Standards liegen nicht vor.

Zur Entwicklung diagnostischer Standards wurden die Einschätzungsergebnisse vieler Untersucher (Rater) erhoben. Dafür evaluierten 48 Rater die Videoaufzeichnungen mit Schnarchepisoden von unterschiedlichen Patienten, die im Zusammenhang mit einer MISE aquiriert worden waren. Zur Beurteilung der Kollapsmuster wurden parallel zwei verschiedene Klassifizierungssysteme vorgelegt, das etablierte VOTE-System und ein neues, noch nicht erprobtes Piktogrammsystem. Bei vorgegebenem Level (Velum, Oropharynx, Zungengrund, Epiglottis) wurden die Kollapsereignisse von den Ratern hinsichtlich ihrer Richtung und Ausprägung eingestuft.

Primäres Ziel der Studie war der Vergleich der beiden Klassifizierungssysteme hinsichtlich der errechneten Interraterreliabilität. Sekundäre Ziele waren die Evaluierung des Einflusses der Kollapslevel und Kollapsmuster und des Einflusses der Rater auf die Ausprägung der Interraterreliabilität.

Der Vergleich des VOTE- und Piktogrammsystems erbrachte annähernd gleiche, zum Teil kongruente Übereinstimmungswerte, so dass von zwei äquivalenten Klassifizierungssystemen der MISE ausgegangen werden kann. Die Evaluation der Interraterreliabilität auf der Ebene der Kollapslevel und Kollapsmuster führte zu den Stärken und Schwächen gegenwärtiger Schnarchdiagnostik: Sequenzen ohne Kollaps oder mit Vibrationen, teilweise auch Kollapse mit maximaler Obstruktion, zeigten eine höhere Interraterreliabilität als Sequenzen mit mittlerer Kollapsausprägung; Epiglottis-, Velum- und Zungengrundlevel erreichten eine stärker ausgeprägte Interraterreliabilität als das Oropharynxlevel. Die Analyse des Klassifikationsverhaltens erfahrener und unerfahrener Rater ergab, dass es keinen nennenswerten Unterschied bei der Ausprägung der Interraterreliabilität in beiden Gruppen gibt, ebenso lässt sich in der Summe keine Präferenz der Klassifizierungssysteme erkennen.

Zusammenfassend lässt sich festhalten, dass mit der vorliegenden Studie erstmalig an einem großen Teilnehmerkollektiv ein direkter Methodenvergleich von Klassifizierungssystemen der MISE mit präziser Dokumentation von Level, Grad und Muster der Kollapsereignisse durchgeführt wurde. Auf der Basis dieser Daten und den mit der Interraterreliabilität bilanzierten Stärken und Schwächen aktueller Schnarchdiagnostik können der Entwicklungs- und Forschungsbedarf formuliert werden und weitergehende Studien erfolgen.

Heidkamp, Clara: Beurteilung der Interraterreliabilität bei der Diagnostik von schlafbezogenen Atmungsstörungen im Rahmen der medikamentös-induzierten-Schlafendoskopie – Anwendung und Vergleich von verschiedenen Klassifikationssystemen. Halle, Univ., Med.-Fak., Diss., 73 Seiten, 2017

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>III</b>
<b>Tabellenverzeichnis .....</b>	<b>V</b>
<b>1 Einleitung .....</b>	<b>1</b>
1.1 Schnarchen als Problemhintergrund.....	1
1.2 Entstehung und Erscheinungsformen des Schnarchens.....	2
1.2.1 Entstehung des Schnarchens .....	2
1.2.2 Erscheinungsformen des Schnarchens .....	2
1.3 Diagnoseverfahren .....	4
1.3.1 Anamnese.....	4
1.3.2 Klinische Untersuchung mit Schwerpunkt HNO - Status .....	4
1.3.3 Spezielle apparative Diagnoseverfahren .....	4
<b>2 Fragestellung und Zielsetzung der Studie.....</b>	<b>10</b>
<b>3 Material und Methoden .....</b>	<b>11</b>
3.1 Material .....	11
3.1.1 Exemplarische Schnarchereignisse .....	11
3.1.2 Kollapsmuster .....	11
3.1.3 VOTE – Klassifikation .....	11
3.1.4 Piktogramm – Klassifikation .....	12
3.2 Methoden .....	14
3.2.1 Allgemeines zur Online – Befragung .....	14
3.2.2 Online – Fragebögen/ Klassifikationsbögen .....	15
3.3 Statistik .....	16
3.3.1 Beurteilerübereinstimmung und Interraterreliabilität .....	16
3.3.2 Statistische Aufbereitung der erhobenen Daten .....	17
<b>4 Ergebnisse.....</b>	<b>18</b>
4.1 Studienteilnehmer .....	18
4.2 Beurteilerübereinstimmung.....	19
4.3 Interraterreliabilität.....	19
4.3.1 Interraterreliabilität auf Epiglottis - Level.....	20
4.3.2 Interraterreliabilität auf Oropharynx – Level.....	25
4.3.3 Interraterreliabilität auf Velum – Level .....	36
4.3.4 Interraterreliabilität auf Zungengrund – Level .....	49
<b>5 Diskussion .....</b>	<b>56</b>

---

5.1	Ausprägung der Interraterreliabilität bei VOTE und Piktogrammen – Vergleich der Klassifikationssysteme.....	56
5.2	Einfluss der Kollapslevel und Kollapsmuster auf die Ausprägung der Interraterreliabilität.....	57
5.3	Einfluss der Professionalität und schlafmedizinischen Erfahrung der Rater auf die Ausprägung der Interraterreliabilität.....	58
5.4	Vergleich mit anderen Studien zur Interraterreliabilität und anderen Formen der Reliabilität .....	61
5.4.1	Rodriguez – Bruno et al., 2009: Test – Retest reliability of drug – induced sleep endoscopy.....	61
5.4.2	Kezirian et al., 2010: Interrater – Reliability of Drug – Induced Sleep Endoscopy .....	62
5.4.3	Gillespie et al., 2013: A Trial of Drug – Induced Sleep Endoscopy in the Surgical Management of Sleep – Disordered Breathing .....	63
5.4.4	Vroegop et al., 2013: Observer Variation in Drug – Induced Sleep Endoscopy: Experienced Versus Non – Experienced Ear, Nose an Throat Surgeons .....	64
5.4.5	Kim et al., 2014: Interrater – Reliability of Sleep Videofluoroscopy for Airway Obstruction in Obstructive Sleep Apnea .....	64
5.5	Klinische Bedeutung der Studie mit kritischen Anmerkungen und Schlussfolgerungen.....	65
	<b>Zusammenfassung.....</b>	<b>67</b>
	<b>Literaturverzeichnis .....</b>	<b>69</b>
	<b>Thesen .....</b>	<b>73</b>
	<b>Tabellarischer Lebenslauf .....</b>	<b>74</b>
	<b>Selbstständigkeitserklärung .....</b>	<b>76</b>
	<b>Erklärung über frühere Promotionsversuche .....</b>	<b>77</b>
	<b>Danksagung.....</b>	<b>78</b>

## Abbildungsverzeichnis

	<u>Seite</u>
<b>Abbildung 1:</b> VOTE - Schema.....	12
<b>Abbildung 2:</b> Piktogrammschema Epiglottislevel.....	13
<b>Abbildung 3:</b> Piktogrammschema Oropharynxlevel.....	13
<b>Abbildung 4:</b> Piktogrammschema Velumlevel .....	14
<b>Abbildung 5:</b> Piktogrammschema Zungengrund .....	14
<b>Abbildung 6:</b> Ausschnitt eines Klassifizierungsbogens, Velum.....	16
<b>Abbildung 7:</b> Ausschnitt eines Klassifizierungsbogens, Velum.....	16
<b>Abbildung 8:</b> Zu beurteilendes Level: Epiglottis, Kollapsmuster: Kein Kollaps.....	20
<b>Abbildung 9:</b> Zu beurteilendes Level: Epiglottis, Kollapsmuster: Vibrationen des cranialen Epiglottisrandes, keine Obstruktion. Roter Pfeil: Verlagerungsrichtung .....	21
<b>Abbildung 10:</b> Zu beurteilendes Level: Epiglottis. Kollapsmuster: Vibrationen des cranialen Epiglottisrandes und der cranialen Hälfte der Epiglottis mit partieller Obstruktion. Roter Pfeil: Verlagerungsrichtung .....	22
<b>Abbildung 11:</b> Zu beurteilendes Level: Epiglottis, Kollapsmuster: Partielles Epiglottisansaugphänomen mit Restlumen am linken Epiglottisrand (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung .....	23
<b>Abbildung 12:</b> Zu beurteilendes Level: Epiglottis, Kollapsmuster: Epiglottisansaugphänomen, Verschluss des Larynxeinganges mit obstruktiver Apnoe. Roter Pfeil: Verlagerungsrichtung .....	24
<b>Abbildung 13:</b> Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Kein Kollaps.....	25
<b>Abbildung 14:</b> Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler Kollaps der Pharynxwand mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung .....	27
<b>Abbildung 15:</b> Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler partieller Kollaps der Tonsillen mit Restlumen. Roter Pfeil: Verlagerungsrichtung.....	28
<b>Abbildung 16:</b> Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler partieller Kollaps der Tonsillen mit Restlumen. Roter Pfeil: Verlagerungsrichtung.....	30
<b>Abbildung 17:</b> Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler Kollaps der Tonsillen mit kompletter Obstruktion. Roter Pfeil: Verlagerungsrichtung.....	31

<b>Abbildung 18:</b>	Zu beurteilendes Level: Caudaler Oropharynx auf Zungengrundniveau, Kollapsmuster: Kompletter Kollaps konzentrisch mit Obstruktion. Roter Pfeil: Verlagerungsrichtung .....	33
<b>Abbildung 19:</b>	Zu beurteilendes Level: Caudaler Oropharynx auf Zungengrundniveau, Kollapsmuster: Kompletter Kollaps konzentrisch mit Obstruktion. Roter Pfeil: Verlagerungsrichtung .....	34
<b>Abbildung 20:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Kein Kollaps .....	36
<b>Abbildung 21:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Vibration des Velums anterior-posterior. Roter Pfeil: Verlagerungsrichtung.....	38
<b>Abbildung 22:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Vibration des Velums anterior-posterior. Roter Pfeil: Verlagerungsrichtung.....	40
<b>Abbildung 23:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Komplette Obstruktion des Velums anterior-posterior. Roter Pfeil: Verlagerungsrichtung ....	42
<b>Abbildung 24:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Partielle konzentrische Obstruktion des Velums mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung.....	43
<b>Abbildung 25:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Komplette konzentrische Obstruktion des Velums. Roter Pfeil: Verlagerungsrichtung .....	45
<b>Abbildung 26:</b>	Zu beurteilendes Level: Velum, Kollapsmuster: Partielle laterale Obstruktion des Velums. Roter Pfeil: Verlagerungsrichtung .....	46
<b>Abbildung 27:</b>	Zu beurteilendes Level: Zungengrund, Kollapsmuster: Kein Kollaps	49
<b>Abbildung 28:</b>	Zu beurteilendes Level: Zungengrund, Kollapsmuster: Partieller Kollaps anterior-posterior mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung .....	50
<b>Abbildung 29:</b>	Zu beurteilendes Level: Zungengrund, Kollapsmuster: Komplette Obstruktion anterior-posterior ohne Restlumen. Roter Pfeil: Verlagerungsrichtung .....	52
<b>Abbildung 30:</b>	Zu beurteilendes Level: Zungengrund, Kollapsmuster: Partieller Kollaps anterior-posterior der hyperplastischen Zungengrundtonsille mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung.....	53
<b>Abbildung 31:</b>	Zu beurteilendes Level: Zungengrund, Kollapsmuster: Komplette Obstruktion anterior-posterior der hyperplastischen Zungengrundtonsille ohne Restlumen. Roter Pfeil: Verlagerungsrichtung .....	54

## Tabellenverzeichnis

	<u>Seite</u>
<b>Tabelle 1:</b> Sequenz 1: Epiglottis, s. Abb. 8 .....	20
<b>Tabelle 2:</b> Sequenz 4: Epiglottis, s. Abb. 9 .....	21
<b>Tabelle 3:</b> Sequenz 5: Epiglottis, s. Abb. 1 .....	22
<b>Tabelle 4:</b> Sequenz 3: Epiglottis, s. Abb. 11 .....	23
<b>Tabelle 5:</b> Sequenz 2: Epiglottis, s. Abb. 12 .....	24
<b>Tabelle 6:</b> Sequenz 6: Oropharynx, s. Abb. 1 .....	26
<b>Tabelle 7:</b> Sequenz 9: Oropharynx, s. Abb. 14. ....	27
<b>Tabelle 8:</b> Sequenz 10: Oropharynx, s. Abb. 15 .....	28
<b>Tabelle 9:</b> Sequenz 11: Oropharynx, s. Abb. 16 .....	30
<b>Tabelle 10:</b> Sequenz 12: Oropharynx, s. Abb. 17 .....	31
<b>Tabelle 11:</b> Sequenz 7: Oropharynx, s. Abb. 18 .....	33
<b>Tabelle 12:</b> Sequenz 8: Oropharynx, s. Abb. 19. ....	34
<b>Tabelle 13:</b> Sequenz 13: Velum, s. Abb. 20.....	36
<b>Tabelle 14:</b> Sequenz 14: Velum, s. Abb. 21.....	38
<b>Tabelle 15:</b> Sequenz 15: Velum, s. Abb. 22.....	40
<b>Tabelle 16:</b> Sequenz 16: Velum, s. Abb. 23.....	42
<b>Tabelle 17:</b> Sequenz 17: Velum, s. Abb. 24.....	43
<b>Tabelle 18:</b> Sequenz 18: Velum, s. Abb. 25.....	45
<b>Tabelle 19:</b> Sequenz 19: Velum, s. Abb. 26.....	46
<b>Tabelle 20:</b> Sequenz 20: Zungengrund, s. Abb. 27 .....	49
<b>Tabelle 21:</b> Sequenz 21: Zungengrund, s. Abb. 28.....	50
<b>Tabelle 22:</b> Sequenz 22: Zungengrund, s. Abb. 29.....	52
<b>Tabelle 23:</b> Sequenz 23: Zungengrund, s. Abb. 30.....	53
<b>Tabelle 24:</b> Sequenz 24: Zungengrund, s. Abb. 31.....	54

# 1 Einleitung

## 1.1 Schnarchen als Problemhintergrund

Schnarchen als altbekanntes Phänomen wird in der Literatur gerne karikiert und als komisches Ereignis verharmlost wie in den Versen von Eugen Roth:

Ein Mensch hört staunend und empört,  
Daß(sic) er, als Unmensch, alle stört,  
Er nämlich bildet selbst sich ein,  
Der angenehmste Mensch zu sein,  
Ein Beispiel macht euch solches klar:  
Der Schnarcher selbst schläft wunderbar.

Die nächtliche Szenerie ist vorstellbar: Das rasselnde, pfeifende, sägende Atemgeräusch eines Schnarchenden! Der betroffene Angehörige liegt stundenlang wach im Bett daneben. Am nächsten Morgen ein "böses Erwachen" mit Kopfschmerzen, Abgeschlagenheit und Tagesmüdigkeit - auch von Beziehungsproblemen ist die Rede (Virkkula et al., 2005, Ulfberg et al., 2000).

Verbreitet war lange Zeit die falsche Vorstellung vom tiefen, festen, eben "wunderbaren" Schlaf des Schnarchers. Inzwischen ist ins allgemeine Bewusstsein gedrungen, dass der Schnarcher „wunderbar“ allenfalls phasenweise in der Nacht schläft, denn regelmäßiges Schnarchen führt repräsentativer Studien zufolge auch nach Ausschluss bekannter Risikofaktoren langfristig häufiger zu Hypertonie, Diabetes mellitus, Hypercholesterinämie, Myokardinfarkt, Apoplex (Zamarrón et al., 1999, Hu et al., 2000, Roopa et al., 2010) und auf der Alltagsebene zu Heiserkeit, Kopfschmerzen, schlechter Schlafqualität mit vermehrtem Auftreten von Alpträumen sowie verminderter Leistungsfähigkeit in Ausbildung und Beruf (Hamdan et al., 2012, Scher et al., 2003, Ficker et al., 1999, de Groen et al., 1993).

Gemäß der Autoren Young (1993), Ohayon et al. (1997) und Stuck et al. (2010) sind je nach Alter und Geschlecht 28 - 62% der Bevölkerung vom regelmäßigen oder unregelmäßigen Schnarchen betroffen. Das lästige, potenziell krankmachende Schnarchen und das hohe Konfliktpotenzial im sozialen Umfeld der Betroffenen stellt eine große Herausforderung für die Medizin nach Beratung, eindeutiger Diagnostik und effizienter Therapie dar, zumal „[...] die individuelle Erwartung an eine Leistungswiederherstellung durch erholsamen Schlaf gewachsen ist“ (Penzel, 2013) und die Schnarchproblematik in der Gesellschaft durch eine alternde Bevölkerung immer weiter ansteigt.

Der Präsident der Deutschen Gesellschaft für Hals - Nasen - Ohren - Heilkunde, Kopf - und Halschirurgie e.V., Thomas Deitmer, konstatiert in diesem Zusammenhang, dass *„die Hals - Nasen - Ohrenheilkunde in Deutschland [...] zunehmend unter Druck [stehe], tradierte Behandlungsformen zu überprüfen, auf eine Evidenz - basierte Basis zu stellen und so bei der kompetitiven Situation um Ressourcen im Gesundheitswesen die Inhalte des Faches für die Zukunft zu etablieren“* (Deitmer, 2014).

Die Forderung einer Weiterentwicklung im Bereich der Hals - Nasen - Ohren - Heilkunde trifft auch auf den Teilbereich der Schnarchdiagnostik und Therapie zu. Indizien für die hier vorliegenden Defizite sind:

1. eine fehlende eindeutige Definition des Phänomens Schnarchen
2. fehlende Parameter für die Ausprägung und Therapie des Schnarchens und untersucherabhängige Einschätzung des Phänomens
3. keine klare begriffliche Trennung von Schnarchen als eigenständigem Phänomen und Schnarchen als Symptom für eine Krankheit (z.B. OSA)
4. Mangel an klinischen Studien (Stuck et al., 2010)

## **1.2 Entstehung und Erscheinungsformen des Schnarchens**

### **1.2.1 Entstehung des Schnarchens**

Schnarchen ist auf eine Erschlaffung der Muskulatur des Gaumens und des Rachens zurückzuführen, die eine Verengung der oberen extrathorakalen Atemwege bewirkt. In der Folge kommt es zu einer Beschleunigung des Atemflusses und zu Vibrationen der Weichteilstrukturen, wodurch die Schnarchgeräusche ausgelöst werden.

Eine gefährliche Variante der schlafbezogenen Atmungsstörungen stellt die Obstruktive Schlafapnoe (OSA) dar. Sie ist gekennzeichnet durch wiederholte Kollapse in den oberen Atemwegen während des Schlafes, die einen teilweisen oder kompletten Verschluss provozieren. Allgemein begünstigen Faktoren wie Alter, Geschlecht, Übergewicht, Lebensgewohnheiten (Alkoholkonsum, Rauchen) und anatomische Gegebenheiten (Engstellung der Nase, Schleimhautüberschuss im Bereich des weichen Gaumens etc) die Entstehung des Schnarchens.

### **1.2.2 Erscheinungsformen des Schnarchens**

#### **1.2.2.1 Primäres Schnarchen**

Primäres Schnarchen (auch: habituelles Schnarchen, ungefährliches Schnarchen, nicht apnoisches Schnarchen) ist gekennzeichnet durch laute Atmungsgeräusche, die als regelmäßig, tieffrequent und gleichförmig beschrieben werden (Pirsig, 1998).

Primäres Schnarchen ist nicht eindeutig von Schnarchen bei OSA abzugrenzen. Deshalb ist eine Begriffsbestimmung schwierig. Die u.a. Definition des Schnarchens entsprechend der ICSD – 3 (Sateia, 2014) stellt einen Anhaltspunkt dar. Sie resultiert aus den Erfahrungen klinischer Praxis und Diagnostik und ist als hypothetisch und vorläufig zu bewerten.

1. „Die/der Betroffene bzw. die Bettpartnerin/der Bettpartner des/der Betroffenen berichtet über atmungsabhängige, meist inspiratorische akustische Phänomene im Schlaf, wobei objektive Parameter zur Definition dieser akustischen Phänomene als „Schnarchen“ derzeit nicht zur Verfügung stehen.
2. Die/der Betroffene klagt nicht über eine Schlafstörung, die ursächlich auf das Schnarchen zurückgeführt werden kann.
3. Eine schlafmedizinische Diagnostik ergibt keinen Hinweis auf das Vorliegen einer schlafbezogenen Atmungsstörung.“

Schnarchen in diesem Sinne gilt nicht als Krankheit und muss aus medizinischer Sicht nicht behandelt werden.

#### **1.2.2.2 Schnarchen bei Obstruktiver Schlafapnoe (OSA)**

Schnarchen gilt als Begleitsymptom der OSA und tritt laut DGSM – Leitlinie (2009, Anm.: Diese Leitlinie wird gerade aktualisiert.) *„bei 95% der Betroffenen auf“*. Schnarchen bei OSA ist unterschiedlich ausgeprägt und kann zum Beispiel wie primäres/ habituelles Schnarchen auftreten. In der Regel wird Schnarchen bei OSA jedoch als laut und unregelmäßig mit dazwischen liegenden Apnoen von mehr als zehn Sekunden Dauer beschrieben (Carskadon et al., 1994, Guilleminault et al., 1989). Eine Objektivierung der Diagnose OSA erfolgt über den Apnoe – Hypopnoe – Index (AHI). Mit Hilfe der drei Stufen des AHI - leicht <15; mittelgradig >15 und <30; schwer > 30 - kann zusammen mit der u.a. typischen Symptomatik der Schweregrad einer OSA eingeschätzt werden.

Bekannte Symptome der OSA sind Tagesmüdigkeit, vermehrte Einschlafneigung am Tage und eine Einschränkung der Konzentration und Leistungsfähigkeit (Hamdan et al., 2012, Scher et al., 2003, Ficker et al., 1999, de Groen et al., 1993). Die eingangs erwähnten Folgeerkrankungen des regelmäßigen Schnarchens (s. Kapitel 1.1) können bei OSA verstärkt prognostiziert werden und sind laut HE et al. (1988) mit einem erheblich erhöhten Mortalitätsrisiko verbunden.

## 1.3 Diagnoseverfahren

### 1.3.1 Anamnese

Grundlage aller diagnostischen Verfahren ist eine genaue Anamnese der Symptomatik und des Schlafverhaltens der Patienten. Zusätzlich sollte die Anamnese die lebensweltlichen Bedingungen im privaten und beruflichen Umfeld der Patienten sowie Vorerkrankungen, Medikamenteneinnahme, Alkoholkonsum etc. erfassen und die Familien – und Fremdanamnese miteinschließen.

Ergänzend können Fragebögen und Tagebuchaufzeichnungen/ Protokolle zur Selbstbeurteilung hinzugezogen werden (DGSM – Leitlinie „Nicht erholsamer Schlaf“, 2009).

### 1.3.2 Klinische Untersuchung mit Schwerpunkt HNO - Status

Bei der Ursachenforschung des Schnarchens ist neben der kardiopulmonalen vor allem die HNO - ärztliche Untersuchung von zentraler Bedeutung, da häufig die anatomischen Gegebenheiten der oberen Atemwege die Entstehung von Schnarchgeräuschen und Obstruktionen begünstigen oder verursachen. Die Untersuchung umfasst eine genaue Betrachtung der Nase und des Nasenrachens, vor allem unter dem Aspekt der „*strömungsrelevanten nasalen Strukturen*“ und „*kann auch eine endoskopische Beurteilung beinhalten*“ (Stuck et al., 2013).

Ferner erfolgt eine Inspektion des Oro- und Hypopharynx sowie des Larynx. Auch die Morphologie des Gesichtsschädels sollte in die Bewertung miteinbezogen werden. Bestandteil der klinischen Untersuchung ist ebenfalls eine allergologische Basisdiagnostik.

### 1.3.3 Spezielle apparative Diagnoseverfahren

Den Goldstandard in der Detektion der schlafbezogenen Atemstörungen stellt die Polysomnographie (PSG) dar. Als vorgeschaltetes Diagnostikinstrument wird jedoch häufig die vereinfachte Methode der Polygraphie (PG) eingesetzt.

PSG und PG sind als quantitative Methoden einzustufen. Als qualitative Methode steht die medikamentös – induzierte Schlafendoskopie (MISE) zur Verfügung, die in den letzten Jahren die Bedeutung der Topodiagnostik der OSA erlangt hat. Beide Methodenformen können ergänzend angewandt werden.

PSG und PG bilden die diagnostische Grundlage für die nächtliche Überdruckbeatmung (CPAP – Beatmung), die laut DGSM – Leitlinie (2009) als effektivste Therapieform der OSA gilt (außer der Tracheotomie in Notfällen). Wird die CPAP – Therapie von Patienten nicht toleriert und/ oder besteht ein alternativer

Therapiewunsch zur Beatmungstherapie, z.B. nach prognerierenden Schienen oder operativer Sanierung, muss eine zielgerichtete Diagnostik mit Hilfe der MISE erfolgen.

### **1.3.3.1 Polysomnographie**

Mit der komplexen Technik der Polysomnographie werden während der ganzen Nacht kontinuierlich die folgenden unterschiedlichen Körperfunktionen überwacht:

- Hirnströme mittels EEG (19 Kanäle)
- Herzaktivität mittels EKG
- Muskelbewegungen mittels EMG (dabei auch Berücksichtigung der Körperlage)
- Augenbewegungen mittels Elektrookulogramm
- Sauerstoffsättigung des Blutes mittels Pulsoxymetrie
- Atemflussmessung nasal oder oral
- Atemexkursionen des Thorax und Abdomen
- Körpertemperatur
- Blutdruckmessungen

Tonaufzeichnungen und Videoaufnahmen können die diagnostischen Maßnahmen ergänzend unterstützen (Moser et al., 2009).

### **1.3.3.2 Polygraphie**

Die Polygraphie wird in der Regel mit portablen Geräten mit bis zu sechs Messkanälen, die unter dem Sammelbegriff „Polygraphiesysteme“ fungieren, durchgeführt. Zur Diagnostik der OSA werden Geräte mit vier bis sechs Kanälen eingesetzt, laut DGSM reichen sie zur Erkennung der OSA aus.

Ihr Leistungsspektrum umfasst die folgenden Messungen:

- Herzaktivität mittels EKG
- Sauerstoffsättigung des Blutes mittels Pulsoxymetrie
- Atemflussmessung nasal oder oral
- Registrierung der Körperlage mittels Lagesensoren (Akzelerometer)

Da bei der Polygraphie keine Schlafstadien mittels EEG gemessen werden, ist die Aussagefähigkeit der Methode begrenzt, so dass vor allem im Zusammenhang mit komorbiden Erkrankungen auf die umfangreichere Polysomnographie zurückgegriffen werden muss.

### **1.3.3.3 Medikamentös induzierte Schlafendoskopie (MISE)**

Mit der MISE können die Entstehungsorte der Vibrationen, der Obstruktionen und der Kollapse in den oberen Atemwegen lokalisiert werden. Die Patienten werden unter kardiorespiratorischer Überwachung in einen künstlichen Schlaf versetzt.

Mittels eines feinen Endoskops, das durch die Nase in den Nasenrachenraum eingeführt wird, lassen sich die Kollapsmuster in mehreren Ebenen (Velum, Oropharynx, Zungengrund, Epiglottis) detektieren und in ihrer Intensität und Ausprägung beurteilen (Steinhart et al., 2000, Johal et al., 2005). Die Voraussetzungen, Vorgehensweisen und erforderlichen Weiterentwicklungen der MISE sind im „European Position Paper on drug – induced – sleep – endoscopy“ (Vito et al., 2014) detailliert zusammengefasst.

Ausgehend von der Indikation für die MISE nach alternativen Therapien zur CPAP – Therapie, z.B. die Uvulopalatopharyngoplastik (UPPP), sehen die Experten konsensuell für eine erfolgreiche Durchführung der MISE die folgenden Bedingungen:

- Notwendigkeit von Voruntersuchungen (s. Kap. 1.3.1 – 1.3.3.2)
- spezielles technisches Equipment für die MISE
- geschultes Personal
- Raum mit kardiorespiratorischer Überwachung als Ausführungsort für die MISE
- Lokalanästhesie der Nase, Anschwellen der Nase, zusätzlich antisekretorische Medikamente
- Positionierung des Patienten in Abhängigkeit davon, ob Schnarchphänomene in Rückenlage und/ oder Seitenlage auftreten
- Überwachungsfenster -> kompletter stabiler Schnarchzyklus (gegenwärtige Kritik an der MISE -> zu kurze Untersuchungszeit, die nur einen kurzen Simulationsausschnitt erfasst), ausreichende Schlaftiefe (BIS 50 – 70 oder CSI)
- Anästhetika

Als Anästhetika stehen Midazolam und Propofol zur Verfügung. Propofol kommt gegenüber dem Midazolam wegen geringerer Beeinflussung des Atemzentrums, schwächerer Muskelrelaxationen und guter Kontrollierbarkeit häufiger zur Anwendung. Gleichermaßen werden beide Anästhetika kombiniert eingesetzt mit den Vorteilen, dass der Patient durch das Propofol schneller einschläft und für das Midazolam ein Antidot vorhanden ist (Vito et al., 2014).

In dem zuvor genannten Positionspapier der Europäischen Gesellschaft für Schlafmedizin wird zur Weiterentwicklung der Diagnoseverfahren primär die

Verständigung auf ein gemeinsames und universell akzeptiertes Klassifikationssystem gefordert.

Weiterhin müssen die klinischen Möglichkeiten und die technischen Begleiterscheinungen der MISE besser erforscht werden, zum Beispiel:

- die Rolle der MISE bei der Stimulation der oberen Atemwege oder prognerierenden Schienen als Therapieoption
- die Optimierung des Potentials der MISE in der OSA – Behandlung
- Überprüfung und Evaluation der Methoden für das Messen der Schlaftiefe während der MISE
- der Einfluss der Mundöffnung oder des Mundschlusses während der MISE in Relation zum Atemwegwiderstand der oberen Atemwege
- die Analyse der Positionierung der Patienten während der MISE
- Untersuchung des Einflusses zentraler Apnoe während der MISE
- Verbesserung der Kenntnisse des Einflusses von Sedativa auf Kollapsmuster und Level

Die letztgenannte Forderung steht im Zusammenhang mit Kritik an der MISE wegen der Entstehung von Muskelrelaxationen, die ein falsch positives obstruktives Schnarchen erzeugen können (Johal et al., 2005, Kotecha et al., 2006, Hewitt et al., 2008, Gillespie et al., 2013).

Mit der Thematik der Muskelrelaxationen während der Narkose beschäftigt sich eine Studie der Universität Halle (Kellner et al., 2013). Sie beschreibt eine Veränderung der Pharynxrelaxationen in Abhängigkeit von der Sedierungstiefe und weist nach, dass sich das Obstruktions - und Schnarchverhalten entsprechend der Propofoldosierung ändert. Werte zur individuellen Dosierung des Anästhetikums und zu den *„therapeutischen Konsequenzen aus einem sedierungsabhängig ändernden Obstruktions - und Vibrationsverhalten“* (ebenda) der oberen Atemwege liegen noch nicht vor.

Als notwendig angesehen wird zusätzlich ein Methodenvergleich der MISE mit anderen Diagnoseverfahren mit präziser Dokumentation der Bewertungsunterschiede von Grad, Level und Muster der Kollapsereignisse. Auch der Zusammenhang von der mit der MISE erstellten Diagnostik und den Behandlungsergebnissen muss analysiert werden.

Von Gillespie et al. liegt aus dem Jahre 2013 eine Studie zu dieser Thematik vor. Die Studie weist den Diagnose – und Therapieerfolg bei den durch die MISE entdeckten Schnarchereignissen nach. 75% der durch die MISE forcierten OPs (UPPP) waren

erfolgreich. Der Einfluss der MISE auf die Entscheidung zu operieren oder die operative Planung zu ändern, war hoch: Nach zuvor erstellter Diagnose mittels Wachendoskopie mit Müller – Manöver, bei der durch negativen Druck das instabilste Segment der oberen Atemwege kollabiert und das Problem auf diese Art und Weise identifiziert wird, wurde nach zusätzlicher Untersuchung mit der MISE in 62% der Fälle der chirurgische Plan geändert. Allerdings gibt es noch keine Langzeitergebnisse zur Beständigkeit der Operationserfolge (Gillespie et al., 2013).

Zusammenfassend kann festgestellt werden, dass die MISE als qualitatives Assessment die dynamische Anatomie der schlafbezogenen Atemstörungen erfasst und die Behandlungsoptionen aufzeigt. Sie stellt eine zur Zeit ersatzlose Alternative im Bereich der (Topo) Diagnostik dar (Barbar - Craig et al., 2000, Kotecha et al., 2008).

#### **1.3.3.4 Klassifikationsmöglichkeiten der MISE**

In dem zuvor genannten Positionspapier der Europäischen Schlafgesellschaft (Vito et al., 2014) werden die aktuell zur Anwendung kommenden Klassifikationsysteme der MISE aufgeführt. Präferenzen im Sinne von Empfehlungen werden nicht angezeigt.

- Pringle, 1993: Unterscheidung der Kollapsereignisse nach Schweregrad und Muster:
  - Grad 1 -> einfaches Gaumen – Schnarchen
  - Grad 2 -> Obstruktion am Weichgaumen
  - Grad 3 -> multisegmentaler intermittierender oropharyngealer Kollaps
  - Grad 4 -> multisegmentaler kontinuierlicher Kollaps
  - Grad 5 -> Zungengrundkollaps
- Camillieri, 1995: Unterscheidung der Kollapsereignisse nach Schweregrad und Muster:
  - Grad 1 -> gaumenbedingtes Schnarchen
  - Grad 2 -> gemischtes Schnarchen
  - Grad 3 -> nicht gaumenbedingtes Schnarchen (Zungengrundschnarchen)
- Kezirian, 2011: Unterscheidung der Kollapsereignisse nach
  - Level: Velum, Oropharynx, Zungengrund, Epiglottis
  - Schweregrad: kein Kollaps, partieller Kollaps, kompletter Kollaps
  - Muster: anterior – posterior, lateral, konzentrisch
- Vincini, 2012: Unterscheidung der Kollapsereignisse nach Level: Nase, Oropharynx, Hypopharynx, Larynx -> unterteilt in

- A: Supraglottis,
- B: Glottis;
- Schweregrad: 0-25% = Grad 1, 25-50%= Grad 2, 50-75%= Grad 3, 75-100%= Grad 4;
- Muster: anterior – posterior, lateral, konzentrisch
  
- Bachar, 2012: Unterscheidung der Kollapsereignisse nach
  - Level: Nase + Nasopharynx, Gaumen inklusive Tonsillen, Zungengrund, Larynx, Hypopharynx;
  - Schweregrad: kein Kollaps = 0, partieller Kollaps = 1, kompletter Kollaps = 2
  
- Gillespie, 2013: Unterscheidung der Kollapsereignisse nach Level und Muster: Gaumen AP, Hypopharynx, Tonsillen, Zungengrund, Epiglottis, Schweregrad entsprechend dem DISE – Index: 0 – 1 – 2 – 3/4, 0 = kein Kollaps, 1 = partieller Kollaps, 2 = kompletter Kollaps, 3/4 = nicht klassifizierbar (betrifft -> Gaumen, Hypopharynx, Tonsillen, Epiglottis) oder bezüglich des Zungengrundes: 3 = kompletter Kollaps mit präsenten Zungengrundtonsillen, 4 = kompletter Kollaps ohne sichtbare Zungengrundtonsillen
  
- Kou, 2013: Unterscheidung der Kollapsereignisse nach
  - Level: retropalatal, retrolingual
  - Schweregrad: 0/1/2;
  - Muster: anterior – posterior, lateral, benachbarte Strukturen kollabieren/vibrieren mit.

## 2 Fragestellung und Zielsetzung der Studie

Ausgangspunkt der vorliegenden Studie war die dargestellte Situation untersucherabhängiger Einschätzungen und fehlender Standards in der Schnarchdiagnostik und Therapie. Die zuvor angeführten Klassifizierungssysteme der MISE spiegeln diesen Zustand wider: Viele Systeme mit unterschiedlichen Kategorien und Schwerpunktsetzungen führen zu unterschiedlichen Ergebnissen und Bewertungen. Konkret lag darin die Herausforderung für die hier gegebene vergleichende Studie der Klassifizierungssysteme.

Die Wahl fiel auf die herkömmliche VOTE – Methode und die noch nicht erprobte Klassifizierungsmethode über Piktogramme. Als Beurteiler (Rater) wurden niedergelassene HNO – Ärzte oder Klinikärzte rekrutiert. Die Studie wurde als prospektive Evaluation der Interraterreliabilität bei der Beurteilung der Kollapsmuster im Rahmen der MISE durchgeführt. Die Rater sollten die Merkmalsausprägungen gleicher Videosequenzen mit Schnarchepisoden mittels VOTE und parallel mittels Piktogrammen beurteilen und klassifizieren. Es sollte nicht nur eingeschätzt werden, ob ein Merkmal (z.B. eine Obstruktion) vorhanden ist oder nicht, sondern auch dessen Manifestationen (Konfiguration, graduelle Ausprägung etc.) erkannt und eingestuft werden.

Primäres Ziel der Arbeit war der Vergleich der Klassifizierungssysteme in Bezug auf die Interraterreliabilität. Sekundäre Ziele waren Untersuchungen zum Einfluss der Kollapslevel und Kollapsmuster auf die Ausprägung der Interraterreliabilität und zum Einfluss der Professionalität der Rater (klinische Tätigkeit, schlafmedizinische Erfahrung, Zusatzqualifikation) auf die Ausprägung der Interraterreliabilität.

## **3 Material und Methoden**

### **3.1 Material**

#### **3.1.1 Exemplarische Schnarchereignisse**

Das in der vorliegenden Studie zu bewertende Material zur Schnarchproblematik bestand aus Videoaufzeichnungen von Patienten des Universitätsklinikums Halle, die in der klinischen Diagnostik bei Schnarchen und/oder OSA mittels MISE untersucht wurden. Die insgesamt 24 Videosequenzen mit zwei bis drei Schnarchepisoden erfassen jeweils eines der folgenden anatomischen Level:

Cranialer Oropharynx/ Velum, caudaler Oropharynx, Tonsillen und lateraler Pharynx, caudaler Oropharynx / Zungengrund, Epiglottis.

Bei der Auswahl der Beispielsequenzen wurde bewusst auf multilokuläre Kollapsmuster verzichtet. Gegenüber der realen Schlafendoskopie, bei der durch Bewegung des Endoskops Verlauf und angrenzende Level der Kollapsereignisse erfasst werden können, erfolgte in der Studie aus Gründen der Realisierbarkeit und der Übersichtlichkeit die modellhafte Fokussierung auf ein Level. Zusätzlich zu den Bildsequenzen wurden parallel Tonsequenzen mit den entsprechenden Schnarchgeräuschen eingefügt.

Hinweis: Aus Gründen der Übersichtlichkeit und Nachvollziehbarkeit der Studienergebnisse werden fotografische Reproduktionen der hier angeführten Bildsequenzen im Ergebnisteil präsentiert (Kapitel 4.3.1 – 4.3.4).

#### **3.1.2 Kollapsmuster**

Kollapsmuster weisen verschiedene Richtungen (Konfigurationen) und verschiedene Ausprägungen (Schweregrade) auf. Die Kollapsrichtungen werden mit den Begriffen anterior – posterior, lateral und konzentrisch gekennzeichnet. Die Kollapsausprägungen werden mit den Kategorien -> kein Kollaps, partieller Kollaps, kompletter Kollaps (VOTE) und mit den graduellen Abstufungen von 0° - IV° (Piktogramme) erfasst.

#### **3.1.3 VOTE – Klassifikation**

Das VOTE – Bewertungssystem wurde von Kezirian, de Vries und Hohenhorst, drei in der medikamentös induzierten Schlafendoskopie erfahrenen, aus den USA, den Niederlanden und aus Deutschland stammenden Forschern im Jahre 2011 entwickelt. Ihre Intention war, sowohl die Vielzahl von Klassifizierungssystemen als auch die unterschiedlichen Bewertungssysteme der Center, die die MISE durchführen, zu

limitieren, um eine bessere Vergleichbarkeit der Systeme zu gewährleisten. Sie entwickelten mit VOTE ein System, welches sich auf die Kernfunktionen konzentriert.

Nach 8000 MISE – Untersuchungen wurden die für die Obstruktionen verantwortlichen Strukturen (Level) sowie die Konfigurationen und möglichen Schweregrade zusammengefasst.

A) Level/ Konfigurationen der Obstruktionen:

- Velum: laterale Pharynxwand, weicher Gaumen, Uvula (ap, lateral, konzentrisch)
- Oropharynx: laterale Pharynxwand, Gaumenmandeln (lateral)
- Tongue base: Zungenbasisprolaps durch Erschlaffung des Musculus genioglossus (nur anterior – posterior möglich)
- Epiglottis: anterior-posterior, lateral

B) Schweregrade der Obstruktionen:

- 0 = keine Obstruktion oder Vibration; 1 = partielle Obstruktion; 2 = komplette Obstruktion; X = nicht beurteilbar -> hier ist die Beurteilung nicht immer eindeutig.

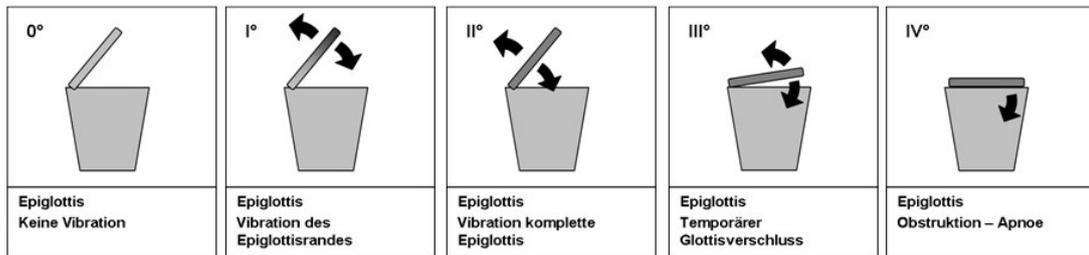
VOTE - Klassifikation				
Anatomische Struktur	Grad der Obstruktion	Konfiguration der Obstruktion / Vibration wenn < 0		
		anterior - posterior	lateral	konzentrisch
Velum				
Laterale Oropharynxwand				
Zungengrund				
Epiglottis				

**Abbildung 1:** VOTE - Schema  
Quelle: Kezirian et al., 2011

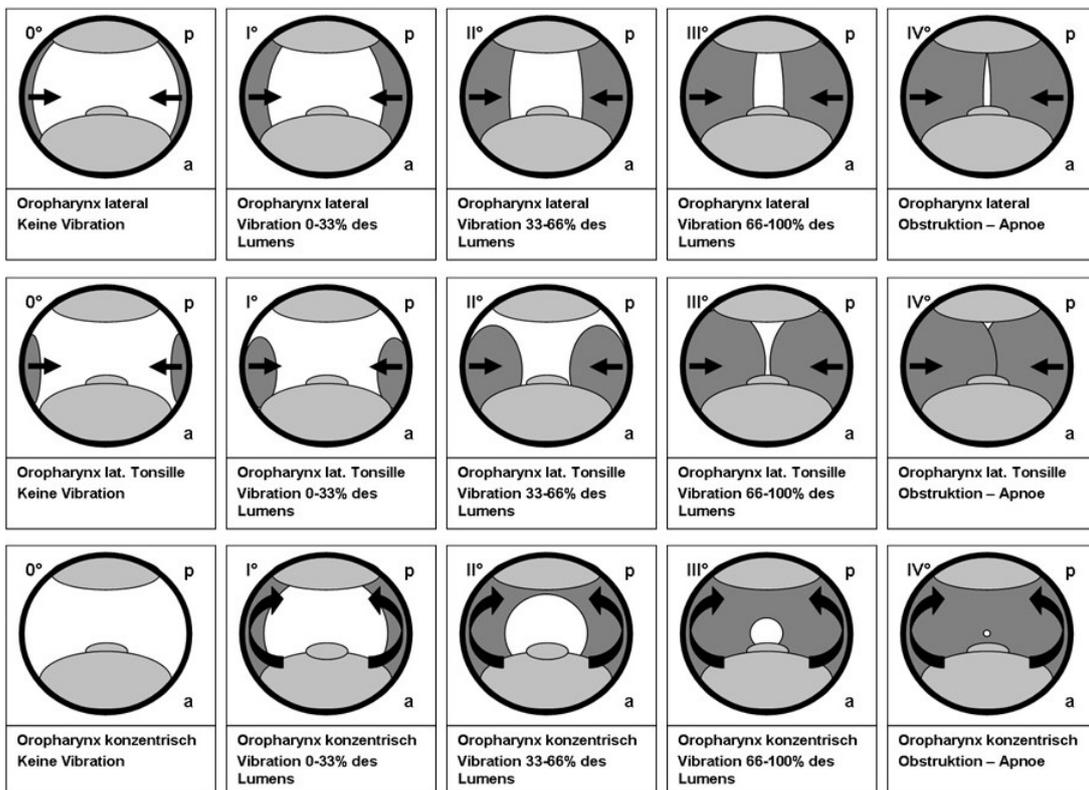
**3.1.4 Piktogramm – Klassifikation**

Mit den Piktogrammen liegt ein schematisiertes bildliches Bewertungssystem vor. Es wurde von Herzog et al. in Greifswald entwickelt und in einem Paper 2006 erstmalig veröffentlicht. Die in den wachendoskopischen Untersuchungen detektierten Schnarchereignisse wurden unter den Gesichtspunkten des Levels, der Konfiguration und des Schweregrades der Kollapse mit Hilfe der Piktogramme klassifiziert. Die in der vorliegenden Studie verwandten Piktogramm – Schemata (s. Abb. 2–5) stellen eine

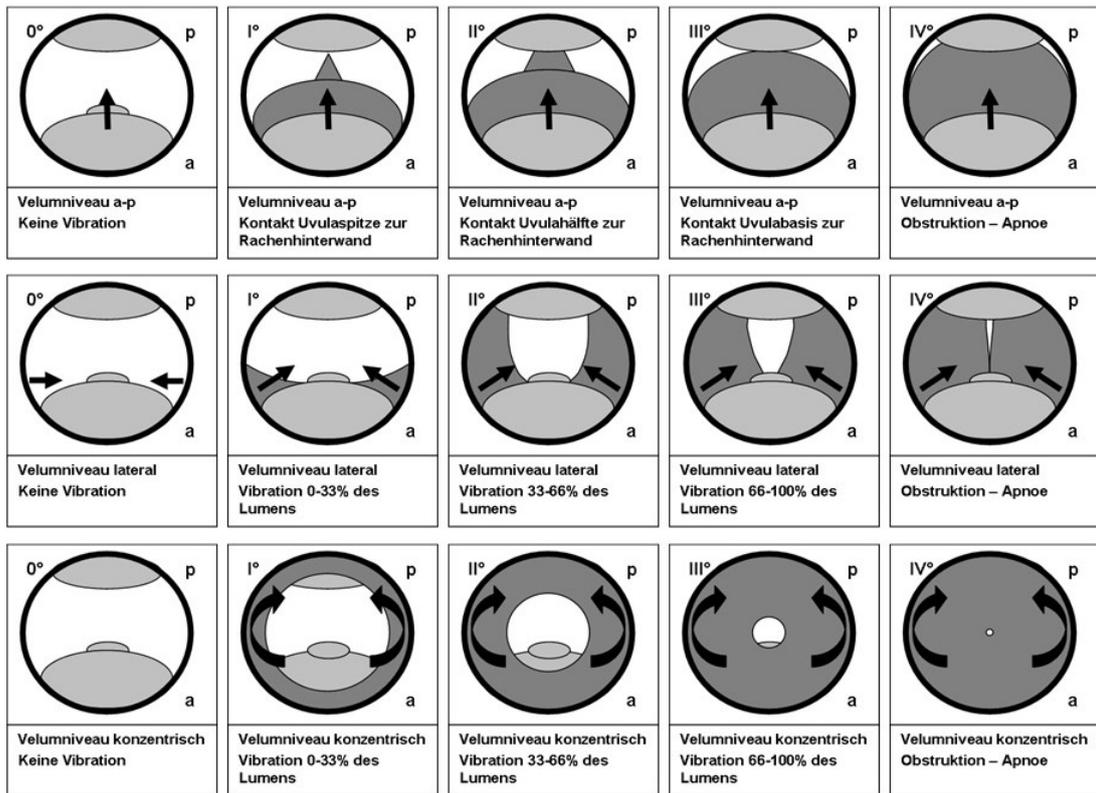
modifizierte Form der Erstentwicklung dar. Sie beziehen sich auf die Epiglottis, den Oropharynx (lateral, Tonsille, konzentrisch), das Velum (anterior – posterior, lateral, konzentrisch) und den Zungengrund (anterior – posterior, Zungengrund – Tonsille). Die Kollapsmuster werden mit den Schweregraden 0° - IV° erfasst.



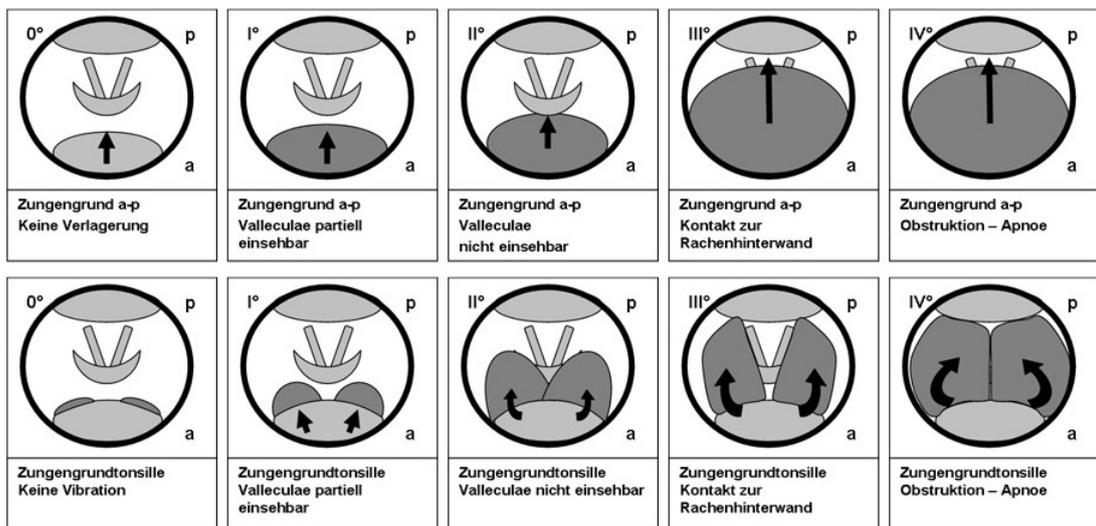
**Abbildung 2:** Piktogrammschema Epiglottislevel  
Quelle: Herzog et al., 2014



**Abbildung 3:** Piktogrammschema Oropharynxlevel  
Quelle: Herzog et al., 2014



**Abbildung 4:** Piktogrammschema Velumlevel  
Quelle: Herzog et al., 2014



**Abbildung 5:** Piktogrammschema Zungengrund  
Quelle: Herzog et al., 2014

## 3.2 Methoden

### 3.2.1 Allgemeines zur Online – Befragung

Die Befragung der Rater wurde mit dem der Martin-Luther-Universität Halle - Wittenberg zur Verfügung stehenden Online- Tool „Lime Survey“ durchgeführt.

Die Rekrutierung der Teilnehmer erfolgte über den Mailverteiler der Arbeitsgemeinschaft „Schlafmedizin“ der Deutschen Gesellschaft für Hals – Nasen – Ohrenheilkunde, Kopf– und Halschirurgie und über den Verteiler der Arbeitsgemeinschaft „Chirurgische Therapieverfahren“ der DGSM sowie über persönliche Ansprache und Ankündigungen auf Fachtagungen.

Technische Probleme bei der Übertragung der Videosequenzen über die Server der Krankenhäuser konnten durch Ausschaltung der Netzwerkkomponente Firewall behoben werden.

Die Umfrage wurde zunächst geöffnet von Anfang Oktober bis Ende November 2014 und wurde dann verlängert bis zum Januar 2015. Bis zu diesem Zeitpunkt konnten 87 Studienteilnehmer rekrutiert werden, von denen aber letztlich nur 48 Personen am vollständigen Studienprogramm teilnahmen.

### **3.2.2 Online – Fragebögen/ Klassifikationsbögen**

Neben den Eingangsfragen zum beruflichen Background der Rater (s. Kapitel 4.1. Studienteilnehmer) enthielten die Fragebögen die beschriebenen beiden Klassifizierungssysteme zur Einstufung der jeweiligen Kollapsrichtung und Ausprägung durch die Rater. Die VOTE – Level der Videosequenzen wurden vorgegeben.

Die Klassifizierungsmöglichkeiten entsprachen der VOTE – Skalierung (Konfiguration: ap, lateral, konzentrisch; Schweregrade: 0 – 1 – 2 – X -> nicht klassifizierbar) sowie der Piktogrammskalierung (Konfiguration: a-p bzw. Tonsille, lateral, konzentrisch, Schweregrade: 0° - IV°). Da die Ziffer 0 sich den Piktogrammen nicht zuordnen ließ, wurde der Einteilung 0° I° II° III° IV° hilfsweise mit den Zahlen 1 2 3 4 5 entsprochen. Auch die Nichtbeurteilbarkeit von Konfiguration und/ oder Schweregrad konnte in beiden Klassifizierungssystemen angezeigt werden.

Das Procedere gestaltete sich wie folgt:

1. Bestätigung eines Buttons, um die Videosequenz abspielen zu können
2. Präsentation des VOTE - Schemas zur Bewertung der Videosequenz: Einschätzung der Konfiguration und des Grades der Obstruktion
3. Präsentation der Piktogramme zur Bewertung der Videosequenz: Einschätzung der Konfiguration und des Grades der Obstruktion

Um die Objektivität der Klassifizierungen zu gewährleisten, wurden die Videosequenzen im Online - Verfahren randomisiert. Beispielsweise folgte auf eine Vibration/ Obstruktion Grad 0 keine Obstruktion Grad I des gleichen Levels.

**abspielen**  
Bitte klassifizieren Sie die Videosequenz:

	Grad der Obstruktion 0 = keine Vibration / Obstruktion 1 = partielle Obstruktion / Vibration 2 = komplette Obstruktion X = nicht beurteilbar	Konfiguration der Obstruktion / Vibration wenn < 0		
Anatomische Struktur	Grad der Obstruktion	anterior - posterior	lateral	konzentrisch
<b>Velum</b>				

Grad der Obstruktion / Konfiguration der Obstruktion	0	1	2	x	anterior - posterior	lateral	konzentrisch	keine Antwort
	<input type="radio"/>							

1	2	3	4	5

Velum a-p	1	2	3	4	5	keine Antwort
	<input type="radio"/>					

1	2	3	4	5

Velum lateral	1	2	3	4	5	keine Antwort
	<input type="radio"/>					

Abbildung 6: Ausschnitt eines Klassifizierungsbogens, Velum

1	2	3	4	5

Velum konzentrisch	1	2	3	4	5	keine Antwort
	<input type="radio"/>					

Abbildung 7: Ausschnitt eines Klassifizierungsbogens, Velum

### 3.3 Statistik

#### 3.3.1 Beurteilerübereinstimmung und Interraterreliabilität

In der wissenschaftlichen Literatur werden Beurteilerübereinstimmung und Interraterreliabilität fälschlicherweise häufig gleichgesetzt.

Die Interraterreliabilität wird zwar aus der Beurteilerübereinstimmung errechnet, ist aber - die Autoren Caspar und Wirtz (2002) weisen explizit darauf hin - von dieser zu unterscheiden.

In Beurteilerübereinstimmungen werden die Konkordanzen zweier oder mehrerer Beurteiler in „Prozentangaben, Streuungen oder Korrelationen“ beschrieben, die Interraterreliabilitäten hingegen geben die „Repräsentativität des Mittelwerts“ an (Bortz,

Lienert, Boehnke, 2008). Die Messmethode der Interrater - Reliabilität analysiert gegenüber der Beurteilerübereinstimmung somit die Untersuchungsmittel.

Die Interrater - Reliabilität liegt als die weitergehende Methode über dem Niveau der Beurteilerübereinstimmung und „*kennzeichnet die Zuverlässigkeit und die Genauigkeit einer Messung*“ (Caspar und Wirtz, 2002).

### 3.3.2 Statistische Aufbereitung der erhobenen Daten

Als erste Maßnahme erfolgte eine Übertragung der Daten in eine Excel – Tabelle. Daraus ließen sich Übersichten mit den absoluten und den prozentualen Häufigkeiten der durch die Experten vorgenommenen Klassifizierungen erstellen sowie eine Visualisierung der absoluten Häufigkeit mittels Graphen (Säulendiagramme) vornehmen.

Aus den damit vorliegenden rechnerisch und graphisch erfassten Häufigkeiten für beide Klassifizierungssysteme wurde der Anteil rein zufälliger Übereinstimmungen mit Hilfe von Fleiss' Kappa  $p_0 = \frac{\sum_i h_{iii}}{N}$  herausgerechnet und die Interrater-Reliabilität für beide Klassifizierungsmethoden ermittelt.

Um eine Vergleichbarkeit der beiden Methoden herbeizuführen, bedurfte es eines 95% Konfidenzintervalls. Dafür wurde auf das Computerprogramm „R“ zurückgegriffen, auf dessen Basis ein Programm entwickelt werden konnte, das es ermöglichte, das 95% Konfidenzintervall für das Fleiss' Kappa zu bestimmen.

Weitere Überlegungen zur Optimierung der Vergleichbarkeit führten dazu, die unterschiedlichen Skalenniveaus der beiden Klassifizierungssysteme VOTE mit drei Niveaus (kein Kollaps, partieller Kollaps, kompletter Kollaps) und Piktogramme mit fünf Niveaus (kein Kollaps = 0°, I°, II°, III°, IV°) in Übereinstimmung zu bringen. Das Skalenniveau für die Piktogramme wurde reduziert auf die neuen Kategorien 0°, I°/II°, III°/IV° und damit auf drei Niveaus gebracht.

## 4 Ergebnisse

### 4.1 Studienteilnehmer

Von den ursprünglich 87 Studienteilnehmern beantworteten 39 zwar die Eingangsfragen zum fachlichen Hintergrund und zur beruflichen Praxis, bewerteten aber die Videosequenzen nicht oder nicht vollständig und/ oder brachen vorzeitig ab.

Die verbleibenden 48 Studienteilnehmer bearbeiteten die gestellten Aufgaben komplett. Von Ihnen war ein zeitlicher Einsatz von 20 bis 30 Minuten zu erbringen.

Der fachliche Hintergrund der 48 Studienteilnehmer ergab folgendes Bild:

#### Zusatzbezeichnung „Schlafmedizin“ der Landesärztekammer:

11 Mediziner

#### Qualifikationsnachweis „Somnologie“ der DGSM:

5 Mediziner

#### Beschäftigung mit der Schlafmedizin in Jahren:

unter einem Jahr -> 9 Mediziner

zwischen 1-5 Jahren -> 22 Mediziner

zwischen 5-10 Jahren -> 10 Mediziner

über 10 Jahre -> 7 Mediziner

#### Erfahrung mit der Durchführung von Schlafendoskopien in Jahren:

unter einem Jahr -> 10 Mediziner

zwischen 1- 5 Jahren -> 24 Mediziner

zwischen 5-10 Jahren -> 11 Mediziner

keine Erfahrung -> 3 Mediziner

#### Eigene Erfahrung mit der Durchführung von Schlafendoskopien:

unter 10 Endoskopien -> 9 Mediziner

10-30 Endoskopien -> 14 Mediziner

30-50 Endoskopien -> 17 Mediziner

70-100 Endoskopien -> 2 Mediziner

über 100 Endoskopien -> 2 Mediziner

Vier der Mediziner führen keine Schlafendoskopien durch.

#### Dokumentation der Schlafendoskopiebefunde:

VOTE -> 15 Mediziner

Piktogramme -> 14 Mediziner

Dokumentation im OP – Bericht -> 13 Mediziner

Eigene Dokumentation -> 4 Mediziner

Sonstige -> 2 Mediziner

## 4.2 Beurteilerübereinstimmung

Für alle Level bei der VOTE - und der Piktogrammklassifikation lassen sich hohe numerische und prozentuale Übereinstimmungen in den Bereichen der Kontrollsequenzen ohne Kollaps sowie bei einigen Kollapsereignissen mit komplettem Verschluss konstatieren. Kollapsereignisse mittlerer Ausprägung weisen dagegen eine große Streuung der Bewertungen durch die Studienteilnehmer auf.

Die Ergebnisse liegen für jede Videosequenz in Tabellenform vor, in der die Anzahl der Rater und die prozentualen Übereinstimmungen auf den verschiedenen Bewertungsniveaus erfasst sind. Gleichzeitig wird damit die gesamte Übereinstimmungsmatrix geboten. Säulendiagramme veranschaulichen die numerische Verteilung der Antworten für VOTE und für die Piktogramme.

Der Übersichtlichkeit wegen werden die Beurteilerübereinstimmungen - reduziert auf die Prozentangaben der vorgenommenen Klassifizierungen - in den Kapiteln 4.3.1 – 4.3.4 zusammen mit den Interraterreliabilitäten präsentiert.

## 4.3 Interraterreliabilität

Die Interraterreliabilität dokumentiert sich in den Fleiss' Kappa – Werten in den nachfolgenden Tabellen (s. 4.3.1 – 4.3.4). Die Fleiss' Kappa Werte wurden für jedes Level errechnet und zwar sowohl für die Skalenniveaus der beiden Ratingmethoden VOTE und Piktogramme als auch für die dem Oropharynx -, dem Velum -, und dem Zungengrund – Level zugehörigen Videosequenzen (Sequenzen 6 – 24).

Für alle Videosequenzen (1 – 24) wurden darüber hinaus auf jeder Stufe der Skalenniveaus reliable Werte ( $p_1$ ) festgesetzt. Für das Oropharynx -, das Velum – und das Zungengrundlevel wurde zudem jeweils ein Gesamtwert ermittelt.

Für jeden Fleiss' – Kappa Wert wurde das entsprechende Konfidenzintervall (Vertrauensintervall) bestimmt.

Die Stärke der Einstufung der Interraterreliabilität in schwach ( $<0,20$ ), leicht ( $0,21 – 0,40$ ), mittelmäßig ( $0,41 – 0,60$ ), gut ( $0,61 – 0,80$ ) und sehr gut ( $0,81 – 1,00$ ) wurde nach Altmann (Grouven et al., 2007) vorgenommen.

Alle Interraterreliabilitäten liegen innerhalb der Intervallgrenzen. Auf Grundlage dieser Fakten kann für alle angegebenen Werte die statistische Signifikanz der Interraterreliabilität festgestellt werden. Die Interraterreliabilität der Piktogramm – Klassifizierung ändert sich in Abhängigkeit von den Skalenniveaus: Die Piktogrammklassifizierung mit angeglichenem dreistufigen Skalenniveaus verzeichnet eine höhere Interraterreliabilität als die ursprüngliche mit fünfrangigen Skalenniveaus.

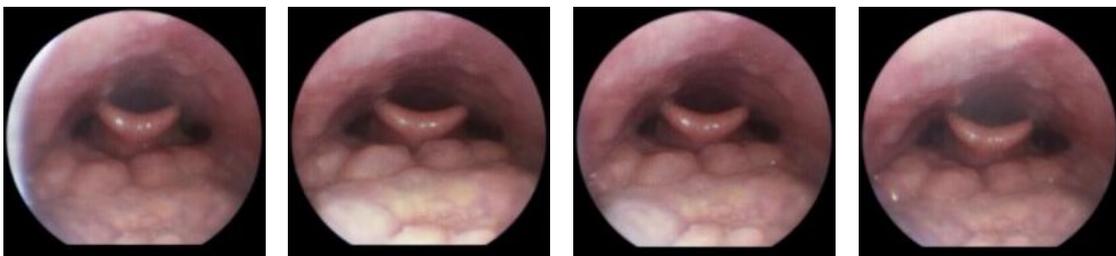
#### 4.3.1 Interraterreliabilität auf Epiglottis - Level

##### 4.3.1.1 Interraterreliabilität auf dem Gesamtlevel Epiglottis

VOTE	
Ausprägung	0,695
Muster	0,806

Piktogramme	5 – rangige Skala	3 – rangige Skala
Epiglottis	0,778	0,838

##### 4.3.1.2 Beurteilerübereinstimmung und Interraterreliabilität auf der Ebene der Sequenzen 1 – 5

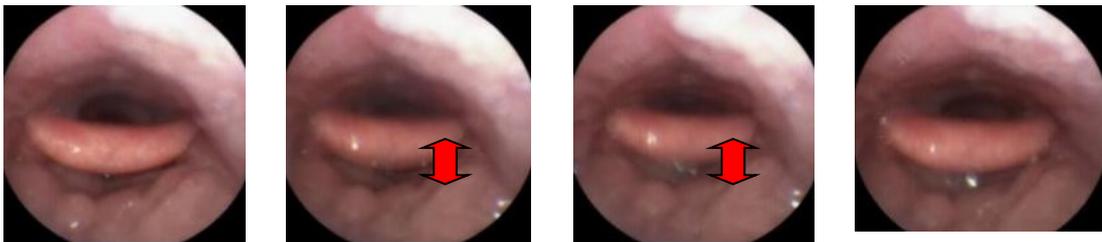


**Abbildung 8:** Zu beurteilendes Level: Epiglottis, Kollapsmuster: Kein Kollaps  
Quelle: Herzog et al., 2014

**Tabelle 1:** Sequenz 1: Epiglottis, s. Abb. 8

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/ 5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	95,83	0,918	
		partiell	4,16		
		komplett	0		
		keine Angabe	0		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/ 5- rangige Skala	3-rangige Skala
	Muster	a-p	0	0,918	
		lateral	4,16		
		konzentrisch	0		
		keine Angabe	95,83		
Piktogramme	Epiglottis	0°	100	1	1
		I°	0		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	0		

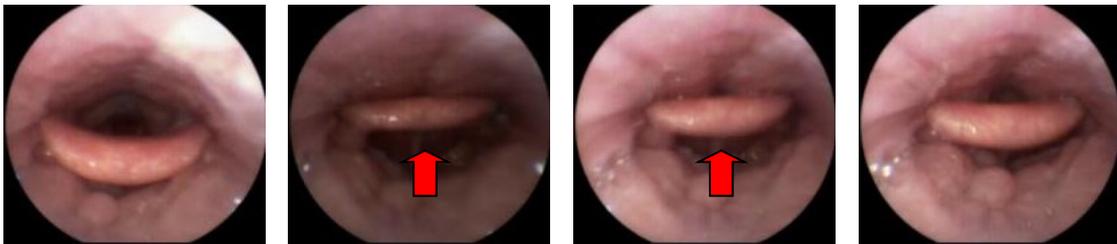


**Abbildung 9:** Zu beurteilendes Level: Epiglottis, Kollapsmuster: Vibrationen des cranialen Epiglottisrandes, keine Obstruktion. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 2:** Sequenz 4: Epiglottis, s. Abb. 9

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	14,58	0,676	
		partiell	81,25		
		komplett	0		
		keine Angabe	4,16		
	Muster	a-p	81,25	0,689	

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		lateral	0		
		konzentrisch	0		
		keine Angabe	18,75		
Piktogramme	Epiglottis	0°	0	0,918	1
		I°	95,83		
		II°	4,16		
		III°	0		
		IV°	0		
		keine Angabe	0		

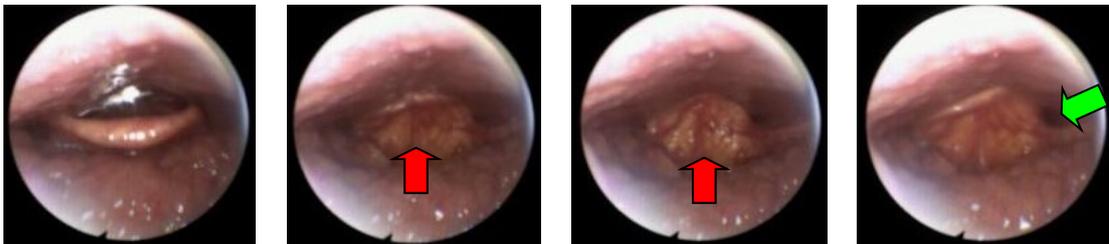


**Abbildung 10:** Zu beurteilendes Level: Epiglottis. Kollapsmuster: Vibrationen des cranialen Epiglottisrandes und der cranialen Hälfte der Epiglottis mit partieller Obstruktion. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 3:** Sequenz 5: Epiglottis, s. Abb. 1

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	2,08	0,801	
		partiell	89,58		
		komplett	2,08		
		keine Angabe	6,25		
	Muster	a-p	95,83	0,918	
		lateral	0		
		konzentrisch	0		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		keine Angabe	4,16		
Piktogramme	Epiglottis	0°	0	0,733	0,809
		I°	4,16		
		II°	85,41		
		III°	8,33		
		IV°	2,08		
		keine Angabe	0		

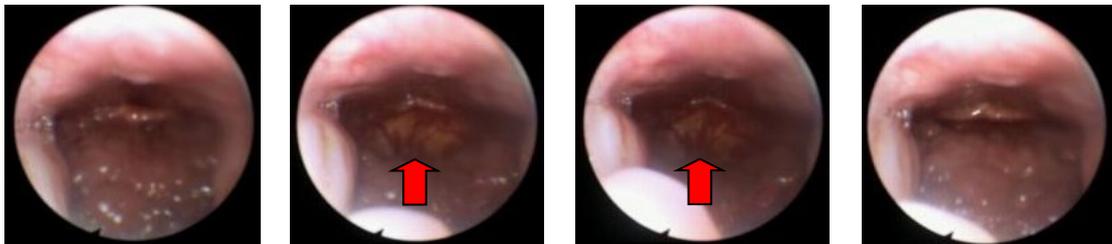


**Abbildung 11:** Zu beurteilendes Level: Epiglottis, Kollapsmuster: Partielles Epiglottisansaugphänomen mit Restlumen am linken Epiglottisrand (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 4:** Sequenz 3: Epiglottis, s. Abb. 11

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,533	
		partiell	31,25		
		komplett	66,66		
		keine Angabe	2,08		
	Muster	a-p	97,91	0,958	
		lateral	0		
		konzentrisch	0		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		keine Angabe	2,08		
Piktogramme	Epiglottis	0°	0	0,645	0,746
		I°	0		
		II°	14,58		
		III°	79,16		
		IV°	6,25		
		keine Angabe	0		



**Abbildung 12:** Zu beurteilendes Level: Epiglottis, Kollapsmuster: Epiglottisansaugphänomen, Verschluss des Larynxeinganges mit obstruktiver Apnoe. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 5:** Sequenz 2: Epiglottis, s. Abb. 12

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,777	
		partiell	0		
		komplett	87,5		
		keine Angabe	12,5		
	Muster	a-p	85,42	0,746	
		lateral	0		
		konzentrisch	0		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		keine Angabe	14,58		
Piktogramme	Epiglottis	0°	0	0,777	0,777
		I°	0		
		II°	0		
		III°	0		
		IV°	87,5		
		keine Angabe	12,5		

#### 4.3.2 Interraterreliabilität auf Oropharynx – Level

##### 4.3.2.1 Interraterreliabilität auf dem Gesamtlevel Oropharynx

VOTE	
Ausprägung	0,446
Muster	0,484

Piktogramme	5 – rangige Skala	3 – rangige Skala
Lateral 0° - IV°	0,348	0,391
Tonsil 0° - IV°	0,599	0,624
Konzentrisch 0° - IV°	0,513	0,537
Gesamt	0,512	0,543

##### 4.3.2.2 Beurteilerübereinstimmung und Interraterreliabilität auf der Ebene der Sequenzen 6 - 12

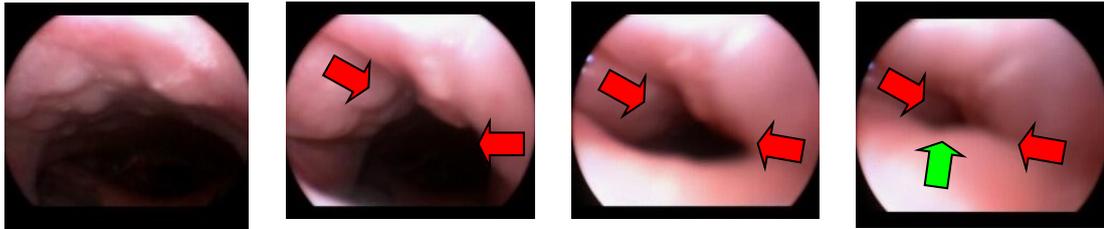


**Abbildung 13:** Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster:  
Kein Kollaps  
Quelle: Herzog et al., 2014

**Tabelle 6:** Sequenz 6: Oropharynx, s. Abb. 1

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	62,5	0,521	
		partiell	37,5		
		komplett	0		
		keine Angabe	0		
	Muster	a-p	0	0,512	
		lateral	39,58		
		konzentrisch	0		
		keine Angabe	60,41		
Piktogramme	lateral	0°	33,33	0,546	0,546
		I°	66,66		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	0		
	tonsillär	0°	95,83	0,918	0,918
		I°	2,08		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	2,08		
	konzentrisch	0°	93,75	0,879	0,879
		I°	4,16		
		II°	0		
		III°	0		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	0		
		keine Angabe	2,08		
Interraterreliabilität gesamt:				0,703	0,703

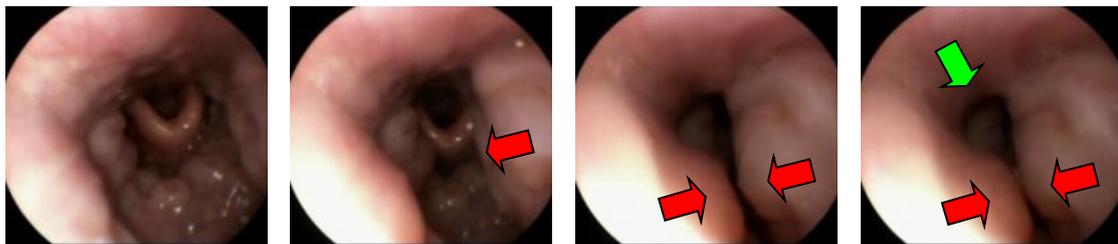


**Abbildung 14:** Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler Kollaps der Pharynxwand mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 7:** Sequenz 9: Oropharynx, s. Abb. 14.

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,469	
		partiell	64,58		
		komplett	18,75		
		keine Angabe	16,66		
	Muster	a-p	0	0,639	
		lateral	77,08		
		konzentrisch	0		
		keine Angabe	22,91		
Piktogramme	lateral	0°	12,5	0,258	0,354
		I°	12,5		
		II°	37,5		
		III°	31,25		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala		
		IV°	0	0,6	0,605		
		keine Angabe	6,25				
	tonsillär	0°	77,08				
		I°	4,16				
		II°	6,25				
		III°	4,1				
		IV°	0				
		keine Angabe	8,33				
	konzentrisch	0°	22,91			0,276	0,334
		I°	10,41				
		II°	18,75				
		III°	43,75				
		IV°	2,08				
		keine Angabe	2,08				
Interraterreliabilität gesamt:				0,283	0,331		



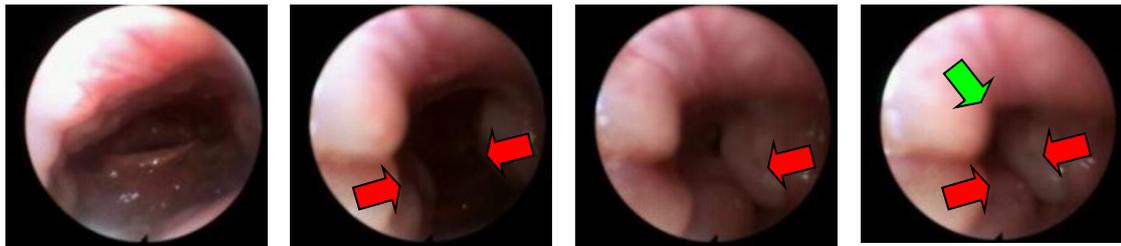
**Abbildung 15:** Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler partieller Kollaps der Tonsillen mit Restlumen. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 8:** Sequenz 10: Oropharynx, s. Abb. 15

Ratingmethoden	Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
----------------	--------------------------------	-------------------------------	--------------------

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	2,08	0,413	
		partiell	56,25		
		komplett	10,41		
		keine Angabe	31,25		
	Muster	a-p	0	0,546	
		lateral	66,66		
		konzentrisch	0		
		keine Angabe	33,33		
Piktogramme	lateral	0°	47,91	0,328	0,368
		I°	31,25		
		II°	6,25		
		III°	8,33		
		IV°	0		
		keine Angabe	6,25		
	tonsillär	0°	0	0,56	0,63
		I°	4,16		
		II°	72,91		
		III°	18,75		
		IV°	2,08		
		keine Angabe	2,08		
	konzentrisch	0°	91,66	0,84	0,84
		I°	2,08		
		II°	2,08		
		III°	0		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	0		
		keine Angabe	4,16		
Interraterreliabilität gesamt:				0,473	0,511

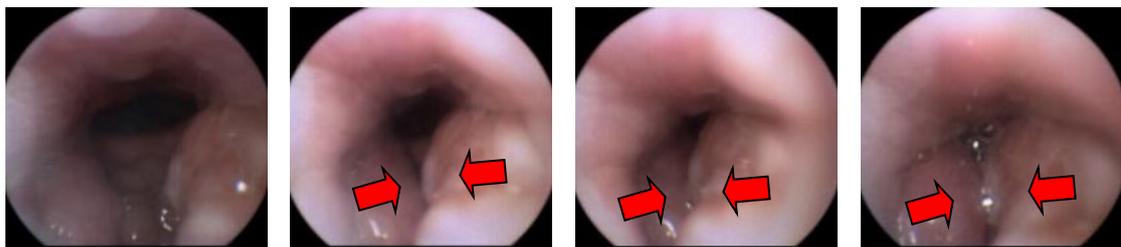


**Abbildung 16:** Zu beurteilendes Level: Cranialer Oropharynx auf Velumniveau, Kollapsmuster: Lateraler partieller Kollaps der Tonsillen mit Restlumen. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 9:** Sequenz 11: Oropharynx, s. Abb. 16

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,449	
		partiell	56,25		
		komplett	6,25		
		keine Angabe	37,5		
	Muster	a-p	0	0,521	
		lateral	62,5		
		konzentrisch	0		
		keine Angabe	37,5		
Piktogramme	lateral	0°	66,66	0,463	0,476
		I°	10,41		
		II°	6,25		
		III°	10,41		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	0	0,504	0,504
		keine Angabe	6,25		
	tonsillär	0°	0		
		I°	0		
		II°	58,33		
		III°	41,66		
		IV°	0		
		keine Angabe	0		
	konzentrisch	0°	79,16		
		I°	8,33		
		II°	0		
		III°	4,16		
		IV°	0		
		keine Angabe	8,33		
Interraterreliabilität gesamt:				0,434	0,434



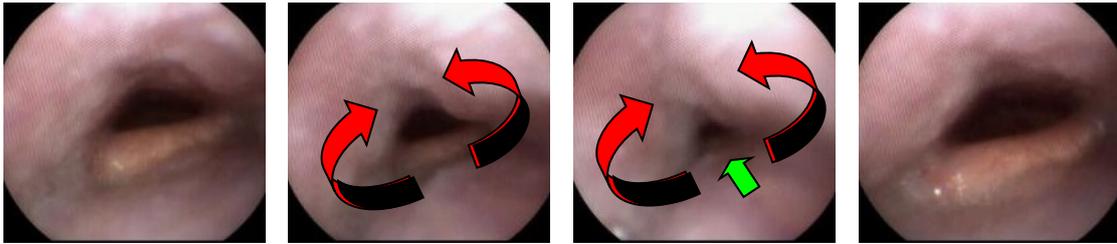
**Abbildung 17:** Zu beurteilendes Level: Craniale Oropharynx auf Velumniveau, Kollapsmuster: Lateraler Kollaps der Tonsillen mit kompletter Obstruktion. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 10:** Sequenz 12: Oropharynx, s. Abb. 17

Ratingmethoden	Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
----------------	--------------------------------	-------------------------------	--------------------

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,521	
		partiell	0		
		komplett	62,5		
		keine Angabe	37,5		
	Muster	a-p	0	0,497	
		lateral	56,25		
		konzentrisch	0		
		keine Angabe	43,75		
Piktogramme	lateral	0°	60,41	0,389	0,419
		I°	10,41		
		II°	4,16		
		III°	8,33		
		IV°	12,5		
		keine Angabe	4,16		
	tonsillär	0°	0	0,84	0,918
		I°	0		
		II°	0		
		III°	4,16		
		IV°	91,66		
		keine Angabe	4,16		
	konzentrisch	0°	81,25	0,663	0,666
		I°	2,08		
		II°	2,08		
		III°	4,16		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	2,08		
		keine Angabe	8,33		
Interraterreliabilität gesamt:				0,53	0,57

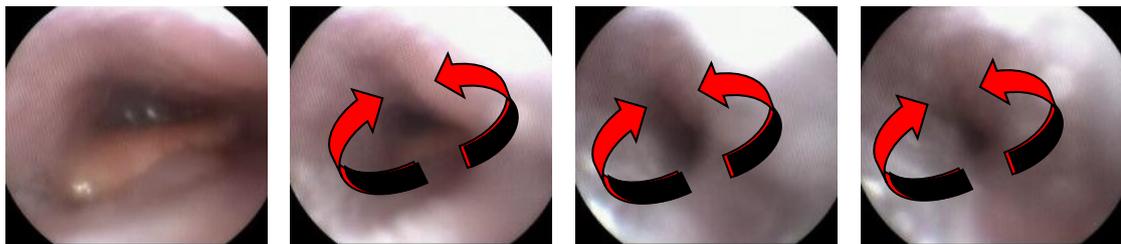


**Abbildung 18:** Zu beurteilendes Level: Caudaler Oropharynx auf Zungengrundniveau,  
Kollapsmuster: Kompletter Kollaps konzentrisch mit Obstruktion. Roter Pfeil:  
Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 11:** Sequenz 7: Oropharynx, s. Abb. 18

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,614	
		partiell	77,08		
		komplett	8,33		
		keine Angabe	14,58		
	Muster	a-p	0	0,663	
		lateral	79,16		
		konzentrisch	0		
		keine Angabe	20,83		
Piktogramme	lateral	0°	8,33	0,451	0,451
		I°	0		
		II°	63,58		
		III°	16,66		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	0	0,575	0,577
		keine Angabe	10,41		
	tonsillär	0°	75		
		I°	4,16		
		II°	2,08		
		III°	6,25		
		IV°	0		
		keine Angabe	12,5		
	konzentrisch	0°	47,91		
		I°	10,41		
		II°	18,75		
		III°	22,91		
		IV°	0		
		keine Angabe	0		
Interraterreliabilität gesamt:				0,345	0,358



**Abbildung 19:** Zu beurteilendes Level: Caudaler Oropharynx auf Zungengrundniveau,  
Kollapsmuster: Kompletter Kollaps konzentrisch mit Obstruktion. Roter Pfeil:  
Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 12:** Sequenz 8: Oropharynx, s. Abb. 19.

Ratingmethoden	Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
----------------	--------------------------------	-------------------------------	--------------------

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,405	
		partiell	39,58		
		komplett	50		
		keine Angabe	10,41		
	Muster	a-p	0	0,663	
		lateral	79,16		
		konzentrisch	0		
		keine Angabe	20,83		
Piktogramme	lateral	0°	6,25	0,269	0,405
		I°	0		
		II°	37,5		
		III°	29,16		
		IV°	22,91		
		keine Angabe	4,16		
	tonsillär	0°	66,66	0,464	0,476
		I°	12,5		
		II°	4,16		
		III°	4,16		
		IV°	2,08		
		keine Angabe	10,41		
	konzentrisch	0°	45,83	0,264	0,325
		I°	6,25		
		II°	10,41		
		III°	18,75		

Ratingmethoden			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	12,5		
		keine Angabe	6,25		
Interraterreliabilität gesamt:				0,243	0,304

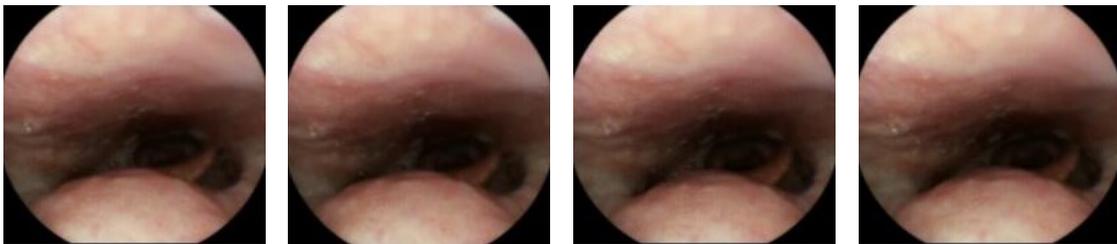
#### 4.3.3 Interraterreliabilität auf Velum – Level

##### 4.3.3.1 Interraterreliabilität auf dem Gesamtlevel Velum

VOTE	
Ausprägung	0,790
Muster	0,731

Piktogramme	5 – rangige Skala	3 – rangige Skala
a-p 0° - IV°	0,606	0,684
lateral 0° - IV°	0,479	0,504
Konzentrisch 0° - IV°	0,565	0,592
Gesamt	0,576	0,618

##### 4.3.3.2 Beurteilerübereinstimmung und Interraterreliabilität auf Ebene der Sequenzen 13 – 19



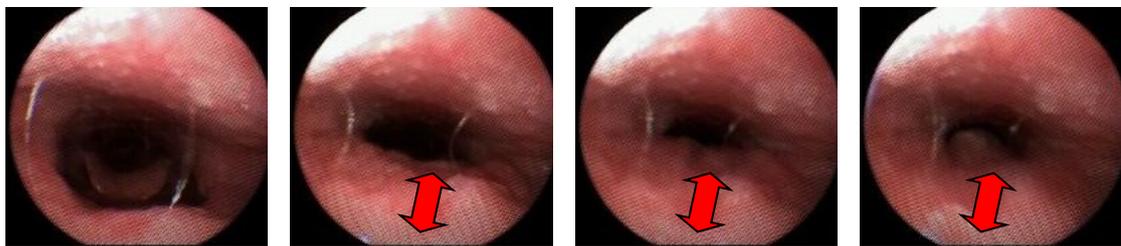
**Abbildung 20:** Zu beurteilendes Level: Velum, Kollapsmuster: Kein Kollaps  
Quelle: Herzog et al., 2014

**Tabelle 13:** Sequenz 13: Velum, s. Abb. 20  
Quelle: Eigene Darstellung

RATINGMETHODEN	Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
----------------	--------------------------------	-------------------------------	--------------------

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	95,83	0,918	
		partiell	2,03		
		komplett	0		
		keine Angabe	2,08		
	Muster	ap	2,08	0,918	
		lateral	2,08		
		konzentrisch	0		
		keine Angabe	95,83		
Piktogramme	a-p	0°	100	1	1
		I°	0		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	0		
	lateral	0°	93,75	0,879	0,879
		I°	4,16		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	2,08		
	konzentrisch	0°	97,91	0,958	0,958
		I°	0		

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	0		
		Gesamt:		0,921	0,921

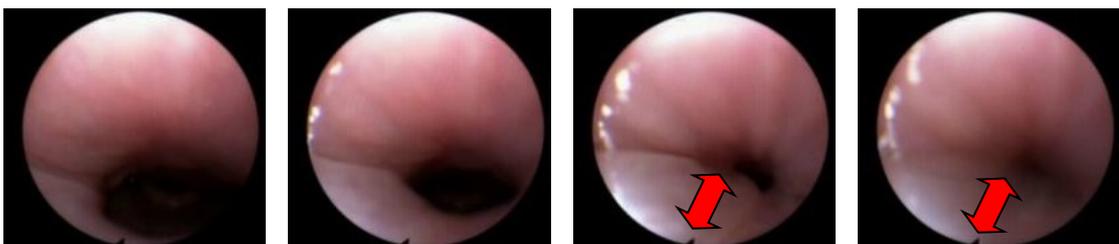


**Abbildung 21:** Zu beurteilendes Level: Velum, Kollapsmuster: Vibration des Velums anterior-posterior. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 14:** Sequenz 14: Velum, s. Abb. 21

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
Pikto- gram- me	Ausprägung	keine	0	1	
		partiell	100		
		komplett	0		
		keine Angabe	0		
	Muster	ap	97,82	0,879	
		lateral	2,17		
		konzentrisch	4,16		
		keine Angabe	0		
a-p	0°	2,08	0,736	0,918	

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		I°	87,82		
		II°	10,63		
		III°	0		
		IV°	0		
		keine Angabe	2,08		
	lateral	0°	83,33	0,698	0,703
		I°	6,25		
		II°	4,16		
		III°	0		
		IV°	0		
		keine Angabe	6,25		
	konzentrisch	0°	89,58	0,802	0,806
		I°	4,16		
		II°	4,16		
		III°	0		
		IV°	0		
		keine Angabe	2,08		
	Gesamt:		0,66	0,738	

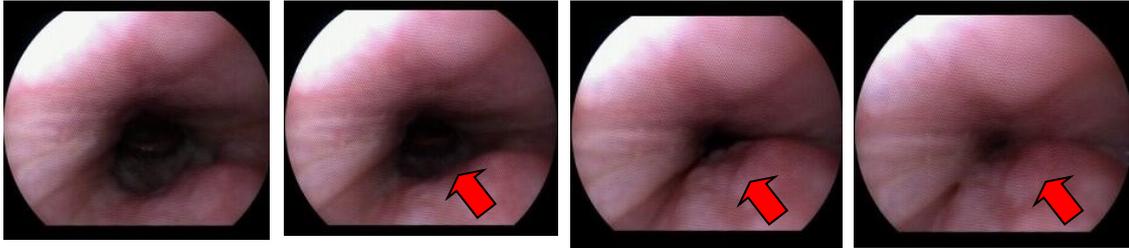


**Abbildung 22:** Zu beurteilendes Level: Velum, Kollapsmuster: Vibration des Velums anterior-posterior. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 15:** Sequenz 15: Velum, s. Abb. 22

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,879	
		partiell	93,75		
		komplett	4,16		
		keine Angabe	2,06		
	Muster	ap	87,5	0,772	
		lateral	2,06		
		konzentrisch	10,41		
		keine Angabe	0		
	Piktogramme	a-p	0°	2,06	
I°			0		
II°			83,33		
III°			10,41		
IV°			2,08		
keine Angabe			2,08		
lateral		0°	41,66	0,341	0,379
		I°	41,66		
		II°	4,16		
		III°	4,16		
		IV°	2,08		
		keine Angabe	6,25		

RATINGMETHODEN		Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
konzentrisch	0°	83,33	0,701	0,701
	I°	2,08		
	II°	0		
	III°	10,41		
	IV°	0		
	keine Angabe	4,16		
	Gesamt:		0,478	0,492

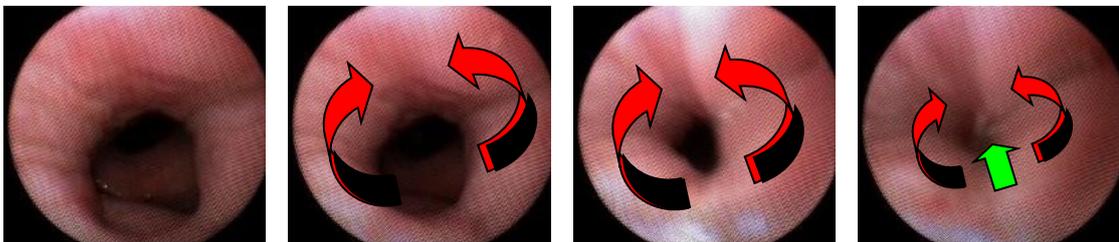


**Abbildung 23:** Zu beurteilendes Level: Velum, Kollapsmuster: Komplette Obstruktion des Velums anterior-posterior. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 16:** Sequenz 16: Velum, s. Abb. 23

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala	
VOTE	Ausprägung	keine	0	0,958		
		partiell	2,08			
		komplett	97,91			
		keine Angabe	0			
	Muster	ap	91,66	0,841		
		lateral	0			
		konzentrisch	6,25			
		keine Angabe	2,08			
Piktogramme	a-p	0°	0	0,879	0,958	
		I°	0			
		II°	0			
		III°	4,16			
		IV°	93,75			
		keine Angabe	2,08			
	lateral	0°	75	0,57		0,578
		I°	8,33			

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		II°	4,16		
		III°	2,08		
		IV°	2,08		
		keine Angabe	8,33		
	konzentrisch	0°	79,16	0,631	0,634
		I°	6,25		
		II°	0		
		III°	2,08		
		IV°	6,25		
keine Angabe		6,25			
		Gesamt:		0,599	0,634



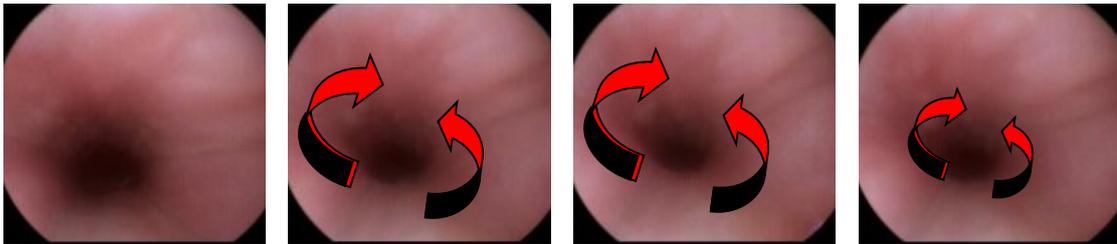
**Abbildung 24:** Zu beurteilendes Level: Velum, Kollapsmuster: Partielle konzentrische Obstruktion des Velums mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 17:** Sequenz 17: Velum, s. Abb. 24

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,57	
		partiell	72,91		
		komplett	20,83		

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		keine Angabe	6,25	0,879	
	Muster	ap	0		
		lateral	2,08		
		konzentrisch	93,75		
		keine Angabe	4,16		
Piktogramme	a-p	0°	66,66	0,461	0,476
		I°	10,41		
		II°	6,25		
		III°	4,16		
		IV°	2,08		
		keine Angabe	10,41		
	lateral	0°	60,41	0,389	0,414
		I°	11,5		
		II°	8,33		
		III°	8,33		
		IV°	2,08		
		keine Angabe	8,33		
	konzentrisch	0°	2,08	0,622	0,655
		I°	0		
		II°	18,75		
		III°	77,08		
IV°		2,08			

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		keine Angabe	0		
		Gesamt:		0,387	0,411

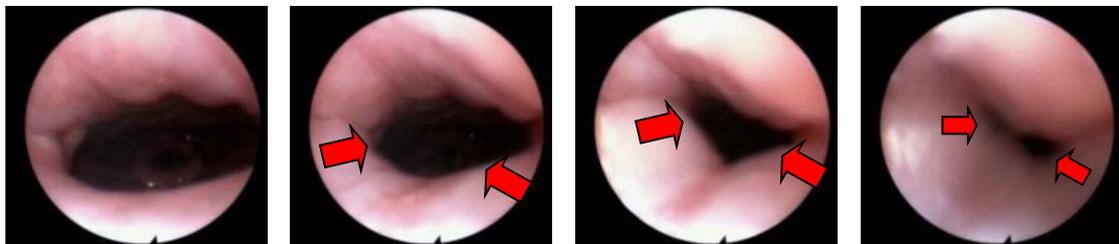


**Abbildung 25:** Zu beurteilendes Level: Velum, Kollapsmuster: Komplette konzentrische Obstruktion des Velums. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 18:** Sequenz 18: Velum, s. Abb. 25

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,496	
		partiell	60,42		
		komplett	37,5		
		keine Angabe	2,08		
	Muster	ap	2,08	0,549	
		lateral	29,17		
		konzentrisch	68,75		
		keine Angabe	0		
Piktogramme	a-p	0°	25	0,379	0,474
		I°	56,25		
		II°	8,33		
		III°	8,25		

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		IV°	0		
		keine Angabe	4,17		
	lateral	0°	2,08	0,386	0,471
		I°	8,33		
		II°	50		
		III°	37,5		
		IV°	0		
		keine Angabe	2,08		
	konzentrisch	0°	45,83	0,285	0,334
		I°	18,75		
		II°	10,42		
		III°	20,82		
		IV°	2,08		
		keine Angabe	2,08		
		Gesamt:		0,26	0,328



**Abbildung 26:** Zu beurteilendes Level: Velum, Kollapsmuster: Partielle laterale Obstruktion des Velums. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 19:** Sequenz 19: Velum, s. Abb. 26

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,88	
		partiell	93,75		
		komplett	6,25		
		keine Angabe	0		
	Muster	ap	4,16	0,493	
		lateral	62,5		
		konzentrisch	33,33		
		keine Angabe	0		
Piktogramme	a-p	0°	20,83	0,347	0,485
		I°	54,16		
		II°	12,5		
		III°	6,25		
		IV°	0		
		keine Angabe	6,25		
	lateral	0°	8,33	0,371	0,387
		I°	0		
		II°	50		
		III°	36,41		
		IV°	2,08		
		keine Angabe	4,16		
	konzentrisch	0°	22,91	0,221	0,316
		I°	25		

RATINGMETHODEN			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		II°	18,75		
		III°	29,16		
		IV°	0		
		keine Angabe	6,16		
		Gesamt:		0,228	0,3

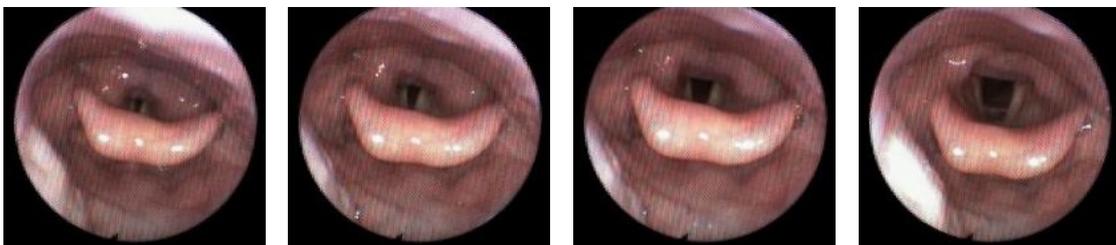
#### 4.3.4 Interraterreliabilität auf Zungengrund – Level

##### 4.3.4.1 Interraterreliabilität auf dem Gesamtlevel Zungengrund

VOTE	
Ausprägung	0,560
Muster	0,691

Piktogramme	5 – rangige Piktogrammskala	3 – rangige Piktogramme
a – p 0° - IV°	0,455	0,541
Zungengrundtonsille 0° - IV°	0,587	0,644
Gesamt	0,550	0,620

##### 4.3.4.2 Beurteilerübereinstimmung und Interraterreliabilität auf der Ebene der Sequenzen 20 – 24

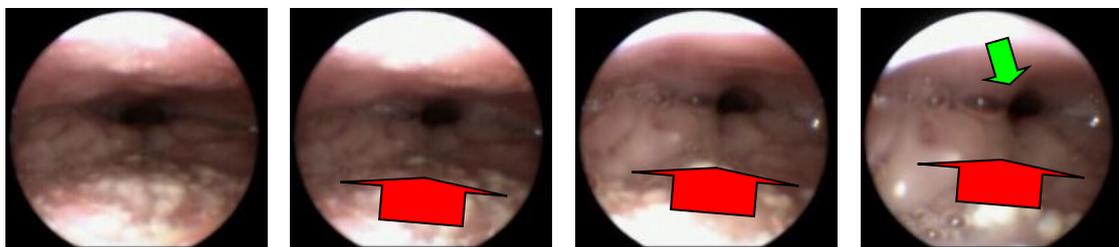


**Abbildung 27:** Zu beurteilendes Level: Zungengrund, Kollapsmuster: Kein Kollaps  
Quelle: Herzog et al., 2014

**Tabelle 20:** Sequenz 20: Zungengrund, s. Abb. 27

Ratingmethode		Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	97,91	0,958
		partiell	0	
		komplett	0	
		keine Angabe	2,08	
	Muster	a-p	2,08	0,958

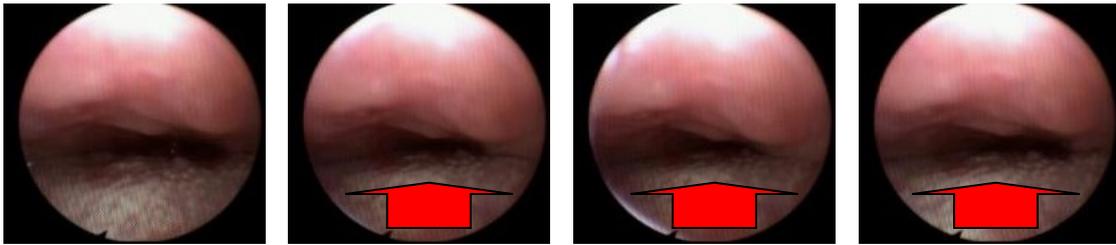
Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		lateral	0		
		konzentrisch	0		
		keine Angabe	97,91		
Piktogramm	a-p	0°	100	1	1
		I°	0		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	0		
	ZuGru - Tonsille	0°	95,833	0,918	0,918
		I°	2,08		
		II°	0		
		III°	0		
		IV°	0		
		keine Angabe	2,08		
Interraterreliabilität gesamt:				0,921	0,921



**Abbildung 28:** Zu beurteilendes Level: Zungengrund, Kollapsmuster: Partieller Kollaps anterior-posterior mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 21:** Sequenz 21: Zungengrund, s. Abb. 28

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala	
VOTE	Ausprägung	keine	6,25	0,496		
		partiell	64,58			
		komplett	29,16			
		keine Angabe	0			
	Muster	a-p	95,83	0,918		
		lateral	4,16			
		konzentrisch	0			
		keine Angabe	0			
Piktogramm	a-p	0°	0	0,448	0,512	
		I°	2,08			
		II°	37,5			
		III°	36,25			
		IV°	4,16			
		keine Angabe	0			
	ZuGru - Tonsille	0°	75	0,582		0,586
		I°	2,08			
		II°	8,33			
		III°	14,58			
		IV°	0			
		keine Angabe	0			
Interraterreliabilität gesamt:				0,342	0,374	

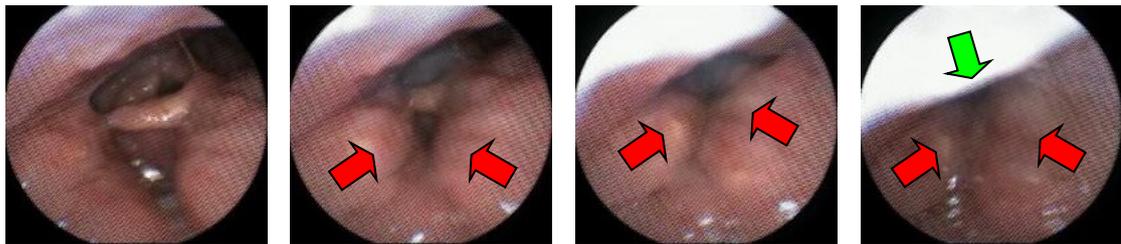


**Abbildung 29:** Zu beurteilendes Level: Zungengrund, Kollapsmuster: Komplette Obstruktion anterior-posterior ohne Restlumen. Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014.

**Tabelle 22:** Sequenz 22: Zungengrund, s. Abb. 29

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala	
VOTE	Ausprägung	keine	6,25	0,639		
		partiell	12,5			
		komplett	79,166			
		keine Angabe	2,08			
	Muster	a-p	91,66	0,844		
		lateral	8,33			
		konzentrisch	0			
		keine Angabe	0			
Piktogramm	a-p	0°	0	0,608	0,777	
		I°	2,08			
		II°	10,41			
		III°	10,41			
		IV°	77,08			
		keine Angabe	0			
	ZuGru - Tonsille	0°	72,91	0,548		0,548
		I°	10,41			
		II°	0			

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		III°	6,25		
		IV°	0		
		keine Angabe	10,41		
Interraterreliabilität gesamt:				0,403	0,492

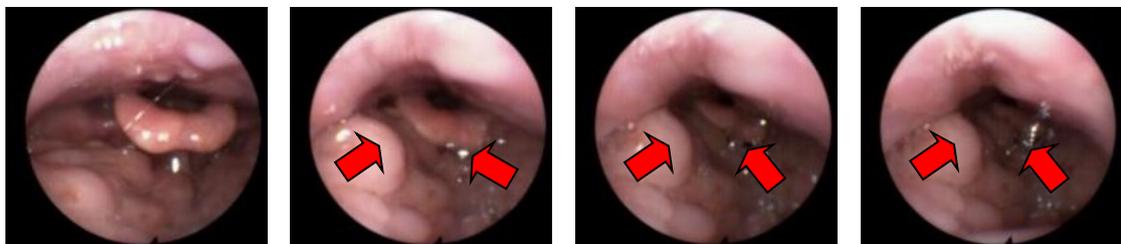


**Abbildung 30:** Zu beurteilendes Level: Zungengrund, Kollapsmuster: Partieller Kollaps anterior-posterior der hyperplastischen Zungengrundtonsille mit Restlumen (grüner Pfeil). Roter Pfeil: Verlagerungsrichtung  
Quelle: Herzog et al., 2014

**Tabelle 23:** Sequenz 23: Zungengrund, s. Abb. 30

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,405	
		partiell	10,41		
		komplett	39,58		
		keine Angabe	50		
	Muster	a-p	31,94	0,49	
		lateral	34,72		
		konzentrisch	0		
		keine Angabe	0		
Piktogramm	a-p	0°	50	0,312	0,379
		I°	20,83		

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
		II°	14,58		
		III°	10,41		
		IV°	2,08		
		keine Angabe	2,08		
	ZuGru - Tonsille	0°	0	0,576	0,74
		I°	2,08		
		II°	10,41		
		III°	75		
		IV°	20,41		
		keine Angabe	2,08		
Interraterreliabilität gesamt:				0,28	0,384



**Abbildung 31:** Zu beurteilendes Level: Zungengrund, Kollapsmuster: Komplette Obstruktion anterior-posterior der hyperplastischen Zungengrundtonsille ohne Restlumen.  
 Roter Pfeil: Verlagerungsrichtung  
 Quelle: Herzog et al., 2014

**Tabelle 24:** Sequenz 24: Zungengrund, s. Abb. 31

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
VOTE	Ausprägung	keine	0	0,577	
		partiell	4,16		
		komplett	72,91		

Ratingmethode			Beurteiler- übereinstimmung	3-rangige/5- rangige Skala	3-rangige Skala
	Muster	keine Angabe	22,91	0,617	
		a-p	63,15		
		lateral	21,05		
		konzentrisch	0		
		keine Angabe	0		
Piktogramm	a-p	0°	18,75	0,193	0,314
		I°	14,58		
		II°	16,66		
		III°	10,41		
		IV°	33,33		
		keine Angabe	6,25		
	ZuGru - Tonsille	0°	14,58	0,581	0,676
		I°	0		
		II°	4,16		
		III°	6,25		
		IV°	75		
		keine Angabe	0		
Interraterreliabilität gesamt:				0,234	0,324

## 5 Diskussion

### 5.1 Ausprägung der Interraterreliabilität bei VOTE und Piktogrammen – Vergleich der Klassifikationssysteme

Die für die Skalenniveaus der vier anatomischen Level Epiglottis, Oropharynx, Velum und Zungengrund errechneten Fleiss – Kappa Werte zeigen sehr gute, gute und mittelmäßige (einmal leichte) Interraterreliabilitäten. Die Ergebnisse der dreirangigen VOTE – Skalen fallen im Vergleich mit den Ergebnissen der fünfrangigen Piktogrammskalen etwas besser aus, bei den angepassten dreirangigen Piktogrammskalen nähern sich die Werte jedoch ziemlich stark an.

Die Interraterreliabilitäten im Einzelnen:

#### **Epiglottis – Level:**

VOTE: gut/ sehr gut

Piktogramme (V): gut

Piktogramme (III): sehr gut

#### **Oropharynx – Level:**

VOTE: mittelmäßig/ mittelmäßig

Piktogramme (V): leicht/ mittelmäßig/ mittelmäßig

Piktogramme (III): leicht/ gut/ mittelmäßig

#### **Velum – Level:**

VOTE: gut/ gut

Piktogramme (V): gut/ mittelmäßig/ mittelmäßig

Piktogramme (III): gut/ mittelmäßig/ mittelmäßig

#### **Zungengrund – Level:**

VOTE: mittelmäßig/ gut

Piktogramme (V): mittelmäßig/ mittelmäßig

Piktogramme (III): mittelmäßig/ gut

**FAZIT:** Beide Klassifizierungssysteme liefern verlässliche und vergleichbare Werte. Das ist umso erstaunlicher, als VOTE ein bekanntes, bewährtes System ist und die Piktogrammklassifizierung neu und nur von den an der Entwicklung beteiligten Insidern erprobt wurde. Besonders beeindruckend sind die Kongruenzen auf dem Epiglottislevel, aber auch auf dem Zungengrund – und Velumlevel sind sie überzeugend.

Beim Oropharynx ist es komplizierter, die Interraterreliabilitäten sind in beiden Klassifizierungssystemen geringer. Die Ursachenanalyse führt zu den Bedingungen von Studien: Für eine Studie wie der vorliegenden gilt grundsätzlich, dass die notwendige Standardisierung mit Beschränkungen verbunden ist. In der Realität können durch die Bewegung des Endoskops die Kollapsverläufe und die angrenzenden Level erfasst und somit die gesamten Außmaße der Kollapsmuster gesehen und eingeschätzt werden. In einer Studie ist es nicht möglich, dynamische Sequenzen so akzentuiert und übersichtlich zu filmen, dass sie die erforderlichen Informationen liefern. Die damit verbundene Eindimensionalität wirkt sich offenbar bei der Beurteilung des hochkomplexen Oropharynx stärker aus als bei den anderen Leveln.

Im Kontext der Oropharynx – Bewertung fallen vier sehr gute Einzelwerte der Piktogrammklassifizierung auf. Es handelt sich um unterschiedliche Kollapsmuster des cranialen Oropharynx auf Velumniveau – ohne Kollaps, partieller Kollaps der Tonsillen, lateraler Kollaps der Tonsillen mit kompletter Obstruktion. Bei solchen aus dem Rahmen fallenden Einzelwerten stellt sich die Frage, ob die Piktogramme eventuell eine intuitive Einschätzung der Kollapsereignisse ermöglichen.

## **5.2 Einfluss der Kollapslevel und Kollapsmuster auf die Ausprägung der Interraterreliabilität**

Weitere Intention dieser Arbeit war es, die Auswirkungen der Kollapslevel und einzelnen Kollapsmuster auf die Interraterreliabilität zu untersuchen.

Es erwies sich, dass das Epiglottislevel eine sehr gute Interraterreliabilität erreichte.

Bei den drei weiteren Leveln – Oropharynx, Velum, Zungengrund – wurden die Interraterreliabilitäten beider Klassifizierungssysteme zusammengerechnet, die sich je nach fünf – oder dreirangiger Piktogrammskalierung unterschiedlich präsentieren. Bei fünfrangigem Piktogrammrating ergeben sich auf dem Oropharynx -, dem Velum – und dem Zungengrundlevel Interraterreliabilitäten mittlerer Stärke. Bei der angeglichenen dreirangigen Piktogrammklassifizierung zeigen dagegen das Velum – und das Zungengrundlevel eine gute Interraterreliabilität.

Die Fleiss – Kappa Werte auf Sequenzebene geben Aufschluss über die Bewertung der Kollapsmuster (Ausprägung und Konfiguration) durch die Rater. Sehr gute und gute Interraterreliabilitäten lassen alle Sequenzen auf Epiglottis – Level erkennen (Sequenzen 1 – 5). Zudem fällt auf, dass auf allen Leveln Sequenzen ohne Kollaps (Sequenzen 1,6,13,20) sehr gute und gute Übereinstimmungswerte erzielen. Auch

Kollapsstrukturen mit Vibrationen (Sequenzen 4,5,14) weisen gute und sehr gute Interraterreliabilitäten auf.

Gute Interraterreliabilitäten lassen sich zudem bei kompletter Obstruktion der Epiglottis (Sequenz 2) und des Velums a-p (Sequenz 16) konstatieren.

Ingesamt zeigen aber stärker ausgeprägte Kollapsereignisse mit unterschiedlichen Mustern eine größere Bandbreite. Erreicht wurden mittelmäßige Interraterreliabilitäten (Sequenzen 10,11,12,15,22), mittelmäßige bis leichte Interraterreliabilitäten (Sequenz 17) und leichte Interraterreliabilitäten (Sequenzen 7,8,9,18,19,21,23,24). Eine schwache Interraterreliabilität kommt nicht vor.

### **Fazit:**

Die bei den einzelnen Kollapsmustern erreichten Interraterreliabilitäten weisen zumindest tendenziell auf die Stärken und Schwächen aktueller Diagnostik im Bereich der Kollapsstrukturen hin. Die mittleren und speziell die leichten Interraterreliabilitäten lassen Bewertungsprobleme erkennen, die sich zum Beispiel mit den Kollapsereignissen des caudalen Oropharynx auf Zungengrundniveau, mit kompletter konzentrischer und partieller lateraler Obstruktion des Velums und partieller Kollapse anterior – posterior mit Restlumen sowie kompletter Obstruktion der hyperplastischen Zungengrundtonsille des Zungengrundlevels identifizieren und benennen lassen. Damit ist ein Ansatz für eine zielgerichtete Weiterentwicklung und Forschung der Diagnoseverfahren auf dem weiten medizinischen Gebiet der Schnarchproblematik gegeben.

### **5.3 Einfluss der Professionalität und schlafmedizinischen Erfahrung der Rater auf die Ausprägung der Interraterreliabilität**

Zur Bearbeitung dieses Zielaspektes der vorliegenden Studie wurde das Klassifikationsverhalten jedes einzelnen Raters unter Berücksichtigung seiner klinischen Erfahrungen und seiner Zusatzqualifikationen analysiert und ausgewertet. Ermöglicht wurden dadurch Aussagen zur Akzeptanz und zur Praktikabilität der Klassifizierungssysteme sowie zu den Präferenzen der Rater für die eine oder die andere Methode, die sich an der Verteilung der Ergebnisse in den beiden Klassifizierungssystemen ablesen lässt.

Eine klare Trennlinie zwischen erfahrenen und nicht erfahrenen Ratern ließ sich nicht ziehen. Außer bei vier absolut unerfahrenen und vier sehr erfahrenen Ratern sind die Grenzen bei den Ratern fließend. Unter dem Gesichtspunkt der Erfahrungen mit selbst durchgeführten Schlafendoskopien wurde eine grobe Einteilung in zwei Gruppen

vorgenommen: 17 Rater mit 30 – 50 durchgeführten Schlafendoskopien und 23 Rater mit 10 bzw. 10 – 30 durchgeführten Schlafendoskopien.

Das Analyseergebnis ist interessant: In der Gruppe der per definitionem erfahrenen Rater zeigt sich zwar insgesamt eine Präferenz der herkömmlichen Methode, aber es wird auch deutlich, dass die neue Methode gut angenommen wurde.

Neun Rater, von denen drei Mediziner über die Zusatzqualifikationen „Schlafmedizin“ verfügen und ein Mediziner über die Zusatzqualifikationen „Schlafmedizin“ und „Somnologie“, erreichten die Interraterreliabilität

- mit 75% der VOTE – Klassifizierungen
- mit 83% der Piktogrammklassifizierungen.

Acht Rater erreichten die Interraterreliabilität

- mit 81,8% der VOTE – Klassifizierungen
- mit 65,5% der Piktogrammklassifizierungen.

Die nachfolgend vorgestellten Klassifikationsergebnisse der weniger erfahrenen Rater zeugen ebenfalls davon, dass beide Methoden gleich gut anwendbar sind (Ausnahme fünf Rater, s.u.) und wahrscheinlich eine „Geschmacksfrage“ darstellen, einerseits die Bevorzugung der rein sprachlichen, definitorischen Kategorien des VOTE – Systems, andererseits die Bevorzugung der schematisierten, visualisierten Formen des Piktogrammsystems. In dieser Gruppe, darunter ein Mediziner mit der Zusatzqualifikation „Schlafmedizin“ und zwei Mediziner mit den Zusatzqualifikationen „Schlafmedizin“ und „Somnologie“, gibt es die folgenden Resultate:

Elf Rater erzielten in den beiden Systemen ähnliche Ergebnisse

- mit 86,5% der VOTE – Klassifizierungen
- mit 82% der Piktogrammklassifizierungen.

Sieben Rater, darunter einer mit der Zusatzqualifikation „Schlafmedizin“, erreichten die Interraterreliabilität

- mit 77,7% der VOTE-Klassifizierungen
- mit 83,6% der Piktogrammklassifizierungen.

Fünf Rater erreichten die Interraterreliabilität

- mit 73,5% der VOTE – Klassifizierungen
- mit 47,5% der Piktogrammklassifizierungen.

Die Klassifizierungen der sehr erfahrenen Rater müssen gesondert interpretiert werden, sie liegen zu weit auseinander, um sie in einem Durchschnittswert darstellen zu können. Sie offenbaren außer in einem Fall mit sehr hoher Zustimmung zu beiden Klassifikationssystemen (Rater IV) eine größere Unabhängigkeit der Rater von diagnoseunterstützenden Klassifikationssystemen und gleichzeitig eine geringere Akzeptanz derselben.

Die Interraterreliabilität der o.a. Rater ergibt sich wie folgt:

Rater I (mehr als 100 Schlafendoskopien, Zusatz „Schlafmedizin“, Somnologie“)

- mit 66,6% der VOTE – Klassifizierungen
- mit nur 24,6% der Piktogrammklassifizierungen

Rater II (mehr als 100 Schlafendoskopien, niedergelassener Arzt)

- mit 70,8% der VOTE – Klassifikation
- mit 42,1% der Piktogrammklassifizierungen

Rater III (70 – 100 Schlafendoskopien)

- mit 72,9% der VOTE – Klassifikation
- mit 35,1% der Piktogrammklassifizierungen

Rater IV (70 – 100 Schlafendoskopien)

- mit 97,9% der VOTE – Klassifikation
- mit 85,96% der Piktogrammklassifizierungen

Im Kontrast der sehr erfahrenen mit den gänzlich unerfahrenen Ratern zeigt sich, dass die Interraterreliabilität der unerfahrenen Rater besser ausfällt, u. U. ein Hinweis darauf, dass die für die Diagnose hilfreichen Klassifizierungssysteme mehr geschätzt werden.

Die Rater dieser Gruppe erreichen die Interraterreliabilität:

Rater I (keine Schlafendoskopien, Zusatzqualifikation Schlafmedizin)

- mit 83,3% der VOTE-Klassifizierungen
- mit 85,96% der Piktogrammklassifizierungen

Rater II (keine Schlafendoskopien)

- mit 79,8% der VOTE-Klassifizierungen
- mit 85,96% der Piktogrammklassifizierungen

Rater III (keine Schlafendoskopien)

- mit 81,25% der VOTE – Klassifizierungen
- mit 78,9% der Piktogrammklassifizierungen

Rater IV (keine Schlafendoskopien)

- mit 81,25% der VOTE – Klassifizierungen
- mit 78,9% der Piktogrammklassifizierungen

**Fazit:** Für die Gesamtheit dieser Studie kann gesagt werden, dass es keinen grundsätzlichen Unterschied bei der Interraterreliabilität zwischen erfahrenen und unerfahrenen Ratern gibt. Die Vorlieben für die eine oder andere Methode lassen kein System erkennen. VOTE ist bekannt und wird a priori angewandt. Die Piktogrammmethodik ist neu, dennoch lassen sich viele Rater – erfahrene und unerfahrene – darauf ein. Eine intellektuelle Begründung gibt es dafür nicht. Die hohen Piktogrammwerte, speziell auch der unerfahrenen Rater, könnten ein Indiz dafür sein, dass die bereits geäußerte Vermutung eines intuitiven Zuganges zu den Piktogrammen zutreffend sein könnte.

#### **5.4 Vergleich mit anderen Studien zur Interraterreliabilität und anderen Formen der Reliabilität**

In den Jahren 2009 – 2014 wurden fünf Studien zur Interraterreliabilität oder auch anderen Formen der Reliabilität – Intrareliabilität, Test – Retest – Reliabilität – bei der Beurteilung von Kollapsmustern durchgeführt. Zielsetzungen, Studiendesign und Teilnehmer unterscheiden sich zum Teil jedoch erheblich von der vorliegenden Studie. Eine direkte Vergleichbarkeit mit der vorliegenden Studie ist aber bei den Bewertungssystemen und den daraus resultierenden Ergebnissen gegeben.

##### **5.4.1 Rodriguez – Bruno et al., 2009: Test – Retest reliability of drug – induced sleep endoscopy**

Ziel dieser Studie war es, die Test – Retest – Reliabilität zu ermitteln, d.h. den Grad der Beurteilerübereinstimmung der Testergebnisse bei denselben Probanden mit demselben Test bei mehreren Testungen festzustellen. Es handelt sich hierbei um eine prospektive Kohortenstudie von 32 OSAS – Patienten, die die positive Drucktherapie (CPAP) nicht toleriert hatten und nun operiert werden sollten. Sie unterzogen sich zwei MISE – Untersuchungen zu unterschiedlichen Zeitpunkten: 1. Untersuchung zur Diagnose, 2. Untersuchung am Tag der geplanten OSAS – Operation. Die Video – Aufzeichnungen der MISE wurden von zwei Ratern, einem verblindeten und einem nicht verblindeten, unter den folgenden drei Gesichtspunkten analysiert:

1. Level der Obstruktion (dichotom: Obstruktion im Gaumen oder im Hypopharynx)

2. Grad der Obstruktion
3. genaue Region des Gaumens oder des Hypopharynx, die primär für die Obstruktion verantwortlich ist.

Verglichen wurden die Bewertungen der Rater unter folgenden Aspekten:

- Patienten – Attributs – Variation zwischen den MISE – Untersuchungen
- Untersucher – Variation in der Bewertung derselben MISE – Untersuchung
- Untersucher – Variation in dem Scoring derselben Untersuchung

Zudem wurden die einzelnen Testungen in Korrelation zueinander gesetzt.

Ergebnisse:

In allen vorgenannten Bewertungskategorien wurde eine gute Test – Retest – Reliabilität des hypopharyngealen Atemwegs erzielt. Die Bewertung des Gaumens erreichte eine geringere Übereinstimmung.

#### **5.4.2 Kezirian et al., 2010: Interrater – Reliability of Drug – Induced Sleep Endoscopy**

Die Studie zur Ermittlung der Interraterreliabilität bei der Beurteilung von Kollapsmustern des Gaumens oder Hypopharynx im Rahmen der MISE wurde durchgeführt an Patienten, welche eine CPAP – Therapie nicht toleriert hatten. Es handelt sich um 108 Patienten über 18 Jahre mit einem AHI > 5/4. Beurteilt wurden die Patienten durch zwei Rater mit dem Verfahren der Test – Retest – Reliabilität, berechnet wurde die Interraterreliabilität mittels Cohens Kappa und gewichtetem Kappa. Es wurden drei Analysen unter den Gesichtspunkten durchgeführt:

1. Obstruktion ja oder nein
2. Grad der Obstruktion
3. spezifische Strukturen

Ergebnisse:

Die Interraterreliabilität für das Vorhandensein einer Obstruktion im Gaumen oder im Hypopharynx war höher als für den Schweregrad einer Obstruktion.

Die Interraterreliabilität bei der Bewertung von primären Strukturen war höher als bei der Bewertung von spezifischen Strukturen.

Die Interraterreliabilität bei der Bewertung des Hypopharynx war höher als bei der Bewertung des Gaumens.

### **5.4.3 Gillespie et al., 2013: A Trial of Drug – Induced Sleep Endoscopy in the Surgical Management of Sleep – Disordered Breathing**

Diese Studie wurde bereits im Zusammenhang mit der Darstellung der MISE angeführt (Kap. 1.3.3.3) Es handelt sich hierbei um eine longitudinale prospektive Studie mit 38 Patienten eines Krankenhauses. Sie litten unter gemäßigten schlafbezogenen Atemstörungen (AHI < 16). Eine dreimonatige CPAP – Therapie verlief erfolglos, sodass sie Aspiranten für eine UPPP waren. Voraussetzung für den Erfolg der UPPP ist eine genaue anatomische Lokalisation der Obstruktionen der oberen Atemwege; als Diagnoseinstrument wurde die MISE favorisiert. Die Kollapsmuster der o.a. Patienten wurden mit Hilfe von VOTE und dem DISE – Score - Index klassifiziert (s. Kapitel 1.3.3.4). Es wurde unterschieden nach:

- Level und Muster: Gaumen a – p, Hypopharynx, Tonsillen, Zungengrund, Epiglottis
- Schweregrad: entsprechen dem DISE – Index: 0 – 1 – 2 – 3/4

Die Studie wurde designed, um die Nullhypothese „Die MISE bringt im Vergleich mit den normalen klinischen Untersuchungen keine zusätzlichen Informationen“ zu testen. Die Studie verfolgte mehrere Intentionen:

1. die Nullhypothese zu widerlegen
2. die Reliabilität der MISE festzustellen in zweifacher Weise:
  - a) als Intrarater – Reliabilität (ein erfahrener HNO – Arzt bewertete die Kollapsereignisse während der MISE und die Aufzeichnung der gleichen Untersuchung zu einem späteren Zeitpunkt nochmals mittels des DISE – Score – Indexes)
  - b) als Interrater – Reliabilität (drei Rater bewerten die Kollapsereignisse mittels des DISE – Score – Indexes unabhängig voneinander; Kappa – Werte und Konfidenzintervall werden errechnet)
3. die Validität der MISE nachzuweisen durch Bewertung der Korrelation zwischen DISE – Scoring – Index und VOTE
4. die Erstellung eines Korrelationsmodells

Ergebnisse:

VOTE zeigt bei dieser Untersuchung eine geringere Übereinstimmung (Intra – und Interrater) als der DISE – Index – Score. Die Gründe dafür sind unklar, eine mögliche Erklärung ist, dass der DISE – Index Score Gaumen und Zungenmandeln umfasst.

#### **5.4.4 Vroegop et al., 2013: Observer Variation in Drug – Induced Sleep Endoscopy: Experienced Versus Non – Experienced Ear, Nose and Throat Surgeons**

Diese prospektive Studie thematisiert die Interraterreliabilität bei der Klassifizierung von Kollapsmustern im Vergleich von erfahrenen und unerfahrenen Bewertern. Die Studie wurde durchgeführt mit insgesamt 97 Ratern verschiedener Krankenhäuser, davon waren 90 unerfahrene und 7 erfahrene HNO – Ärzte (alle Rater verblindet). Zur Bewertung lagen sechs Videos mit MISE – Aufzeichnungen von Kollapsmustern der oberen Atemwege vor. Sie stammten von Patienten, die eine CPAP – Therapie nicht toleriert und sich deshalb einer MISE unterzogen hatten. Bewertet wurden die Kollapsmuster mit unterschiedlichen Bewertungssystemen, die an den Krankenhäusern entwickelt worden waren.

Ergebnisse: Anders als in der vorliegenden Studie erreichten die erfahrenen Rater eine höhere Interraterreliabilität als die unerfahrenen.

Insgesamt wurden für die Epiglottis, Zungengrund und anders – als in der diesseitigen Studie – auch für den Oropharynx bessere Werte erzielt als für den Hypopharynx (auch bei den unerfahrenen Ärzten). Letzteres kann damit zusammenhängen, dass der Hypopharynx in der klinischen Praxis nicht isoliert betrachtet wird. Als Ursache der niedrigen Übereinstimmung beim Hypopharynx können u.a. die unterschiedlichen, an verschiedenen Krankenhäusern etablierten Bewertungssysteme sein. Die nicht einheitlichen Scoringsysteme sind insgesamt als Mangel der Studie zu betrachten.

#### **5.4.5 Kim et al., 2014: Interrater – Reliability of Sleep Videofluoroscopy for Airway Obstruction in Obstructive Sleep Apnea**

Die Studie wurde mit 394 OSAS – Patienten, die für eine UPPP in Frage kamen, und drei Ratern des gleichen Krankenhauses durchgeführt. Ausgangspunkt dieser Studie war die Notwendigkeit einer präzisen anatomischen Diagnostik für den Erfolg der UPPP. Ziel war ein Vergleich der Schlaf – Videofluoroskopie mit anderen Untersuchungsverfahren zur Detektion des OSAS (Friedman Staging System, Müller Manöver, dynamisches CT, Schlaf – MRT...) und speziell ein Vergleich mit der MISE. Die Bewertungen durch die Rater erfassten das Vorliegen eines Kollapses (ja oder nein?), die Kollapsrichtung und die graduelle Ausprägung eines Kollapses. Erhoben wurde sowohl die Intrarater – als auch die Interraterreliabilität mittels Cohen's Kappa. Die höchste Interraterreliabilität gab es bei dem Vorhandensein einer Obstruktion des weichen Gaumens im Velopharynx – Level, die geringste Interraterreliabilität für den weichen Gaumen im Oropharynx – Level. Außer auf Oropharynx – Niveau wurden bei der Schlaf – Videofluoroskopie bessere Übereinstimmungswerte erreicht als bei der MISE, was wohl der Dynamik der Methode geschuldet ist.

Im Vergleich der zwei erfahrenen Rater zu dem unerfahrenen Rater war die Interraterreliabilität am höchsten beim weichen Gaumen auf Velopharynx – Niveau, die geringste bei den Gaumenmandeln im Oropharynx. Die schlechtere Bewertung des Oropharynx entspricht den Ergebnissen der hier gegebenen Studie: die Anatomie des Oropharynx ist sehr komplex. Sie ist wegen überlappender Bewegungen des Gaumens schwer einzuschätzen. Ausgenommen aus der Bewertung wurden laterale Bewegungen der Pharynxwand, sie konnten nicht eindeutig klassifiziert werden. Die Kritik an der Studie bezieht sich auf die geringe Anzahl der Rater, die zudem noch am gleichen Krankenhaus tätig waren.

### **5.5 Klinische Bedeutung der Studie mit kritischen Anmerkungen und Schlussfolgerungen**

Mit der vorliegenden Studie wurde erstmalig ein direkter Methodenvergleich der Klassifizierungssysteme der MISE durchgeführt.

Die Leistung des bekannten VOTE – Klassifizierungssystems besteht darin, dass es fundamentale Aspekte der MISE – Untersuchung reflektiert. Nach Einschätzung der Initiatoren handelt es sich bei VOTE um ein einfaches Bewertungssystem, welches ein tieferes Verständnis der Kollapsstrukturen erzeugt und Vergleichbarkeit schafft. Kritisch wird gesehen, dass mit der Systematisierung eine „Über – Simplifizierung“ produziert wird, die verhindern kann, dass alle Interaktionen erfasst werden.

Der Vorteil des neuen Piktogrammsystems ist seine Praktikabilität. Die Piktogramme stellen die Kollapsmuster in einer anschaulichen grafischen Gestalt dar, die somit leichter standardisiert auszuwerten und zu vergleichen sind als VOTE. Bei der fünfsträngigen Version des Piktogrammsystems besteht gegenüber der VOTE – Skalierung zudem der Vorzug, mit den vorliegenden Abstufungen die Vibrationen der extrakardialen Atemwege präziser erfassen zu können, ein Vorzug, der allerdings bei der dreirängigen Piktogrammversion relativiert wird. Die vorliegende Untersuchung stellt unter Beweis, dass die Piktogramme als Alternative zu VOTE gesehen werden können. Sie stellen ein ebenso einfaches Bewertungssystem dar, das wie das VOTE – System auch einer studienbedingten Simplifizierung ggfs. komplexerer Vorgänge unterliegt.

Die in der Summe in beiden Klassifikationssystemen annähernd gleichen, individuell jedoch sowohl bei erfahrenen als auch bei unerfahrenen Ratern unterschiedlichen Klassifikationen, weisen darauf hin, dass es für die klinische Praxis sinnvoll und hilfreich ist, unterschiedliche methodische Bewertungszugänge anzubieten. Die nicht gravierend voneinander abweichenden Gesamtergebnisse in den beiden Systemen

VOTE und Piktogramme bestätigen zugleich die Effizienz der beiden Klassifizierungssysteme.

Die Klassifizierungsergebnisse beider Systeme auf der Ebene der Kollapslevel und Kollapsmuster reflektieren die Probleme der Schnarchdiagnostik, sie machen Umfang und Ausmaß der Probleme offenbar. Aus den hohen, mittleren und niedrigen Übereinstimmungswerten der Rater lässt sich eine Stärken – Schwächen – Bilanz gegenwärtiger Schnarchdiagnostik ziehen, aus der der Weiterentwicklungs – und Forschungsbedarf abgeleitet werden kann.

## Zusammenfassung

Die medikamentös-induzierte Schlafendoskopie ist ein häufig eingesetztes Verfahren zur Lokalisation der schlafbezogenen Atmungsstörungen. Die Einschätzung der damit detektierten Kollapsereignisse ist untersucherabhängig und bedarf objektiver Kriterien. Das war der Hintergrund für eine mit dieser Studie durchgeführte prospektive Evaluation der Interraterreliabilität bei der Beurteilung von Kollapsmustern im Rahmen der MISE. 48 Rater bewerteten im Onlineverfahren 24 Videosequenzen mit Schnarchepisoden von unterschiedlichen Patienten des Universitätsklinikums Halle, die im Rahmen einer MISE – Untersuchung aquiriert worden waren. Zur Beurteilung der Kollapsmuster wurden parallel zwei verschiedene Klassifizierungssysteme vorgelegt, das etablierte VOTE – System und ein neues, noch nicht erprobtes Piktogramm – System.

Bei vorgegebenem Level (Velum, Oropharynx, Zungengrund, Epiglottis) wurden die Kollapse von den Ratern hinsichtlich ihrer Richtung und Ausprägung eingestuft. Pro Level wurde jeweils eine Sequenz ohne Kollaps und eine Sequenz mit maximaler Obstruktion angeboten. Die übrigen Sequenzen auf den vier Leveln wiesen unterschiedliche Kollapsausprägungen auf. Die Interraterreliabilität, eingestuft in schwach ( $<0,20$ ), leicht ( $0,21 - 0,40$ ), mittelmäßig ( $0,41 - 0,60$ ), gut ( $0,61-0,80$ ) und sehr gut ( $0,81 - 1,0$ ) wurde mittels Fleiss' Kappa bestimmt. Neben den spezifischen Fragen zu den Kollapsereignissen beantworteten die Rater auch persönliche Fragen zum fachlichen Hintergrund und zur klinischen Erfahrung.

Primäres Ziel der Studie war der Vergleich der beiden Klassifizierungssysteme hinsichtlich der errechneten Interraterreliabilität. Sekundäre Ziele waren die Evaluierung des Einflusses der Kollapslevel und Kollapsmuster und des Einflusses der Professionalität der Rater auf die Ausprägung der Interraterreliabilität.

Der Vergleich der Systeme erfolgte auf der Ebene der Skalenniveaus (VOTE = kein Kollaps, partieller Kollaps, kompletter Kollaps; Piktogramme =  $0^\circ - IV^\circ$  bzw. in modifizierter Form =  $0^\circ, I^\circ/II^\circ, III^\circ/IV^\circ$ ). Als Ergebnis lassen sich kongruente oder annähernd gleiche Übereinstimmungswerte in beiden Systemen verzeichnen mit sehr guter Interraterreliabilität der Skalenniveaus auf Epiglottislevel, guter bis mittelmäßiger Interraterreliabilität auf Velum – und Zungengrundlevel und mittelmäßiger bis leichter Interraterreliabilität auf Oropharynx – Level. Die schlechteren Ergebnisse des Oropharynx erklären sich einerseits durch die Komplexität des Oropharynx und andererseits durch die Methode, die wegen fehlender Dynamik die ausschließliche Betrachtung nur eines Levels erlaubte.

Als Ergebnis lässt sich konstatieren, dass mit VOTE und den Piktogrammen zwei äquivalente Klassifizierungssysteme der MISE vorliegen.

Der zweite Zielaspekt der Studie erforderte die Erhebung der Interraterreliabilität für die gesamten Kollapsstrukturen jedes einzelnen Levels und für die jeder einzelnen Sequenz. Die Übereinstimmungswerte auf Levelebene entsprechen denen der Skalenniveaus: sehr gute und gute Interraterreliabilität auf Epiglottislevel, gute bis mittelmäßige auf Velum- und Zungengrundlevel, mittelmäßige auf Oropharynxlevel. Auf der Ebene der Sequenzen erzielten die Sequenzen des Epiglottislevels, die Sequenzen ohne Kollaps und die Sequenzen mit Vibrationen auf allen Leveln sowie ein kompletter Kollaps des Velums a-p sehr gute und gute Interraterreliabilitäten. Die übrigen Sequenzen mit mittlerer und stärkerer Ausprägung wiesen hohe Streuungen auf und erreichten mittelmäßige und leichte Interraterreliabilitäten. Eine schwache Interraterreliabilität kam nicht vor. Die weitere Intention der Studie, die Ermittlung des Einflusses der schlafmedizinischen Erfahrung der Rater, ihrer klinischen Tätigkeit und Zusatzqualifikation auf die Ausprägung der Interraterreliabilität zeigte, dass die o.a. Voraussetzungen keinen nennenswerten Einfluss auf die Interraterreliabilität haben. Das gilt auch für die Präferenz der Klassifizierungssysteme bei erfahrenen und unerfahrenen Ratern. Bei den erfahrenen Ratern gab es eine geringfügige Bevorzugung des VOTE – Systems. Insgesamt lässt sich festhalten, dass sich das Klassifikationsverhalten erfahrener und unerfahrener Rater nicht grundsätzlich voneinander unterscheidet. Die neue Methode der Piktogrammklassifizierung wurde in beiden Gruppen gut angenommen, die erreichte Interraterreliabilität war in beiden Bewertungssystemen vergleichbar.

Erstmalig wurde mit der hier gegebenen Studie an einem großem Teilnehmerkollektiv ein direkter Methodenvergleich von Klassifizierungssystemen der MISE mit präziser Dokumentation von Level, Grad und Muster der Kollapsereignisse durchgeführt. Auf der Basis dieser Daten können weiterführende Studien zur Validierung und Optimierung der Klassifizierungssysteme erfolgen. Darüberhinaus werden mit den errechneten Interraterreliabilitäten die Stärken und Schwächen gegenwärtiger Diagnostik von Kollapsereignissen offengelegt und der Entwicklungs – und Forschungsbedarf im diagnostischen Bereich der Schnarchproblematik artikuliert.

## Literaturverzeichnis

- Babar-Craig H, Rajani NK, Bailey P, Kotecha BT. (2012) Validation of sleep nasendoscopy for assessment of snoring with bispectral index monitoring. *Eur Arch Otorhinolaryngol* 269: 1277-1279.
- Becker H, Ficker J, Fietze I, Geisler P, Happe S, Hornyak M, Kotterba S, Orth M, Podszus T, Raschke F, Randerath W, Rühle KH, Stiasny-Kolster K, Walther B, Wiater A, Arzt M, Schulz R, Wessendorf T. (2009) "S3-Leitlinie Nicht Erholsamer Schlaf/Schlafstörungen." Deutsche Gesellschaft für Schlafforschung und Schlafmedizin (DGSM). *Somnologie* 13: 4– 160.
- Bortz J, Lienert GA, Boehnke K: Verteilungsfreie Methoden in der Biostatistik. 3. Aufl. Springer, Berlin, 2008, S. 57-77.
- Carskadon M, Dement W. Normal human sleep: An overview. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. 4th ed. Philadelphia: Elsevier Saunders; 2005. pp. 13–23.
- Deitmer, T. (2014) "Willkommensgruß Zur 85. Jahrestagung Der Deutschen Gesellschaft Für Hals – Nasen – Ohren - Heilkunde, Kopf - und Halschirurgie E.v.." <http://www.egms.de/de/meetings/hnod2014/> (Stand: 12.06.2016).
- de Groen JH, Op den Velde W, Hovens JE, Falger PR, Schouten EG, van Duijn H. (1993) Snoring and anxiety dreams. *Sleep* 16: 35-36.
- De Vito A, Carrasco Llatas M, Vanni A, Bosi M, Braghiroli A, Campanini A, de Vries N, Hamans E, Hohenhorst W, Kotecha BT, Maurer J, Montevecchi F, Piccin O, Sorrenti G, Vanderveken OM, Vicini C. (2014) European position paper on drug-induced sedation endoscopy (DISE). *Sleep Breath* 18: 453-465.
- El Badawey MR, McKee G, Heggie N, Marshall H, Wilson JA. (2003) Predictive value of sleep nasendoscopy in the management of habitual snorers. *Ann Otol Rhinol Laryngol* 112: 40-4.
- Ficker JH, Wiest GH, Lehnert G, Meyer M, Hahn EG. (1999) Are snoring medical students at risk of failing their exams? *Sleep* 22: 205-209.
- Fleiss JL, Gurland BJ, Cooper JE. (1971) Some contributions to the measurement of psychopathology. *Br J Psychiatry* 119: 647-656.
- Georgalas C, Garas G, Hadjihannas E, Oostra A. (2010) Assessment of obstruction level and selection of patients for obstructive sleep apnoea surgery: an evidence-based approach. *J Laryngol Otol* 124: 1-9.
- Gillespie MB, Reddy RP, White DR, Discolo CM, Overdyk FJ, Nguyen SA. (2013) A trial of drug-induced sleep endoscopy in the surgical management of sleep-disordered breathing. *Laryngoscope* 123: 277-282.
- Grouven U, Bender R, Ziegler A, Lange S. (2007) [The kappa coefficient]. *Dtsch Med Wochenschr* 132 Suppl 1: e65-68.
- Guilleminault C, Quera-Salva MA, Powell NB, Riley RW. (1989) Maxillo-mandibular surgery for obstructive sleep apnoea. *Eur Respir J* 2: 604-612.

- Hamdan AL, Al-Barazi R, Kanaan A, Al-Tamimi W, Sinno S, Husari A. (2012) The effect of snoring on voice: a controlled study of 30 subjects. *Ear Nose Throat J* 91: 28-33.
- He J, Kryger MH, Zorick FJ, Conway W, Roth T. (1988) Mortality and apnea index in obstructive sleep apnea. Experience in 385 male patients. *Chest* 94: 9-14.
- Herzog M, Kellner P, Plossl S, Glien A, Rohrmeier C, Kuhnel T, Plontke S, Herzog B. (2015) Drug-induced sleep endoscopy and simulated snoring in patients with sleep-disordered breathing: agreement of anatomic changes in the upper airway. *Eur Arch Otorhinolaryngol* 272: 2541-2550.
- Herzog M, Metz T, Schmidt A, Bremert T, Venohr B, Hosemann W, Kaftan H. (2006) The prognostic value of simulated snoring in awake patients with suspected sleep-disordered breathing: introduction of a new technique of examination. *Sleep* 29: 1456-1462.
- Hewitt RJ, Dasgupta A, Singh A, Dutta C, Kotecha BT. (2009) Is sleep nasendoscopy a valuable adjunct to clinical examination in the evaluation of upper airway obstruction? *Eur Arch Otorhinolaryngol* 266: 691-697.
- Hu FB, Willett WC, Manson JE, Colditz GA, Rimm EB, Speizer FE, Hennekens CH, Stampfer MJ. (2000) Snoring and risk of cardiovascular disease in women. *J Am Coll Cardiol* 35: 308-313.
- Johal A, Battagel JM, Kotecha BT. (2005) Sleep nasendoscopy: a diagnostic tool for predicting treatment success with mandibular advancement splints in obstructive sleep apnoea. *Eur J Orthod* 27: 607-614.
- Jones TM, Ho MS, Earis JE, Swift AC, Charters P. (2006) Acoustic parameters of snoring sound to compare natural snores with snores during 'steady-state' propofol sedation. *Clin Otolaryngol* 31: 46-52.
- Kellner P, Meinhardt R, Herzog B, Christel A, Plöchl S, Bieseke L, Herzog M. (2013) "Zusammenhang von Narkosetiefe Und pharyngealer Relaxation bei der Medikamentös Induzierten Schlaf-Endoskopie bei Patienten mit schlafbezogenen Atemstörungen. *Anästhesie Intensivmedizin* 54: 3–6.
- Kezirian EJ, Hohenhorst W, de Vries N. (2011) Drug-induced sleep endoscopy: the VOTE classification. *Eur Arch Otorhinolaryngol* 268: 1233-1236.
- Kezirian EJ, White DP, Malhotra A, Ma W, McCulloch CE, Goldberg AN. (2010) Interrater reliability of drug-induced sleep endoscopy. *Arch Otolaryngol Head Neck Surg* 136: 393-397.
- Kim DK, Lee WH, Lee CH, Rhee CS, Kim JW. (2014) Interrater reliability of sleep videofluoroscopy for airway obstruction in obstructive sleep apnea. *Laryngoscope* 124: 1267-1271.
- Kotecha BT, Hannan SA, Khalil HM, Georgalas C, Bailey P. (2007) Sleep nasendoscopy: a 10-year retrospective audit study. *Eur Arch Otorhinolaryngol* 264: 1361-1367.

- Moser D, Anderer P, Gruber G, Parapatics S, Loretz E, Boeck M, Kloesch G, Heller E, Schmidt A, Danker-Hopfe H, Saletu B, Zeitlhofer J, Dorffner G. (2009) Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. *Sleep* 32: 139-149.
- Ohayon MM, Guilleminault C, Priest RG, Caulet M. (1997) Snoring and breathing pauses during sleep: telephone interview survey of a United Kingdom population sample. *BMJ* 314: 860-863.
- Penzel, T. (2013) „Zukunft der Schlafmedizin“. Vortrag auf der 21. Jahrestagung der Deutschen Gesellschaft für Schlaforschung und Schlafmedizin
- Pirsig W, Hörmann K, Siegert R, Verse T. (1998) "Primäres Schnarchen, Obstruktive Schlafapnoe (OSA) und Obstruktives Schnarchen." *HNO*: 729–30
- Plößl S, Herzog B, Glien A, Plontke S, Herzog M. (2016) Survey of otorhinolaryngologists on their clinical examinations performed in patients with suspected sleep-disordered breathing. *Eur Arch Otorhinolaryngol* 273: 1045-1053.
- Pringle MB, Croft CB. (1993) A grading system for patients with obstructive sleep apnoea--based on sleep nasendoscopy. *Clin Otolaryngol Allied Sci* 18: 480-484.
- Reddy KT, Ingrams DR, Pringle MB, Croft CB. (1993) Sleep disorders. Nasendoscopy before surgery for snoring. *BMJ* 306: 1477.
- Rodriguez-Bruno K, Goldberg AN, McCulloch CE, Kezirian EJ. (2009) Test-retest reliability of drug-induced sleep endoscopy. *Otolaryngol Head Neck Surg* 140: 646-651.
- Roopa M, Deepa M, Indulekha K, Mohan V. (2010) Prevalence of sleep abnormalities and their association with metabolic syndrome among Asian Indians: Chennai Urban Rural Epidemiology Study (CURES-67). *J Diabetes Sci Technol* 4: 1524-1531.
- Sateia MJ. (2014) International classification of sleep disorders-third edition: highlights and modifications. *Chest* 146: 1387-1394.
- Scher AI, Lipton RB, Stewart WF. (2003) Habitual snoring as a risk factor for chronic daily headache. *Neurology* 60: 1366-1368.
- Shrout PE, Fleiss JL. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86: 420-428.
- Steinhart H, Kuhn-Lohmann J, Gewalt K, Constantinidis J, Mertzlufft F, Iro H. (2000) Upper airway collapsibility in habitual snorers and sleep apneics: evaluation with drug-induced sleep endoscopy. *Acta Otolaryngol* 120: 990-994.
- Steinhart H, Kuhn-Lohmann JC, Gewalt K, Constantinidis J, Mertzlufft F, Petak M, Iro H. (2000) [Pharyngolaryngoscopic findings in patients with obstructive sleep apnea syndrome and primary snoring]. *HNO* 48: 917-921.
- Stuck BA, Abrams J, de la Chaux R, Dreher A, Heiser C, Hohenhorst W, Kuhnel T, Maurer JT, Pirsig W, Steffen A, Verse T. (2010) Diagnosis and treatment of snoring in adults--S1 guideline of the German Society of Otorhinolaryngology, Head and Neck Surgery. *Sleep Breath* 14: 317-321.

Ulfberg J, Carter N, Edling C. (2000) Sleep-disordered breathing and occupational accidents. *Scand J Work Environ Health* 26: 237-242.

Virkkula P, Bachour A, Hytonen M, Malmberg H, Salmi T, Maasilta P. (2005) Patient- and bed partner-reported symptoms, smoking, and nasal resistance in sleep-disordered breathing. *Chest* 128: 2176-2182.

Vroegop AV, Vanderveken OM, Wouters K, Hamans E, Dieltjens M, Michels NR, Hohenhorst W, Kezirian EJ, Kotecha BT, de Vries N, Braem MJ, Van de Heyning PH. (2013) Observer variation in drug-induced sleep endoscopy: experienced versus nonexperienced ear, nose, and throat surgeons. *Sleep* 36: 947-953.

Wirtz M, Caspar F: Beurteilerübereinstimmung und Beurteilerreliabilität. 1. Aufl. Hogrefe, Göttingen, 2002, S. 16, 23, 33

Young T. (1993) Analytic epidemiology studies of sleep disordered breathing--what explains the gender difference in sleep disordered breathing? *Sleep* 16: S1-2.

Zamarron C, Gude F, Otero Otero Y, Rodriguez-Suarez JR. (1999) Snoring and myocardial infarction: a 4-year follow-up study. *Respir Med* 93: 108-112.

## Thesen

Die Einschätzung der mit der medikamentös induzierten Schlafendoskopie (MISE) detektierten schlafbezogenen Atmungsstörungen ist untersucherabhängig, verbindliche Standards liegen nicht vor.

Die Einschätzungsergebnisse vieler unterschiedlicher Rater bei der Beurteilung von Kollapsmustern können die faktische Grundlage für die Entwicklung diagnostischer Standards bilden und diese befördern.

Mit der Vorlage eines neuen grafischen Klassifikationssystems über Piktogramme zusätzlich zu dem herkömmlichen begrifflichen VOTE-System wird ein alternativer Zugang zur Bewertung von Kollapsereignissen ermöglicht.

Die gleichzeitige Anwendung von zwei verschiedenen Klassifikationssystemen der MISE zur Einstufung von Schweregraden und Konfiguration der präsentierten Kollapsereignisse ermöglicht durch den direkten Vergleich die Identifizierung diagnostischer Stärken und Schwächen.

Im Ergebnis zeigen sich in beiden Klassifikationssystemen vergleichbare Werte:

A. Hohe Übereinstimmungswerte bei Kollapsereignissen auf allen Leveln bei den Kontrollsequenzen ohne Kollaps sowie bei den Sequenzen mit Vibrationen und teilweise bei Sequenzen mit kompletten Verschluss.

B. Hohe Streuung bei der Einschätzung von Kollapsereignissen mittlerer Ausprägung.

Mit VOTE und den Piktogrammen liegen zwei äquivalente Klassifizierungssysteme der MISE vor mit präziser Dokumentation von Level, Schweregrad und Konfiguration beispielhafter Kollapsstrukturen.

Es gibt keinen quantitativen Unterschied von erfahrenen und unerfahrenen Ratern bei der Präferenz der Methode. VOTE und Piktogramme werden gleichermaßen gut angenommen.

Die erreichten Interraterreliabilitäten- bei 24 Beispielsequenzen zwei Drittel mit sehr guter, guter und mittelmäßiger Interraterreliabilität, ein Drittel mit leichter Interraterreliabilität sind ein Indikator für den Zustand gegenwärtiger Schnarchdiagnostik. Problematische Level, Konfigurationen und Schweregradeinschätzungen werden offengelegt und benannt. Damit ist eine Basis eine gezielte Forschung und diagnostische Weiterentwicklung gegeben.

## **Selbstständigkeitserklärung**

Hiermit erkläre ich, dass ich diese Arbeit ohne unerlaubte Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

---

Halle (Saale), 15.05.2017

Clara Heidkamp

## **Erklärung über frühere Promotionsversuche**

Hiermit erkläre ich, dass ich keinen vorausgegangenen Promotionsversuch unternommen habe, und dass kein Promotionsversuch an einer anderen wissenschaftlichen Einrichtung läuft.

---

Halle (Saale), 15.05.2017

Clara Heidkamp

## Danksagung

Herrn Professor Dr. med. habil. Stefan Plontke danke ich sehr herzlich für die Möglichkeit im Universitätsklinikum für Hals-Nasen-Ohrenheilkunde, Kopf- und Halschirurgie der Martin-Luther-Universität Halle - Wittenberg promovieren zu dürfen.

Herrn PD Dr. med. habil. Michael Herzog, meinem Doktorvater und Betreuer, danke ich in besonderer Weise – für die Vergabe des Promotionsthemas, für die Bereitstellung des neuen, von Herzog et al. entwickelten Klassifikationssystems der MISE mittels Piktogrammen sowie für die Überlassung unverzichtbarer Dokumente. Dr. Herzogs Anregungen für die strukturelle und inhaltliche Gestaltung meiner Promotionsschrift und die stets motivierende Ansprache und konstruktive Kritik waren äußerst hilfreich. Dafür danke ich ihm ebenso wie für die mehrfache Durchsicht und die zeitraubenden Korrekturen meiner Arbeit.

Frau Dr. Beatrice Herzog gilt ebenfalls mein größter Dank für die unentbehrliche Hilfe bei der Statistik. Ein auf der Basis des Computerprogrammes R entwickeltes neues Programm zur Berechnung des 95%-Konfidenzintervalls machte den komplizierten Vergleich der Klassifikationssysteme VOTE und Piktogramme erst möglich.

Herrn Professor Dr. rer. nat. Torsten Rahne danke ich für das Lesen meiner Arbeit, für vorsichtige Kritik und die Beantwortung von Fragen.

Meinen Eltern, Margret und Rolf Heidkamp, danke ich für die fortwährende liebevolle Unterstützung und den verlässlichen Beistand bei der Entwicklung dieser Arbeit.