European Archives of Oto-Rhino-Laryngology (2024) 281:817–826
https://doi.org/10.1007/s00405-023-08313-1

LARYNGOLOGY

# Validity and reliability of the Group for Learning Useful and Performant Swallowing (GLUPS) tool

**Jerome R. Lechien[1,2,3,4] · Auriane Blouin[1] · Robin Baudouin[3] · Laura Bousard[1] · Alexandra Rodriguez[2] · Marie Verhasselt[2] · Gaetan Cavelier[2] · Grégoire Vialatte de Pemille[3] · Marta P. Circiu[3] · Lise Crevier-Buchman[3] · Stephane Hans[3] · Jan Vanderwegen[5] · Didier Dequanter[2]**

Received: 8 July 2023 / Accepted: 23 October 2023 / Published online: 6 December 2023


## Abstract

**Introduction** To validate the Group for Learning Useful and Performant Swallowing (GLUPS), a clinical tool dedicated to videofluoroscopy swallowing study (VFSS).

**Methods** Forty-five individuals were recruited from January 2022 to March 2023 from the Department of Otolaryngology Head and Neck Surgery of University Hospital Saint-Pierre (Brussels, Belgium). Subjects underwent VFSS, which was rated with GLUPS tool by two blinded otolaryngologists and one speech-therapist. VFSS were rated twice with GLUPS within a 7-day period to assess test–retest reliability.

**Results** Twenty-four patients and twenty-one controls completed the evaluations. The internal consistency ($\alpha = 0.745$) and the test–retest reliability ($r_s = 0.941$; $p = 0.001$) were adequate. GLUPS reported a high external validity regarding the significant correlation with the Penetration–Aspiration Scale ($r_s = 0.551$; $p = 0.001$). Internal validity was adequate, because GLUPS score was significant higher in patients compared to controls ($6.21 \pm 4.42$ *versus* $2.09 \pm 2.00$; $p = 0.001$). Interrater reliability did not report significant differences in the GLUPS sub- and total score among the independent judges. The mean GLUPS score of individuals without any evidence of VFSS abnormalities was 2.09/23 (95% CI 1.23–2.95), which supported that a GLUPS score $\geq 3.0$ is suggestive of pathological VFSS.

**Conclusions** GLUPS is a clinical instrument documenting the abnormal findings of oral and pharyngeal phases at the VFSS. GLUPS demonstrated high reliability and excellent criterion-based validity. GLUPS may be used in clinical practice for the swallowing evaluation at the VFSS.

**Keywords** Dysphagia · Otolaryngology · Head neck · Surgery · Videofluoroscopy · Swallowing · Tool · Radiograph

Jerome R. Lechien and Auriane Blouin have similarly contributed to the publication and have to be joined as co-first author.

Jan Vanderwegen and Didier Dequanter have to be joined as co-senior author.

✉ Jerome R. Lechien
jerome.lechien@umons.ac.be

[1] Division of Laryngology and Bronchoesophagology, Condorcet School of Speech Therapy, EpiCURA Hospital, Saint-Ghislain, Belgium

[2] Department of Otorhinolaryngology and Head and Neck Surgery, CHU Saint-Pierre, Brussels, Belgium

[3] Department of Otorhinolaryngology and Head and Neck Surgery, School of Medicine, Foch Hospital, UFR Simone Veil, Université Versailles Saint-Quentin-en-Yvelines (Paris Saclay University), Paris, France

[4] Department of Otorhinolaryngology and Head and Neck Surgery, Elsan Polyclinic of Poitiers, Poitiers, France

[5] Department of Speech, Language and Audiology, Thomas More University College of Applied Sciences, Antwerp, Belgium

## Introduction

Dysphagia affects up to 16% of the general population [1] and up to 30% of acutely hospitalized patients [2]. To date, Fiberoptic Endoscopic Evaluation of Swallowing (FEES) and Videofluoroscopic Swallow Studies (VFSS) are considered as the most reliable instrumental assessments for studying anatomical and physiological features of swallowing [3]. FEES should report a slight advantage over VFSS to detect aspiration, penetration, and residues but only few studies compared both approaches with validated and reliable instruments [1]. The findings resulting from FEES and VFSS are often subjective and depend on the practitioner experience [4]. A recent systematic review of instrumental assessments of dysphagia suggested that most clinical instruments for rating FEES and VFSS findings reported fair psychometric properties or incomplete assessment of the several steps of swallowing (i.e., oral, pharyngeal and esophageal) [5]. Moreover, there is no consensus of standardized criteria to evaluate swallow features at the FEES and VFSS [5].

In the present study, we aimed to validate the Group for Learning Useful and Performant Swallowing (GLUPS), a clinical tool dedicated to the analysis of videofluoroscopy swallowing study (VFSS).

## Methods

### Development of GLUPS

The Group for Learning Useful and Performant Swallowing (GLUPS) tool was developed by practitioners involved in a scientific Belgian group of swallowing, which includes otolaryngologists and speech therapists. The GLUPS content was based on expert experience and mainly consists of an evaluation of oral, and pharyngeal steps of swallowing at the VFSS (Fig. 1). GLUPS includes 3 items for liquid and semisolid or 4 items for solid oral phase, as well as 11, and 3 items related to the evaluation of pharyngeal, and esophageal steps at the lateral radiograph view, respectively (Fig. 1). Five additional items are assessed at the face radiograph view. At the exception of the mastication, all items may be evaluated during the intake of liquid, semi-solid or solid contrast solution. Each item is evaluated with a score of 1 in case of abnormality (KO or yes) or a score of 0 if there is no abnormality. The scores of oral, pharyngeal, esophageal, face radiograph stasis are rated on /3, /11, /3 and /5 when the VFSS is carried out with liquid or semi-solid contrast product. For solid contrast, GLUPS is subdivided into oral (/4), pharyngeal (/11), esophageal (/3) and face radiograph stasis (/5) scores. Thus, the GLUPS total score ranges from 0 to 22 or 0 to 23 according to the product texture. In the pharyngeal phase, an item is dedicated to the Penetration–Aspiration Scale (PAS), which is a validated score for penetration/aspiration [6]. The PAS score is not included in the total score of GLUPS but only provided as an additional score.

In addition, GLUPS includes a management/therapeutic part, in which the otolaryngologist or the speech therapist may indicate recommendations for patient diet (Fig. 1). To improve the interrater reliabilities, authors proposed a scoring system for items with a low possibility of interrater disparity. A scoring system that classified signs as "mild", "moderate", or "severe" was avoided. Authors proposed a scoring system that is as descriptive as possible ("yes" *versus* "no"). The present validation of GLUPS was conducted according to a checklist of recommendations designed to obtain valid and reliable clinical instruments (Appendix 1) [7].

### Subjects and setting

Adult patients with oropharyngeal dysphagia related to otolaryngological disorder and individuals without swallowing disorders and normal FEES were prospectively recruited from January 2022 to March 2023 from the Department of Otolaryngology Head and Neck Surgery of CHU Saint-Pierre (Brussels, Belgium). All patients complained of dysphagia at the consultation and benefited from FEES. The exclusion criteria included patients with base of tongue surgery, or severe neck trauma, pregnancy, psychiatric illness, or allergy to contrast product. In the control group, individuals reported normal FEES and they did not have swallowing disorder. The local ethics committee approved the study protocol (CHUSP, n°BE076201837630).

### Videofluoroscopic swallowing study

The VFSS was carried out with an X-ray flat panel detector system (Zexira®, Toshiba, Tokyo, Japan) regarding the revised Logemann protocol, in which the individual was in a sitting position, allowing the lateral view of upper aerodigestive tract through the swallowing process, and in face view for the stasis score and esophageal step. Subjects received 3 mL of Prontobario® (barium sulfate; Bracco Osterreich GmBH, Vienna, Austria) by spoon, which is a contrast product that can be mixed with a variable quantity of water to be solid, semi-solid or liquid. Patients chewed solids, while liquid and semi-solid solutions were kept in the oral cavity a few seconds before swallowing. If patient was unable to open the mouth voluntarily, the speech pathologist helped him/her to open the mouth and another practitioner put the liquid material into the mouth to induce a swallowing reflex.

**PATIENT I.D.**

**DATE OF VFSS**……
Physician…….
CCC-SLP……..

| | LIQUID | | SEMI-SOLID | | SOLID | |
|---|---|---|---|---|---|---|
| **I. LATERAL RADIOGRAPH** | | | | | | |
| 1. Oral Step | | | | | | |
| Chewing | | | | | OK | KO |
| Tongue propulsion | OK | KO | OK | KO | OK | KO |
| Posterior leakage | No | Yes | No | Yes | No | Yes |
| Oral stasis | No | Yes | No | Yes | No | Yes |
| 2. Pharyngeal Step | | | | | | |
| Swallowing reflex initiation | OK | KO | OK | KO | OK | KO |
| Soft palate elevation | OK | KO | OK | KO | OK | KO |
| Tongue base posterior movement | OK | KO | OK | KO | OK | KO |
| Food progression (constrictor muscle contraction) | OK | KO | OK | KO | OK | KO |
| Lowering of epiglottis | OK | KO | OK | KO | OK | KO |
| Laryngeal elevation | OK | KO | OK | KO | OK | KO |
| Apnea | OK | KO | OK | KO | OK | KO |
| Vallecula stasis | No | Yes | No | Yes | No | Yes |
| Piriform sinus stasis | No | Yes | No | Yes | No | Yes |
| Primary penetration | No | Yes | No | Yes | No | Yes |
| Secondary penetration | No | Yes | No | Yes | No | Yes |
| Rosenbek Penetration-Asp. | ……../8 | | ……/8 | | ……/8 | |
| 3. Esophageal Step | | | | | | |
| Cricopharyngeal sphincter relaxation | OK | KO | OK | KO | OK | KO |
| Esophageal motility | OK | KO | OK | KO | OK | KO |
| Esophageal stenosis | No | Yes | No | Yes | No | Yes |
| **II. FACE RADIOGRAPH** | | | | | | |
| Oral cavity stasis | No | Yes | No | Yes | No | Yes |
| Right vallecula stasis | No | Yes | No | Yes | No | Yes |
| Left vallecula stasis | No | Yes | No | Yes | No | Yes |
| Right piriform sinus stasis | No | Yes | No | Yes | No | Yes |
| Left piriform sinus stasis | No | Yes | No | Yes | No | Yes |
| Comments : …… | | | | | | |
| Scores: Oral step: | ……./3 | | ……./3 | | ……./4 | |
| Pharyngeal step: | ……./11 | | ……./11 | | ……./11 | |
| Esophageal step: | ……./3 | | ……./3 | | ……./3 | |
| Face Rx stasis: | ……./5 | | ……./5 | | ……./5 | |
| Total score: | ……/22 | | ……/22 | | ……/23 | |

| **III. PROPOSITION OF MANAGEMENT** | |
|---|---|
| Liquid | ☐ Liquid authorized - ☐ Liquids authorized with a practitioner (under control)<br>☐ Need of thickening of liquids (gel or cream)<br>☐ Interdiction of all liquids |
| Semi-liquid (=cream) | ☐ Authorized - ☐ Authorized under control - ☐ Interdiction of all semi-liquids |
| Solid | ☐ Authorized - ☐ Authorized under control - ☐ Interdiction of all solids |
| Treatment (medications) | ☐ Ok - ☐ on pill at a time - ☐ Crushing pills |

| **IV. FACILITATION POSITION** | **V. MANŒUVRE** |
|---|---|
| ☐ Anterior flexion of the head<br>☐ Right position of the head – ☐ Left position (rotation)<br>☐ Right position of the head - ☐ Left position (tilt)<br>☐ Other …….. | ☐ Susglottic<br>☐ Super-susglottic<br>☐ Repeated swallowing<br>☐ Other…….. |

**Fig. 1** The Group for Learning Useful and Performant Swallowing (GLUPS) tool

VFSS recordings were assessed by judges retrospectively in one or two sessions (for the test–retest reliability). Judges assessed VFSS without clinical information, including the disease of patient, history, or medication.

### Statistical methods

Statistical analyses were performed using the Statistical Package for the Social Sciences for Windows (SPSS version 22,0; IBM Corp, Armonk, NY, USA). A level of significance of $p < 0.05$ was used. The following psychometric properties were evaluated:

### Intra- and interrater reliabilities

Internal consistency was measured using Cronbach's alpha for all items for patients and controls. Because VFSS protocol is carried out by a physician in our country, the intrarater reliability was evaluated through test–retest blinded evaluations of VFSS by two otolaryngologists within a 7-day interval. The Spearman's rank correlation coefficient ($r_s$) was used for the test–retest reliability. The correlation coefficient was considered as low, moderate and strong for $r_s < 0.30$, 0.30–0.60, and $r_s > 0.60$, respectively.

The interrater reliability (concordance analysis) was assessed by comparing the GLUPS of 3 blinded experienced practitioners (two otolaryngologists and one speech therapist) with Kendall's W (coefficient of concordance) and Friedman test. Kendall's W was used to assess the similarity between the judges' ratings of the GLUPS. Friedman test was used to evaluate the similarity (absolute agreement) of the values of the scores given by the judges.

### Validity

Convergent validity was evaluated through an analysis of the correlation between the GLUPS and the PAS of all individuals (Spearman's rank correlation coefficient). A statistical comparison of the GLUPS items and total score of dysphagic patients and asymptomatic individuals was carried out with the Mann–Whitney *U* test to assess the internal validity.

### Normative data

The GLUPS threshold for determining the presence and absence of significant abnormalities at the VFSS was examined through a calculation of the mean and standard deviation of GLUPS of individuals without swallowing disorders at the FEES.

## Results

Forty-five subjects were recruited from the Department of Otolaryngology Head and Neck Surgery of the CHU Saint-Pierre (Brussels, Belgium). Twenty-four individuals had dysphagia and abnormal FEES. The mean age of patients was $68.2 \pm 12.0$ years. There were 13 males and 11 females. All patients underwent VFSS. The VFSS was carried out in 21 individuals without evidence of swallowing disorders at FEES, including 16 females and 5 males (mean age of $53.1 \pm 12.0$ years). The clinical features of patients are described in Table 1. The internal consistency was high according to the Cronbach's alpha analysis ($\alpha = 0.745$). The test–retest reliability of GLUPS was moderate-to-high and high for items and the total score of GLUPS ($r_s = 0.941$; $p = 0.001$), respectively (Table 2). Note that only few patients reported abnormalities at the esophageal examination, which supports the high intrarater reliability of esophageal GLUPS. The interrater reliability (concordance analysis) is reported in Table 3. Interrater reliability did not report significant differences in the GLUPS sub- and total score among the three independent judges. The judge-to-judge analysis is reported in Appendix 2. According to subscores, the interrater reliability ranged from 0.239 to 0.990. The GLUPS score was significantly correlated with the PAS ($r_s = 0.551$; $p = 0.001$), which suggests an adequate convergent validity. The comparison of GLUPS between dysphagic patients and controls is described in Table 4. The significant differences in GLUPS scores (oral, pharyngeal, total) between patients and controls suggest an adequate internal validity. The mean GLUPS score of individuals without swallowing disorder was 2.09 (95% CI 1.23–2.95), which means that a GLUPS score $\geq 3.0$ may be suggestive of pathological VFSS.

## Discussion

The use of valid and reliable clinical instruments is mandatory for the assessment of VFSS due to interindividual subjectivity [8]. In the present study, we assessed the psychometric properties of GLUPS, a clinical instrument rating the phases of swallowing at the VFSS.

The concurrent internal consistency and test–retest reliabilities of the GLUPS were both > 0.700, indicating excellent reliability. The internal consistency was not evaluated for the other clinical instruments documenting VFSS, which limits the comparison with the literature findings [5]. Indeed, Swan et al. summarized the current psychometric properties of FEES and VFSS clinical instruments,

**Table 1** Patient features

| Features | Patients ($N=24$) | Controls ($N=21$) |
|---|---|---|
| Age (mean years, SD) | 68.2 ± 12.0 | 53.1 ± 12.0 |
| *Gender* (*N*, %) | | |
| Females | 11 (46) | 16 (76) |
| Males | 13 (54) | 5 (24) |
| BMI | 22.8 ± 4.00 | 25.5 ± 5.93 |
| *VFSS indications* (*N*, %) | | |
| Idiopathic dysphagia | 12 (50) | – |
| Aspiration and head and neck cancer | 3 (12) | – |
| Idiopathic aspirations | 2 (8) | – |
| Post-radiation dysphagia | 2 (8) | – |
| Suspicion of esophageal stenosis | 1 (4) | – |
| Dysphagia and giant goiter | 1 (4) | – |
| Dysphagia and cervical arthrodesis | 1 (4) | – |
| Dementia | 1 (4) | – |
| Cervical vertebrae fracture | 1 (4) | – |
| *Comorbidities* (*N*, %) | | |
| Hypertension | 11 (46) | 4 (19) |
| Hypercholesterolemia | 5 (21) | 3 (14) |
| Smoking | 4 (17) | 2 (9) |
| Alcohol overuse | 4 (17) | 1 (5) |
| Chronic obstructive pulmonary disease | 4 (17) | 0 (0) |
| Arrythmia | 3 (12) | 0 (0) |
| Reflux | 3 (12) | 1 (5) |
| Type 2 diabetes | 3 (12) | 2 (9) |
| Gastric ulcer | 2 (8) | 0 (0) |
| Asthma | 2 (8) | 3 (14) |
| Stroke | 2 (8) | 1 (5) |
| Type 1 diabetes | 1 (4) | 1 (5) |
| Chronic pancreatitis | 1 (4) | 0 (0) |
| Chronic kidney insufficiency | 1 (4) | 1 (5) |
| Breast cancer | 0 (0) | 2 (9) |

*SD* standard deviation, *VFSS* videofluoroscopic swallowing study

and reported that the internal consistency was not evaluated for clinical instruments available in the literature [5]. The intra-rater reliability of GLUPS was comparable to other VFSS instruments, including the University California San Francisco (UCSF) standardized grading form (coefficient = 0.886–0.910) [9], or the Dynamic Imaging Grade of Swallowing Toxicity (DIGEST) instrument (coefficient = 0.82–0.84) [10]. The different statistical approach used for the validation of the modified barium swallowing impairment (MBSimp) makes difficult the comparison [11]. For the PAS, the intra-rater reliability ranged from 0.79 to 0.89 according to studies [6, 12, 13].

An important challenge in the proposition of new clinical instrument is the development of an instrument that exhibited high interrater reliability. Indeed, because some VFSS findings are nonspecific, the rating of VFSS score may be influenced by the practitioner experience. Consequently, most VFSS instruments reported poor interrater reliability [5]. The content of GLUPS and the simple scoring system of items (OK *versus* KO) was developed to be as descriptive as possible, reducing the risk of subjectivity between judges. In this study, the Friedman analysis did not report significant differences between judges in the oral, pharyngeal, esophageal and stasis scoring of GLUPS. The GLUPS interrater reliability analysis between judges reported coefficients ranging from 0.57 to 0.99, which corroborate those of some VFSS instruments, including the UCSF standardized grading form (0.890–0.958) [9], the PAS (coefficient = 0.78–0.84) [6], the DIGEST (coefficient = 0.82–0.84) [10], functional dysphagia scale (FDS) (coefficient = 0.73)

**Table 2** Test–retest reliability

| Items | $r_s$ | *p* value |
|---|---|---|
| *Lateral radiograph VFSS* | | |
| *Oral step* | | |
| Chewing | 0.699 | 0.001 |
| Tongue propulsion | 0.583 | 0.001 |
| Posterior leakage | 0.702 | 0.001 |
| Oral cavity stasis | 0.430 | 0.004 |
| Oral score | 0.822 | 0.001 |
| *Pharyngeal step* | | |
| Swallowing initiation | 0.729 | 0.001 |
| Soft palate closure | 0.990 | 0.001 |
| Base of tongue posterior movement | 0.831 | 0.001 |
| Food progression (pharyngeal muscle contraction) | 0.990 | 0.001 |
| Epiglottis movement (lowering) | 0.990 | 0.001 |
| Elevation of larynx | 0.807 | 0.001 |
| Apnea | 0.990 | 0.001 |
| Vallecula stasis | 0.598 | 0.001 |
| Piriform stasis | 0.694 | 0.001 |
| Primary penetration | 0.807 | 0.001 |
| Secondary penetration | 0.855 | 0.001 |
| Pharyngeal score | 0.979 | 0.001 |
| PAS | 0.775 | 0.001 |
| *Esophageal step* | | |
| Cricopharyngeal sphincter relaxation | 0.999 | 0.001 |
| Esophageal motility | 0.999 | 0.001 |
| Esophageal stenosis | 0.999 | 0.001 |
| Esophageal score | 0.806 | 0.001 |
| *Face radiograph VFSS* | | |
| Oral cavity stasis | 0.990 | 0.001 |
| Right vallecula stasis | 0.807 | 0.001 |
| Left vallecula stasis | 0.807 | 0.001 |
| Right piriform stasis | 0.990 | 0.001 |
| Left piriform stasis | 0.990 | 0.001 |
| Stasis score | 0.990 | 0.001 |
| GLUPS total score | 0.941 | 0.001 |

*GLUPS* Group for Learning Useful and Performant Swallowing, *PAS* Penetration–Aspiration Scale, *VFSS* videofluoroscopic swallowing study

*There was no esophageal abnormality in our cohort, supporting why intra-rater reliability of esophageal GLUPS was high

[14], while there were no available data for others (e.g., MBSImp [11]). In some European countries, the VFSS is only assessed by otolaryngologists, whereas speech therapists play a key role in the management of dysphagia (rehabilitation). Interestingly, we observed adequate interrater reliability between trained otolaryngologists (judges 1 and 2) and an experienced speech therapist, (judge 3) which may support the role of speech therapist in the assessment of VFSS. GLUPS is the only instrument that proposes several options for dysphagia management by the speech therapist, which supports the involvement of speech therapists in the VFSS assessment. Based on the GLUPS scores and features, the speech therapist may propose some swallowing facilitation techniques and may recommend certain food textures.

Convergent validity was evaluated through a correlation study between GLUPS and PAS. The choice of PAS was based on its high intra- and interrater reliabilities, which supports that PAS is a reliable clinical instrument for VFSS [6, 12]. The GLUPS convergent validity was adequate regarding the moderate but significant association with PAS. As for the internal consistency, it was difficult to compare our data with those of the other VFSS instruments, because authors did not assess the internal consistency through a similar statistical approach [5]. Moreover, the lack of assessment of internal consistency in other studies may be attributed to the lack of validated VFSS clinical instruments.

The sub- and total scores of GLUPS were significantly higher in dysphagic patients compared to controls, indicating high convergent validity. Convergent validity was not assessed in most previous studies. Indeed, the studies validating FDS [14], videofluoroscopic dysphagia scaler (VDS) [15], DIGEST [10], and MBSImp [11], did not include controls who underwent VFSS, which limits the comparison with the literature. The evaluation of VFSS in healthy individuals is, however, an important issue, because some items remain non-specific and may be found in an asymptomatic population (e.g., vallecula stasis, oral posterior leakage). In the present study, we did not find significant differences in

**Table 3** Interrater reliability

| Items | Judge 1 | Judge 2 | Judge 3 | Kendall W Coefficient | Friedman *p* value |
|---|---|---|---|---|---|
| Oral score | 0.60 ± 1.00 | 0.56 ± 0.92 | 0.73 ± 1.14 | 0.008 | NS |
| Pharyngeal score | 1.62 ± 2.02 | 1.51 ± 1.69 | 1.58 ± 1.71 | 0.005 | NS |
| Esophageal score | 0.07 ± 0.25 | 0.09 ± 0.29 | 0.16 ± 0.47 | 0.020 | NS |
| Stasis score | 0.20 ± 0.84 | 0.22 ± 0.88 | 0.22 ± 0.88 | 0.001 | NS |
| GLUPS | 2.49 ± 2.89 | 2.38 ± 2.60 | 2.69 ± 2.94 | 0.021 | NS |

*GLUPS* Group for Learning Useful and Performant Swallowing, *NS* non-significant

**Table 4** Internal validity

| Items | Patients | Controls | *p* value |
|---|---|---|---|
| *Lateral radiograph VFSS* | | | |
| *Oral step (Mean ± SD)* | | | |
| Chewing | 0.03 ± 0.14 | 0.02 ± 0.07 | NS |
| Tongue propulsion | 0.33 ± 0.33 | 0.10 ± 0.20 | 0.002 |
| Posterior leakage | 0.28 ± 0.37 | 0.12 ± 0.23 | NS |
| Oral cavity stasis | 0.24 ± 0.27 | 0.10 ± 0.21 | 0.045 |
| Oral score | 0.88 ± 0.90 | 0.33 ± 0.63 | 0.003 |
| *Pharyngeal step (Mean ± SD)* | | | |
| Swallowing initiation | 0.29 ± 0.32 | 0.17 ± 0.22 | NS |
| Soft palate closure | 0.01 ± 0.01 | 0.01 ± 0.01 | NS |
| Base of tongue posterior movement | 0.16 ± 0.23 | 0.07 ± 0.15 | NS |
| Food progression (pharyngeal muscle contraction) | 0.14 ± 0.21 | 0.01 ± 0.01 | 0.002 |
| Epiglottis movement (lowering) | 0.17 ± 0.19 | 0.04 ± 0.07 | 0.020 |
| Elevation of larynx | 0.12 ± 0.18 | 0.02 ± 0.09 | 0.005 |
| Apnea | 0.03 ± 0.10 | 0.01 ± 0.01 | NS |
| Vallecula stasis | 0.82 ± 0.33 | 0.33 ± 0.38 | 0.001 |
| Piriform stasis | 0.59 ± 0.43 | 0.07 ± 0.18 | 0.001 |
| Primary penetration | 0.08 ± 0.23 | 0.01 ± 0.01 | 0.029 |
| Secondary penetration | 0.10 ± 0.20 | 0.01 ± 0.01 | 0.008 |
| Pharyngeal score | 2.42 ± 1.80 | 0.69 ± 0.75 | 0.001 |
| PAS | 1.27 ± 1.28 | 0.66 ± 0.14 | 0.001 |
| *Esophageal step (Mean ± SD)* | | | |
| Cricopharyngeal sphincter relaxation | 0.09 ± 0.26 | 0.03 ± 0.09 | NS |
| Esophageal motility | 0.03 ± 0.07 | 0.01 ± 0.04 | NS |
| Esophageal stenosis | 0.07 ± 0.20 | 0.01 ± 0.01 | NS |
| Esophageal score | 0.19 ± 0.38 | 0.03 ± 0.10 | NS |
| *Face radiograph VFSS (Mean ± SD* | | | |
| Oral cavity stasis | 0.08 ± 0.20 | 0.01 ± 0.01 | NS |
| Right vallecula stasis | 0.75 ± 0.43 | 0.01 ± 0.01 | NS |
| Left vallecula stasis | 0.75 ± 0.43 | 0.01 ± 0.01 | NS |
| Right piriform stasis | 0.91 ± 0.14 | 0.01 ± 0.01 | NS |
| Left piriform stasis | 0.91 ± 0.14 | 0.01 ± 0.01 | NS |
| Stasis score | 0.43 ± 1.17 | 0.01 ± 0.01 | 0.05 |
| GLUPS total score | 6.21 ± 4.42 | 2.09 ± 2.00 | 0.001 |

*GLUPS* Group for Learning Useful and Performant Swallowing, *NS* non-significant, *PAS* Penetration–Aspiration Scale, *VFSS* videofluoroscopic swallowing study

*There was no esophageal abnormality in our cohort of patients, supporting why there was no significant difference between patients and controls in the esophageal assessment

esophageal and stasis items between dysphagic patients and controls, which may be attributed to the non-specificity of some findings (stasis) or the profile of included dysphagic patients. Indeed, most patients were recruited from the department of otolaryngology, and reported oropharyngeal dysphagia rather than esophageal dysphagia. This specific population profile needs to be considered in future studies using GLUPS for the evaluation of swallowing steps. GLUPS should require future validation studies in patients with esophageal dysphagia to evaluate the convergent validity.

The primary limitation of the study was the focus on patients with oropharyngeal dysphagia, which may bias the validation study of esophageal phase of GLUPS. This limitation was found in other studies investigating the psychometric properties of other clinical instruments. Precisely, VDS and FDS were validated in stroke patients only [14, 15], whereas Hutchenson et al. validated the DIGEST scale in cancer patients [10]. Only Martin-Harris et al. included outpatients with heterogeneous medical and surgical diagnoses, including esophageal and oropharyngeal dysphagia [11]. The low number of patients included in this study is another limitation. However, the analysis of 24 and 21 consecutive VFSS recordings is a process that requires an important concentration. Thus, the judge may rate the VFSS recordings differently according to its level of fatigue and concentration. In that way, we decided to include a maximum of 50 subjects to limit the evaluation bias related to the fatigue of the judge. The consideration of a control group with individuals without evidence of abnormalities at the FEES is the main strength of our study. This group allowed the analysis of additional psychometric properties, such as convergent validity and the determination a threshold score for normal *versus* abnormal examinations.

## Conclusion

GLUPS is a clinical instrument documenting the abnormal findings of oral and pharyngeal phases at the VFSS. GLUPS demonstrates high reliability and excellent criterion-based validity. GLUPS may be used in clinical practice for the swallowing evaluation at the VFSS.

## Appendix 1

See below Table 5.

**Table 5** Definition of the measurement properties of signs of instruments analyzed in the study

| Domain | Definition |
|---|---|
| **Conceptual model** | |
| *Construct definition* | It provides a rationale for and description of the concepts and target population that a measure is intended to assess |
| *Target population* | |
| *Expected subscales* | |
| **Content validity** | It refers to evidence that an instrument is appropriate for its intended use. Items and conceptual domains must be relevant to the targeted population |
| *Content expert involved* | The instrument's development of signs must include direct input from experts. There should be a clear description of the process by which included signs were derived |
| *Description of item development* | The items described in the instrument must reflect the most common signs encountered in the disease |
| **Reliability** | The degree to which scores are free from random (measurement) error |
| *Internal consistency reliability* | Extent to which items within each domain are interrelated.* |
| *Test–retest reliability* | Stability of scores over time when no change is expected in the concept of interest.* |
| *Concordance* | The degree of agreement among raters |
| **Construct validity** | It refers to whether an instrument measures intended theoretic constructs or traits and directly affects the appropriateness of the measurement-based inferences |
| *Responsiveness to change* | The extent to which an instrument detects meaningful changes over time that have occurred after baseline.*** |
| *Convergent validity* | The degree to which the sign score correlates with other instruments measuring the same construct or with related clinical indicators.** |
| *Known-groups validity* | The extent to which the instrument can discriminate between groups that are known to differ on the variables being measured.*** |
| **Interpretability and scoring** | The degree to which the meaning of the scores can be easily understood |
| *Plan for scoring measure* | A description of how to score the measure should be provided (sum, algorithm) |
| *Plan for missing data* | A prespecified plan for managing missing responses can mitigate the risk of bias resulting from the necessity to exclude cases with missing data |
| *Scaling described* | The process of distributing the full range of respondents' possible scores with respect to the measured attribute |

*=consistent: > 0.70 for group-level comparisons; **= < 0.30 = low correlation; 0.30 to 0.60 = moderate correlation; > 0.60 = strong correlation (Pearson or Spearman analysis); ***=large change: > 0.80; moderate change: 0.50–0.79; small change: 0.2–0.49

## Appendix 2

See below Table 6.

**Table 6** Interrater reliability from judge to judge

| | Judge 1 | Judge 2 | Coefficient | *p* value |
|---|---|---|---|---|
| Oral score | 0.60 ± 1.00 | 0.56 ± 0.92 | 0.567 | 0.001 |
| Pharyngeal score | 1.62 ± 2.02 | 1.51 ± 1.69 | 0.657 | 0.001 |
| Esophageal score | 0.07 ± 0.25 | 0.09 ± 0.29 | 0.285 | NS |
| Stasis score | 0.20 ± 0.84 | 0.22 ± 0.88 | 0.664 | 0.001 |
| GLUPS | 2.49 ± 2.89 | 2.38 ± 2.60 | 0.750 | 0.001 |
| | Judge 1 | Judge 3 | Coefficient | *p* value |
| Oral score | 0.60 ± 1.00 | 0.73 ± 1.14 | 0.492 | 0.001 |
| Pharyngeal score | 1.62 ± 2.02 | 1.58 ± 1.71 | 0.540 | 0.001 |
| Esophageal score | 0.07 ± 0.25 | 0.16 ± 0.47 | 0.239 | NS |
| Stasis score | 0.20 ± 0.84 | 0.22 ± 0.88 | 0.664 | 0.001 |
| GLUPS | 2.49 ± 2.89 | 2.69 ± 2.94 | 0.571 | 0.001 |
| | Judge 2 | Judge 3 | Coefficient | *p* value |
| Oral score | 0.56 ± 0.92 | 0.73 ± 1.14 | 0.494 | 0.001 |
| Pharyngeal score | 1.51 ± 1.69 | 1.58 ± 1.71 | 0.586 | 0.001 |
| Esophageal score | 0.09 ± 0.29 | 0.16 ± 0.47 | 0.647 | 0.001 |
| Stasis score | 0.22 ± 0.88 | 0.22 ± 0.88 | 0.990 | 0.001 |
| GLUPS | 2.38 ± 2.60 | 2.69 ± 2.94 | 0.607 | 0.001 |

*GLUPS* Group for Learning Useful and Performant Swallowing, *NS* non-significant

**Author contributions** JRL: design, acquisition of data, data analysis and interpretation, drafting, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. AB: design, acquisition of data, data analysis and interpretation, drafting, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. DD: design, acquisition of data, data analysis and interpretation, drafting, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JV: design, acquisition of data, data analysis and interpretation, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. SH: design, acquisition of data, data analysis and interpretation, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. LB: data analysis and interpretation, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. GVP: data analysis and interpretation, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. MV: data analysis and interpretation, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. GC: data analysis and interpretation, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. RB: data analysis and interpretation, drafting, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. AR: data analysis and interpretation, drafting, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. MC: development of GLUPS, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. LCB: development of GLUPS, final approval, and accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** None.

**Data availability** Data are available on request,

## Declarations

**Conflict of interest** The authors have no conflicts of interest.

**Informed consent** Patients consented to the study.

## References

1. Giraldo-Cadavid LF, Leal-Leaño LR, Leon-Basantes GA, Bastidas AR, Garcia R, Ovalle S, Abondano-Garavito JE (2017) Accuracy of endoscopic and videofluoroscopic evaluations of swallowing for oropharyngeal dysphagia. Laryngoscope 127(9):2002–2010. https://doi.org/10.1002/lary.26419
2. Cichero JA, Heaton S, Bassett L (2009) Triaging dysphagia: nurse screening for dysphagia in an acute hospital. J Clin Nurs 18(11):1649–1659
3. Heijnen BJ, Böhringer S, Speyer R (2020) Prediction of aspiration in dysphagia using logistic regression: oral intake and self-evaluation. Eur Arch Otorhinolaryngol 277(1):197–205. https://doi.org/10.1007/s00405-019-05687-z
4. Chan MKK, Cheng PPJ (2017) A comparison of two rating protocols for videofluoroscopic swallowing studyby inexperienced judges. Int J Speech Lang Pathol 19(6):562–568. https://doi.org/10.1080/17549507.2016.1254681
5. Swan K, Speyer R, Scharitzer M, Farneti D, Brown T, Cordier R (2022) A Visuoperceptual Measure for Videofluoroscopic swallow studies (VMV): a pilot study of validity and reliability in adults with dysphagia. J Clin Med 11(3):724. https://doi.org/10.3390/jcm11030724
6. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL (1996) A penetration-aspiration scale. Dysphagia 11(2):93–98. https://doi.org/10.1007/BF00417897
7. Lechien JR, Schindler A, De Marrez LG et al (2019) Instruments evaluating the clinical findings of laryngopharyngeal reflux: a systematic review. Laryngoscope 129(3):720–736
8. Mahoney AS, Khalifa Y, Lucatorto E, Sejdić E, Coyle JL (2022) Cervical vertebral height approximates hyoid displacement in videofluoroscopic images of healthy adults. Dysphagia 37(6):1689–1696. https://doi.org/10.1007/s00455-022-10414-8
9. Curtis JA, Laus J, Yung KC, Courey MS (2016) Static endoscopic evaluation of swallowing: transoral endoscopy during clinical swallow evaluations. Laryngoscope 126(10):2291–2294
10. Hutcheson KA, Barrow MP, Barringer DA, Knott JK, Lin HY, Weber RS et al (2017) Dynamic Imaging Grade of Swallowing Toxicity (DIGEST): scale development and validation. Cancer 123(1):62–70
11. Martin-Harris B, Brodsky MB, Michel Y, Castell DO, Schleicher M, Sandidge J et al (2008) MBS measurement tool for swallow impairment—MBSImp: establishing a standard. Dysphagia 23(4):392–405
12. Kelly AM, Drinnan MJ, Leslie P (2007) Assessing penetration and aspiration: how do videofluoroscopy and fiberoptic endoscopic evaluation of swallowing compare? Laryngoscope 117(10):1723–1727. https://doi.org/10.1097/MLG.0b013e318123ee6a

13. Daniels SK, Schroeder MF, McClain M, Corey DM (2006) Dysphagia in stroke: development of a standard method to examine swallowing recovery. J Rehabil Res Dev 43(3):347
14. Han TR, Paik N-J, Park JW (2001) Quantifying swallowing function after stroke: a functional dysphagia scale based on videofluoroscopic studies. Arch Phys Med Rehabil 82(5):677–682
15. Han TR, Paik N-J, Park J-W, Kwon BS (2008) The prediction of pe sistent dysphagia beyond six months after stroke. Dysphagia 23(1):59–64