

Research Article

Structural Validity, Internal Consistency, and Rater Reliability of the Modified Barium Swallow Impairment Profile: Breaking Ground on a 52,726-Patient, Clinical Data Set

Alex E. Clain,^a  Munirah Alkhuwaiter,^a Kate Davidson,^b  and Bonnie Martin-Harris^{a,c} ^aRoxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL ^bDepartment of Otolaryngology-Head & Neck Surgery, Medical University of South Carolina, Charleston ^cEdward Hines, Jr. VA Hospital, Hines, IL

ARTICLE INFO

Article History:

Received October 13, 2021

Revision received December 24, 2021

Accepted January 6, 2022

Editor-in-Chief: Bharath Chandrasekaran

Editor: Michelle Ciucci

https://doi.org/10.1044/2022_JSLHR-21-00554

ABSTRACT

Purpose: The purpose of this study was to extend the assessment of the psychometric properties of the Modified Barium Swallow Impairment Profile (MBSImP). Here, we re-examined structural validity and internal consistency using a large clinical-registry data set and formally examined rater reliability in a smaller data set.

Method: This study consists of a retrospective structural validity and internal consistency analysis of MBSImP using a large data set ($N = 52,726$) drawn from the MBSImP Swallowing Data Registry and a prospective study of the interrater and intrarater reliability of a subset of studies ($N = 50$) rated by four MBSImP-trained speech-language pathologists. Structural validity was assessed via exploratory factor analysis. Internal consistency was measured using Cronbach's alpha for each of the multicomponent MBSImP domains, namely, the oral and pharyngeal domains. Interrater reliability and intrarater reliability were measured using the intraclass correlation coefficient (ICC).

Results: The exploratory factor analysis showed a two-factor solution with factors precisely corresponding to the scale's oral and pharyngeal domains, consistent with findings from the initial study. Component 17, that is, the esophageal domain, did not load onto either factor. Internal consistency was good for both the oral and pharyngeal domains ($\alpha_{\text{oral}} = .81$, $\alpha_{\text{pharyngeal}} = .87$). Interrater reliability was found to be good with $\text{ICC}_{\text{interrater}} = .78$ (95% confidence interval [CI]: .76, .80). Intrarater reliability was good for each rater, $\text{ICC}_{\text{Rater-1}} = .82$ (95% CI [.77, .86]), $\text{ICC}_{\text{Rater-2}} = .83$ (95% CI [.79, .87]), $\text{ICC}_{\text{Rater-3}} = .87$ (95% CI [.83, .90]), and $\text{ICC}_{\text{Rater-4}} = .87$ (95% CI [.83, .90]).

Conclusions: This study leverages a large-scale, clinical data set to provide strong, generalizable evidence that the MBSImP assessment method has excellent structural validity and internal consistency. In addition, the results show that MBSImP-trained speech-language pathologists can demonstrate good interrater and intrarater reliability.

Correspondence to Alex E. Clain: alexanderclain2022@u.northwestern.edu. **Disclosure:** Bonnie Martin-Harris and Kate Davidson report grants from National Institutes of Health/National Institute on Deafness and Other Communication Disorders, during the conduct of the study, and royalties from Northern Speech Services through the Medical University of South Carolina Foundation for Research Development, outside the submitted work. In addition, Bonnie Martin-Harris has a patent US 62/710,324 issued. The other authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.

The modified barium swallow study (MBSS) using videofluoroscopic (VFS) imaging represents the most commonly used diagnostic approach for instrumental assessment of swallowing impairment (Martino et al., 2004; Pettigrew & O'Toole 2007; Rumbach et al., 2018). MBSSs are assessed by speech-language pathologists (SLPs) and radiologists who use visual inspection and their clinical judgment to make informed decisions about the nature of

each patient's swallowing impairment and about an appropriate management plan. The use of visual inspection and clinical judgment means that there is a perceptual and subjective aspect to clinician appraisals of MBSSs. Thus, it is of critical importance to verify that MBSS assessments satisfy the requirements of subjective tests, that is, that they satisfy the required psychometric properties. In particular, it is important to test that MBSS assessments are measuring what they intend to measure, referred to as psychometric validity, and that there is agreement across and within clinicians on scoring, referred to as psychometric rater reliability (Lambert et al., 2002; Souza et al., 2017). Testing these psychometric properties becomes especially important when clinicians are using standardized assessment methods. Standardized dysphagia assessments provide several benefits to clinicians and researchers including increased ease of comparing impairment across patients and of tracking impairment trajectories within patients. However, for these benefits to come to fruition, assessment methods must be held to a high standard that includes rigorous testing of their psychometric properties. This study represents a continuation of the effort to examine the psychometric properties of the standardized Modified Barium Swallow Impairment Profile (MBSImP™; Martin-Harris et al., 2008), specifically its structural validity, internal consistency, and rater reliability.

MBSImP includes a standardized scale for identifying the nature and severity of physiologic swallowing impairment based on clinician ratings of VFS images obtained during MBSS and has had broad, global uptake in the field of dysphagia (Northern Speech Services, 2016). MBSImP includes 17 physiological components rated across 12 swallowing tasks of varying bolus consistencies, bolus volumes, and presentation methods of standardized, customized, commercially prepared, and stable contrast materials (see Method for further details). The 17 components are each rated on an ordinal scale, where the scale ranges from no impairment (score of 0) to severe impairment (score of 2, 3, or 4, depending on the component). Both clinically and in research studies, the most common scoring method is referred to as Overall Impression (OI) scoring (e.g., Arrese et al., 2017, 2019; Clark et al., 2020; Gullung et al., 2012; Hutcheson et al., 2012; Im et al., 2019; Im & Ko, 2020; Martin-Harris et al., 2015; O'Rourke et al., 2017; Wilmskoetter et al., 2018; Xinou et al., 2018). An OI score is determined separately for each MBSImP component by selecting the worst performance (highest score on an ordinal scale) across all the swallowing tasks, which for a single MBSS, results in one OI score for each of the 17 MBSImP components. As OI scoring represents the primary method for using MBSImP both clinically and in research studies, it is these OI scores that will form the basis of all analyses in this article.

Structural Validity

Structural validity refers to the extent to which statistical groupings of items adhere to hypothesized groupings of items (Lambert et al., 2002). For MBSImP, there are 17 components, hypothesized to form three domains: Components 1–6 form the oral domain, Components 7–16 form the pharyngeal domain, and Component 17 forms the esophageal domain. The structural validity of the multicomponent domains, namely, the oral and pharyngeal domains, was tested and confirmed by factor analysis in a study by Martin-Harris et al. (2008) in a cohort of 300 patients. In this study, we use a much larger sample ($N = 52,726$) of clinical patient visits to provide a large-scale test of the structural validity of MBSImP.

Internal Consistency

Internal consistency represents the degree to which sets of items cohere together to measure single constructs. No prior studies have directly investigated the internal consistency of the MBSImP scale. As mentioned above, Martin-Harris et al. (2008) conducted a factor analysis in which MBSImP scores showed a two-factor structure, congruent with the two hypothesized, multicomponent domains of the scale: oral and pharyngeal. The separation of the components into their hypothesized domains tells us that the domains are valid but does not tell us how well the components of each domain cohere together to form a unified construct. It is common practice in both clinical settings and research to *assume* that each domain forms a unified construct and to then sum up the component scores of each domain into “total” scores, that is, an oral total and a pharyngeal total, to characterize the “overall impairment” of each domain (e.g., Arrese et al., 2017; Clark et al., 2020; Hutcheson et al., 2012; Im & Ko, 2020; O'Rourke et al., 2017). This assumption of a unified construct and use of total scores relies on each domain having good internal consistency. Thus, to test whether this assumption holds, here we assess the internal consistency of each of the multicomponent domains of MBSImP.

Interrater and Intrarater Reliability

Rater reliability, in general, is the degree to which there is agreement between repeated ratings of the same MBSS, where interrater reliability represents the agreement between different raters rating the same study and intrarater reliability represents within-rater agreement when rerating a previously rated study. In a clinical setting, good interrater reliability and intrarater reliability of an assessment method provide assurance that one can trust both another clinician's ratings and one's own previous ratings to be similar to what one would judge. In a research setting, poor reliability of an assessment method results in false reductions in the maximum possible effect size and sensitivity/specificity when examining the

relationship between that assessment method and any outcome measure (Lachin, 2004). Thus, good rater reliability is of critical importance to clinicians and researchers and should be known for any assessment method. At the time of writing, interrater reliability and intrarater reliability of the MBSImP rating scale are ensured by requiring clinicians to train and reach at least 80% agreement with a gold-standard rater's scores before they can be registered to use the method. This current method of training to a reliability threshold with a gold standard guarantees that the clinicians using the method at least agree with the standards set by the MBSImP development team. However, although requiring clinicians to train to reach 80% agreement to a gold standard assures *some* level of agreement across raters, it remains to be examined to what degree this train-to-a-threshold method translates to good interrater and intrarater reliability. We will thus be assessing the interrater and intrarater reliability of MBSImP.

Intent of This Study

A recent review of the dysphagia literature showed that, in general, there is a lack of formal reporting of psychometric properties for existing VFS (and fiber endoscopic) swallowing assessment tools (Swan et al., 2019). This general lack of reporting led those authors to conclude that "there is insufficient evidence to recommend any individual measure included in the review as valid and reliable to interpret videofluoroscopic swallowing studies...." (p. 29). The authors of this article acknowledge the importance of assessing and reporting the psychometric properties of VFS measures of swallowing impairment. This study therefore represents both an extension of prior work on the psychometric properties of MBSImP (Martin-Harris et al., 2008) and a response to the review by Swan et al. (2019). In particular, the aims of this article are to (a) provide a further assessment of the structural validity and internal consistency of MBSImP and (b) provide a formal assessment of the interrater and intrarater reliability of MBSImP.

Method

This study consists of (a) a retrospective structural validity and internal consistency analysis of patient records and (b) a prospective study of the interrater and intrarater reliability of MBSImP-trained clinicians. Structural validity and internal consistency were assessed based on 52,726 patient records drawn from a centralized repository, the MBSImP Swallowing Data Registry (SDR). Interrater reliability and intrarater reliability were assessed based on the ratings of four SLP clinicians on 50 VFS recordings from MBSSs and reratings of 12 such recordings (24% of the total cohort).

The 52,726-Patient Data set

The data set for the structural validity and internal consistency analyses was drawn from the MBSImP SDR. The MBSImP SDR is a centralized electronic health record system that allows MBSImP-registered clinicians to easily store the MBSImP and other de-identified patient data that come from clinical practice. Thus, the MBSImP SDR consists of data from patient visits entered by clinicians trained and registered to use MBSImP. See Supplemental Material S1–S3 for screenshots of the data-entry forms used by clinicians to enter data into the MBSImP SDR. We only included initial visits for each patient to prevent within-subject correlations from affecting the analysis. This left us with a sample size of $N = 52,726$. Table 1 shows the demographic information for this sample.

The clinicians who entered data into the MBSImP SDR ($N = 6,532$) had a wide range of clinical experience. The median number of years treating patients as an SLP was 10 years (5th percentile = 2 years, 95th percentile = 30 years). The median number of years since completing training and becoming registered to use MBSImP was 3 years (5th percentile = < 1 year, 95th percentile = 9 years).

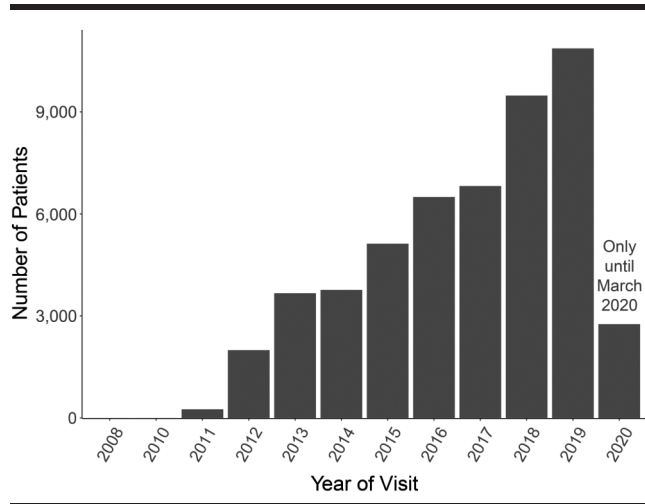
The data were derived from patients with a wide range of dysphagia-associated diagnoses. Patient diagnosis, however, was an optional field for clinicians to enter. Thus, diagnosis information was only available for a

Table 1. Demographic information for the 52,726-patient data set drawn the Modified Barium Swallow Impairment Profile Swallowing Data Registry.

Characteristic	<i>n</i>	%
Sex		
Male	27,282	52%
Female	23,828	45%
Unknown	1,616	3%
Age		
18–30	1,204	2%
31–40	1,571	3%
41–50	3,046	6%
51–60	6,987	13%
61–70	11,243	21%
71–80	12,605	24%
81–90	9,827	19%
91+	4,115	8%
Unknown	2,128	4%
Ethnicity		
Not Hispanic/Latino	35,317	67%
Hispanic/Latino	2,206	4%
Unknown/not reported	15,203	29%
Race		
White	35,879	68%
Black/African American	6,243	12%
Asian	1,050	2%
American Indian/Alaskan Native	269	0.5%
Native Hawaiian/Pacific Islander	148	0.3%
More than one race	289	0.5%
Unknown/other/not reported	8,848	16.7%

Note. $N = 52,726$.

Figure 1. A bar chart of the number of patients entered into the present sample as a function of the year their data were input into the Modified Barium Swallow Impairment Profile Swallowing Data Registry. Note that the data for 2020 were only that which were collected up until March of that year.



subset of patients. Of the patients for whom diagnosis information was available, five of the most common diagnoses were head and neck cancer ($n = 2,157$), stroke ($n = 3,791$), chronic obstructive pulmonary disorder ($n = 2,033$), Parkinson’s disease ($n = 1,053$), and dementia ($n = 1,424$). The 30 most prevalent diagnoses and their respective sample sizes can be viewed in Supplemental Material S4. The data also come from hospitals and medical centers across the world. These include all 50 states, multiple Canadian provinces, Australia, Norway, the United Kingdom, Singapore, and South Korea, among many others. The full list of the geographic locations can be viewed at <https://www.mbsimp.com/clinicians.cfm>.

The data were collected over 9 years, starting in 2008 with a steady increase in rate of data collections through the present. The number of patient visits input into the data set from the beginning of 2008 to March of 2020 is shown in Figure 1.

The 50-Patient Data set

The second data set for this study consists of 50 VFS recordings from MBSSs. These 50 VFS recordings were randomly drawn from the “high-frame rate” (30 fps) recordings of a prior study that was designed to assess possible differences in swallowing assessment between high- and low-frame rate VFSs. The only inclusion criterion for patients was that they were referred for an MBSS, and there was no exclusion criterion. The 50 patients fell into the following seven diagnostic categories: general medicine ($n = 24$), head and neck cancer ($n = 15$), neurology ($n = 5$), gastroenterology ($n = 4$), general ear, nose, and throat ($n = 2$), pulmonary ($n = 2$), and cardiac ($n = 1$).

See Table 2 for the demographic information of the 50-patient sample.

Four SLPs were selected to rate these 50 VFS recordings. In selecting the four SLP raters, we intentionally selected raters with widely varying levels of clinical experience and familiarity with MBSImP. All raters held master’s degrees, were American Speech-Language-Hearing Association certified, and were MBSImP registered. The amount of clinical practice experience ranged from 4 to 30 years, and MBSImP registration ranged between 2 and 10 years. The number of monthly MBSSs conducted by the raters varied between 0–25 and 0–20 during clinical practice and clinical research, respectively. These raters also came from differing institutions. Two raters were based in the Medical University of South Carolina, one rater was from Northwestern University, and one rater was based in private practice. See Table 4 for details on each rater.

Reliability Sample Size Justification

To choose the number of patients for this study, we focused on what size sample would allow our results to generalize across dysphagic patient populations. In the generalizability theory, generalizability is ensured when the variability (the variance) of the patient sample accurately estimates the variability of the entire patient population (Webb & Shavelson, 2005). We do not have data from the entire dysphagic patient population; we do, however, have a patient database of 52,726 patients that we

Table 2. Demographic information for the 50-patient dataset.

Characteristic	<i>n</i>	%
Sex		
Male	26	52%
Female	24	48%
Age		
18–30	2	4%
31–40	2	4%
41–50	5	10%
51–60	7	14%
61–70	13	26%
71–80	9	18%
81–90	10	20%
91+	2	4%
Ethnicity		
Not Hispanic/Latino	46	92%
Hispanic/Latino	4	8%
Unknown/not reported	0	0%
Race		
White	47	94%
Black/African American	3	6%
Asian	0	0%
American Indian/Alaskan Native	0	0%
Native Hawaiian/Pacific Islander	0	0%
More than one race	0	0%
Unknown/other/not reported	0	0%

Note. $N = 50$.

can use to estimate the generalizability of smaller samples. In order to estimate these smaller samples' generalizability, we have done the following: (a) assumed the variability of the 52,726-patient database represents the true population variance of MBSImP scores of dysphagic patients, (b) calculated the variance of the population MBSImP scores for each component, (c) took 10,000 subsamples (with replacement) of the big data set with varying sample sizes (10–100 patients in steps of 10), (d) calculated the variance of MBSImP component for each subsample, (e) calculated the absolute difference and percent difference between the population variance and each subsample variance, and (f) plotted the median difference and median percent difference (error bars represent the 5th and 95th percentiles) as a function of sample size (see Figures 2A and 2B, respectively). Note that the median difference (see Figure 2) is roughly centered on zero for all sample sizes, showing the subsamples are not systematically overestimating or underestimating the variance of the population. However, also note that in Figure 2, the range of values (i.e., the size of the error bars) decreases with sample size. This decrease in the range of values can also be seen in the percent difference plot (see Figure 2B), where the median percent difference of the subsample versus population variance decreases as a function of sample size.

In Figures 2A and 2B, we observe that sample sizes above 50 patients provide diminishing returns in terms of improving accuracy of variance estimates. Furthermore, with 50 patients, the subsamples had only a median of 6% difference in variance relative to the variance of the 52,726-patient data set, meaning that the variance of a 50-patient sample has a 94% accuracy in estimating the

variance of a 52,726-patient sample. We deemed a median of 94% accuracy acceptable for this study and therefore chose a sample size of 50 patients.

Similarly to choosing our number of patients, in selecting our raters, we considered how to best estimate the *variance between raters* that occurs in our rater population of interest, that is, clinicians and clinical researchers. Our total number of raters ($N = 4$) was limited by logistical constraints, but we attempted to best estimate the variance between raters by selecting raters with widely varying levels of expertise and familiarity with MBSImP. We have confidence that our raters are representative of the true rater population because the present raters' range of experience (4–30 years practicing clinically; 2–10 years MBSImP registered) mirrors the 5th–95th percentiles of experience of raters in the MBSImP SDR (2–30 years practicing clinically; 0–9 years MBSImP registered).

MBSImP Protocol

Clinicians must be trained in the use of the MBSImP and meet a baseline threshold of scoring reliability (Martin-Harris et al., 2008) before they are permitted to enter data into the MBSImP SDR in order to maintain fidelity of the data. This training in MBSImP includes both the MBSImP rating scale and the MBSImP bolus administration protocol.

The MBSImP rating scale is a standardized and validated scale of the severity of impairment for 17 physiologic components of swallowing (see Table 3 for the list of components). Scores on the rating scale range from 0 to 2, 3, or 4 depending on which physiologic component is being assessed. The number of impairment levels for each component was determined by an expert panel consensus. For

Figure 2. (A) The median difference (error bars represent the 95th and 5th percentiles) between the variance of the 52,787-patient data set and the variance of subsamples with sample sizes ranging from 10 to 100. (B) The median *percent* difference between the variance of the 52,726-patient data set and the variance of the subsamples, demonstrating that large sample sizes provide better estimates of the between-subjects variance but have substantially diminishing returns for sample sizes greater than ~50 patients. pop. = population.

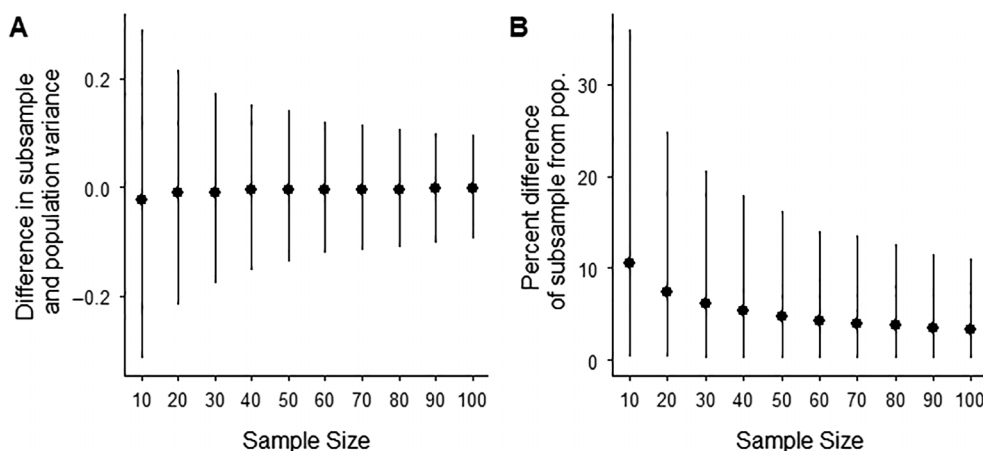


Table 3. The hypothesized domains and the factor loadings of each Modified Barium Swallow Impairment Profile component.

Hypothesized domain	Comp. no.	Component name	Factor 1	Factor 2
Oral	1	Lip closure	0.77	-0.08
	2	Tongue control during bolus hold	0.74	-0.02
	3	Bolus preparation/mastication	0.72	0.07
	4	Bolus transport/lingual motion	0.76	-0.01
	5	Oral residue	0.71	0.05
	6	Initiation of pharyngeal swallow	0.52	0.18
Pharyngeal	7	Soft palate elevation	0.11	0.46
	8	Laryngeal elevation	0.08	0.75
	9	Anterior hyoid excursion	0.05	0.76
	10	Epiglottic movement	-0.08	0.81
	11	Laryngeal vestibular closure	0.05	0.75
	12	Pharyngeal stripping wave	-0.05	0.81
	13	Pharyngeal contraction (A/P view)	-0.01	0.64
	14	Pharyngoesophageal segment opening	-0.11	0.75
	15	Tongue base retraction	0.26	0.59
	16	Pharyngeal residue	0.04	0.81
Esophageal	17	Esophageal clearance, upright position (A/P view)	-0.05	0.19

Note. Factor loadings greater than 0.4 are marked bold and indicate which components can be considered a substantive part of each factor. Comp. = component; A/P = anterior/posterior.

each MBSImP component, these experts identified ordered levels of impairment severity that each represented a unique structural movement, bolus flow, or both, related to the physiology of that component. As some components had fewer unique levels than others, this resulted in differing numbers of impairment levels for different components (Martin-Harris et al., 2008).

The MBSImP bolus administration protocol consists of 12 swallow trials with the following consistencies, volumes, and presentation methods: four thin-liquid (< 15 cps) trials (two 5 ml via teaspoon, a cup sip [20 ml], and sequential swallow from cup [40 ml]), four nectar-thick (150–450 cps) trials (two 5 ml via teaspoon with one from the typical lateral view and one from an anterior/posterior [A/P] view, a cup sip, and a sequential swallow), one thin-honey (800–1800 cps) trial (5 ml via teaspoon), two pudding (4500–7000 cps) trials (two 5 ml via teaspoon with one from the typical lateral view and one from an A/P view), and one solid trial (a half-portion of a Lorna Doone cookie coated with 3 ml of pudding barium). All trials are administered using standardized, “ready-to-use” barium contrast (VARIBAR, barium sulfate 40% weight/volume; Bracco Diagnostics, Inc.).

Furthermore, each component is only scored for the swallow trials for which it can be assessed. For example, for the sequential swallow trials, patients are not asked to hold a liquid bolus in the oral cavity; therefore, Component 2 (tongue control during bolus hold) cannot be assessed and is not scored. Similarly, Component 3 (bolus preparation/mastication) is only scored for the solid bolus trial. For the two swallow trials in the A/P view, the viewing plane provides a perspective ideal for scoring Components 13

(pharyngeal contraction) and 17 (esophageal clearance), but this also means no other components are scored from this view.

In clinical practice and typically in research, the MBSImP scores from the individual swallow trials are represented by the OI score for each of the 17 components, which is the most severe score on that component across all swallowing tasks/bolus consistencies. As such, the analyses in this study are based on OI scores. See Supplemental Material S5 for the distribution of OI scores per component in the 52,726-patient data set and Supplemental Material S6 for the distribution of OI scores per component in the 50-patient data set.

Interrater and Intrarater Protocol

Four SLPs rated approximately five MBSSs per week over a 10-week period. Raters scored each patient’s MBSS using the Swallow By Swallow (SbS) method where a score is given for each relevant MBSImP component for each swallowing task given during the protocol. This scoring method results in a total of 127 possible SbS scores for a single MBSS. From these scores, an OI score, that is, the most severe score across swallowing tasks, was computed for each MBSImP component. All analyses in this study were conducted using OI scores.

The two raters with the least experience with MBSImP (Raters 1 and 4) were allowed to meet with the gold-standard rater once a week to ask questions about scoring for continuing education purposes, but no scores could be changed post hoc based on these meetings. This question-asking protocol was allowed because, in “real-world” settings, MBSImP-registered clinicians also have

the opportunity to send questions to Northern Speech Services and receive guidance on scoring. After all 50 studies were scored, each rater waited 2 weeks and then rerated 12 studies (~24% of the total) in a like manner to the initial rating for use in the intrarater analysis.

Statistical Analysis

The data from the 52,726-patient data set were used to assess the structural validity and internal consistency of MBSImP. Structural validity was assessed by submitting the 17 MBSImP component scores of all patients in the data set to an exploratory factor analysis. The exploratory factor analysis was computed using polychoric correlations and minimum residuals estimation appropriate to MBSImP components scores as ordinal (Holgado-Tello et al., 2008) and nonnormally distributed (Cudeck, 2000), respectively. Furthermore, an oblimin rotation was used because the hypothesized oral and pharyngeal domains were expected to be correlated. The number of factors (2) was chosen based on the hypothesized number of multi-component domains in the scale, that is, the oral and pharyngeal domains, and also based on a prior study that showed a two-factor solution for MBSImP was adequate (Martin-Harris et al., 2008). An MBSImP component was considered to be a substantive part of a factor if it had a factor loading of > 0.4 , a criterion set after loadings were computed. To assess the internal consistency, two values of Cronbach's alpha (Cronbach, 1951) were computed: one Cronbach's alpha from the components in the oral domain (Components 1–6) and one Cronbach's alpha for the pharyngeal domain.

The data from the 50-patient data set were used to assess the interrater and intrarater reliability of MBSImP. Interrater reliability and intrarater reliability were assessed using the intraclass correlation coefficient (ICC; Fleiss & Cohen, 1973). ICC was used because it is calculated based on differences in variances across and within raters, and this approach is well suited in handling the differing numbers of severity levels across the 17 components of MBSImP. To compute the ICC, we used a two-way random-effects model at the single-rater level, with absolute agreement as the measure of interest. We chose to use a two-way random-effects model as the goal of the model was generalization of the results to clinicians outside of the present sample; this model was chosen to be at the "single-rater" level as our intention was to assess agreement between individual clinicians; and finally, the model was chosen to have "absolute agreement" as its basis because our goal is to assess to what extent clinicians have precisely the same score and not just whether the relative position of scores is consistent across clinicians (Koo & Li, 2016). All statistical analyses were conducted using the "psych" package (Revelle, 2021) in the R programming language (R Core Team, 2019).

Results

Structural Validity

An exploratory factor analysis revealed that MBSImP has a two-factor solution that exactly corresponds to the hypothesized oral and pharyngeal domains. Factor 1 had loadings between 0.52 and 0.77 for oral-related components (1–6) and loadings of at most 0.18 for all the other components. This factor accounted for 19% of the total variance in MBSImP scores. Factor 2 had loadings between 0.46 and 0.81 for all pharyngeal-related components (7–16) and loadings less than 0.26 for all the other components. This factor accounted for 32% of the total variance. Component 17, the only esophageal-related component, had loadings of at most 0.19 for both factors. Exact loadings for each factor and MBSImP component are shown in Table 3. The two factors were correlated with a strength of $r = .56$.

Internal Consistency

Cronbach's α was .81 (95% confidence interval [CI; .808, .812]) for the oral domain and .87 (95% CI [.868, .872]) for the pharyngeal domain.

Interrater and Intrarater Reliability

Interrater reliability across all four clinicians as measured by the ICC was .78 (95% CI [.76, .80]). Intrarater reliability, also measured by ICC for the four clinicians, ranged from .82 to .87 (see Table 4).

Discussion

The assessment of the MBSS using VFS imaging by clinicians requires the use of clinical judgment. Codifying clinical judgment into standardized assessment methods offers a range of benefits including ease of comparison between patients and of tracking of patient trajectories. However, in order to reap the benefits of standardization, it requires that these methods have the necessary psychometric properties (Lambert et al., 2002). A previous investigation of MBSImP showed that multiple aspects of the psychometric validity of MBSImP were quite good, including content validity, hypothesis testing, and structural validity (Martin-Harris et al., 2008; $N = 300$). This study complements this prior study by extending the analysis of structural validity and internal consistency to a large-scale, clinical data set ($N = 52,726$) and by adding a formal assessment of rater reliability. Both the prior study and this study use OI scores as the basis for analysis due to OI scoring being the main way that MBSImP is used both clinically and in research studies. Thus, the results here show that MBSImP, as it is used clinically and in research studies, has good structural validity, internal consistency, and interrater and intrarater reliability.

Table 4. Clinical, research, and Modified Barium Swallow Impairment Profile (MBSImP) experience along with the intrarater reliability scores and associated confidence intervals of the four speech-language pathologist raters in this study.

Rater	Graduation year of master's	Years MBSImP registered	Years practicing clinically	MBSS performed/rated each month		Intrarater reliability ICC	ICC 95% confidence interval
				For clinical purposes	For research purposes		
1	2016	2	4	0	10–15	.82	.77, .86
2	2014	6	6	20–25	0–10	.83	.79, .87
3	2007	10	14	5	10–15	.87	.83, .90
4	1990	3	30	0	20	.87	.83, .90

Note. MBSS = modified barium swallow study; ICC = intraclass correlation coefficient.

Structural Validity and Internal Consistency

Reports of the structural validity and internal consistency of VFS assessment methods are rare. In part, this may be because many studies of the psychometric properties of VFS assessment methods only examine one or two physiological outcomes (e.g., Hind et al., 2008; Hutcheson et al., 2017; Karnell & Rogus, 2005; Kelly et al., 2006; Mann et al., 2000; Rommel et al., 2015; Rosenbek et al., 1996). Only examining one or two items in a study means that it is either impossible or not very meaningful to examine the correlations between multiple items, which is the basis of structural validity and internal consistency analyses. Even in psychometric studies of VFS assessment methods that examine multiple items (e.g., Frowen et al., 2008; Gibson et al., 1995; Kim et al., 2012; Lee et al., 2017; Martin-Harris et al., 2008; McCullough et al., 2001; Scott et al., 1998; Stoeckli et al., 2003), few examine structural validity and internal consistency (i.e., only Frowen et al., 2008, and Martin-Harris et al., 2008, in the cited examples). Therefore, this study may serve as a cue to other researchers developing and using VFS assessment methods that it is important to assess structural validity and internal consistency.

The present examination of the structural validity and internal consistency of MBSImP represents a large-scale extension of the work done by Martin-Harris et al. (2008). This prior study showed that, in a sample of 300 patients, MBSImP had a two-factor structure, corresponding to the hypothesized oral and pharyngeal domains. However, due to missing data and small factor loadings, four components were removed from the analysis (Component 1 for inconsistent visualization of the lips, Component 7 for lack of sufficient variability in the sample, and Components 13 and 17 for a small sample size due to omission of the A/P view). This study extends the analysis of structural validity to a substantially larger sample ($N = 52,726$), allowing us to include all 17 components in the structural validity analysis and to assess the internal consistency of each of the multicomponent domains.

The results of this analysis show that, in a large-scale, clinical data set, MBSImP demonstrates good structural

validity and internal consistency. The present results suggest that MBSImP is structurally valid, as shown by a two-factor structure corresponding to the oral and pharyngeal domain components, respectively, with Component 17 (esophageal clearance, representing the esophageal domain) being separable from both of those domains. The good structural validity demonstrated here provides further evidence that the statistical structure of MBSImP is very much aligned with the hypothesized structure of MBSImP.

The present results also show that MBSImP has good internal consistency in that Cronbach's alpha reaches sufficiently high levels for each of the multicomponent domains. The good internal consistency of each domain suggests that the components of each domain, as a group, are measuring a latent "overall impairment" variable. This existence of a latent "overall impairment" variable for each domain provides support for the legitimacy of using oral total and pharyngeal total scores (summed scores of the components of each domain), which are designed to measure this overall oral and pharyngeal impairment, respectively.

The oral and pharyngeal domains are structurally valid, but they alone do not fully characterize a patient's impairment. This is evident in the result that the total explained variance of the oral and pharyngeal factors accounted for 51% of the variance in the MBSImP assessment method, leaving 49% of the variance to be accounted for by the individual components. This split in the sources of variance in MBSImP means that a patient's impairment is operating at two relatively separable levels: (a) the domain level characterized by the oral and pharyngeal total scores and (b) the component level characterized by which components are impaired. Each of these levels can potentially be independently influenced by patient-specific factors. This result highlights the importance of conducting assessments and analyses at both the domain level (oral and pharyngeal total scores) and the individual-component level.

Furthermore, although the oral and pharyngeal domains form separable factors, they should be considered separable but related, as the correlation between the factors was $r = .56$. Esophageal function, that is, Component

17 (esophageal clearance), however, was clearly separable from the oral and pharyngeal domains in this data set because it did not load onto either factor and showed weak, although not necessarily null, correlations with each oral or pharyngeal component of MBSImP (see Table 3 and Supplemental Material S7). The existence of potentially nonnull correlations between the esophageal domain and the oral and pharyngeal domains is consistent with the detection of such correlations in a small case series study by Gullung et al. (2012).

These structural validity and internal consistency results are likely to be highly generalizable due to the data set's large sample size, clinical nature, wide variety of diagnoses, representative demographics, and diversity of data collection locations. The large sample size of the data set helps ensure that the correlations that form the basis of these results are not being driven by sampling errors such as being overly affected by individual patient scores. The present data set is also derived from clinical patient visits, which means that the results are likely to be generalizable to our population of interest, that is, any patient referred to an MBSS conducted by an MBSImP-registered clinician. In addition, the 52,726-patient data set includes scores of patients with any and all diagnoses entered into the MBSImP SDR including, but not limited to, head and neck cancer, stroke, chronic obstructive pulmonary disorder, Parkinson's disease, and dementia. This wide variety of diagnoses allows the results to be generalized across populations with differing diagnoses. Furthermore, the demographics of the present data are roughly representative of the demographics of the U.S. population (United States Census Bureau, 2021), admittedly with some underrepresentation of patients who are Hispanic, female, or Asian. This rough representativeness of the present data suggests that the present results are fairly generalizable across U.S. demographics (see Limitations for further discussion). Finally, the present data were collected from a wide array of locations and institutions across the world. This diversity of data collection locations lends support to the applicability of the present results regardless of geographic location.

Interrater and Intrarater Reliability

The literature on the reliability of MBSS assessment scales is quite heterogeneous. Depending on the structures of the scale, studies use a variety of statistics to describe reliability including kappa for dichotomous/binary scales (Bryant et al., 2012; Gibson et al., 1995; Kim et al., 2012; Lee et al., 2017; McCullough et al., 2001; Stoeckli et al., 2003); weighted kappa for ordinal scales with a uniform number of levels across items (Bryant et al., 2012; Hutcheson et al., 2017); and ICC for continuous variables, ordinal scales with varying numbers of levels, or, in one case, a single ordinal item (Frowen et al., 2008; Kim

et al., 2012; Rommel et al., 2015). Similarly to this study, Kim et al. (2012) assessed the interrater reliability of the Videofluoroscopic Dysphagia Scale (VDS) using ICC for items with varying numbers of severity levels. That study found that, for VDS, ICC is .556, which was deemed a "moderate" level of agreement. Relative to the moderate level of agreement of VDS with an ICC of .556, the interrater ICC of .78 for MBSImP found in this study can be deemed "good." In addition, the interrater reliability in this study of ICC = .78 (95% CI [.76, .80]) can also be considered "good" as per the guidelines set out by Koo and Li (2016), where an ICC can be deemed "good" if it is > .75 and < .9. Along these same lines, the intrarater ICCs in this study ranging from .82 to .87 can all be considered "good." The intrarater ICC range of .82–.87 is only slightly greater than the interrater ICC of .78, suggesting that each of the present raters agreed with other raters only slightly less than they agreed with themselves. Furthermore, the two raters with more years of clinical experience (14 and 30 years) showed higher intrarater ICCs (.87 and .87) than the two raters with fewer years of experience (4 and 6 years), potentially suggesting that SLPs' intrarater agreement may improve with experience. However, this result should be interpreted with caution because the CIs of all raters are overlapping. A future study could specifically aim to test this hypothesis that intrarater reliability improves with years of clinical experience.

Together, these results suggest that the standardized MBSImP training that requires clinicians to reach 80% agreement with a gold-standard rater translates to good interrater and intrarater reliability on previously unseen MBSS. Furthermore, these SLPs had widely varying levels of experience across a range of domains (i.e., years as an SLP, years as MBSImP certified, number of studies rated per month, home institution) providing evidence that, for these clinicians, it was possible for all experience levels and backgrounds to demonstrate good intrarater reliability using MBSImP.

In summary, we have shown in a large-scale, clinical data set that MBSImP has excellent structural validity and internal consistency. In addition, we have provided a formal test of the rater reliability of MBSImP-trained raters and demonstrated that the standardized MBSImP training can result in good interrater and intrarater reliability.

Limitations

The following are this study's limitations:

1. All results in this article are based on OI scores representing the worst performance (highest score on an ordinal scale) across all the swallowing tasks in the MBSImP method. Thus, the results show

satisfactory psychometric properties of these OIs, but caution should be applied before generalizing these findings to the psychometric properties of MBSImP ratings for individual swallows. As SBS scoring is likely confined to research contexts, the authors recommend that researchers using SBS scoring conduct their own independent rater reliability testing. Furthermore, as the hypothesized structure of MBSImP is at the component level and thus at the level of OI scores, structural validity and internal consistency do not apply to SBS scoring level.

2. The present rater reliability assessment of MBSImP represents the reliability of the *entire* MBSImP scale. A future study with a larger sample of raters and patients would be necessary to investigate the reliability of the individual MBSImP components and would provide substantial value to the field.
3. There is some underrepresentation of Asian, Hispanic, and female populations relative to the U.S. Census (United States Census Bureau, 2021). Interestingly, the representation of Asians in the present sample is quite similar to that seen in the dysphagic population of a study on the prevalence of dysphagia that used a professional survey company to obtain a representative sample of the U.S. population (Adkins et al., 2020). Furthermore, in the present sample, 29% of patients did not have their ethnicity input into the MBSImP SDR. It is then possible that the apparent underrepresentation of Hispanics/Latinos maybe due to nonreporting rather than true underrepresentation. In addition, it has been reported that men are more likely to seek care for dysphagia than women (Adkins et al., 2020). This difference in care-seeking behavior may at least partially account for the apparent underrepresentation of women in the present sample. Nonetheless, the apparent underrepresentation of patients who report being Asian, Hispanic, and female could impose a limitation on the generalizability of the results. However, this limitation would only occur if there were systematic differences in MBSImP scores for populations with versus without these demographic identifiers. Future work should compare the MBSImP scores of populations with different demographic identifiers to uncover whether such systematic differences exist.

Author Contributions

Alex E. Clain: Lead role – Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing.
Munirah Alkhuwaiter: Supporting role – Conceptualization,

Formal analysis, Methodology, Validation, Writing – review & editing.
Kate Davidson: Supporting role – Conceptualization, Data curation, Investigation, Project administration, Supervision, Writing – review & editing.
Bonnie Martin-Harris: Supporting role – Conceptualization, Project administration, Supervision, Methodology, Writing – original draft, Writing – review & editing.

Acknowledgments

The authors thank the speech-language pathologist raters for their participation and dedication throughout this project. They are (in alphabetical order) Priscilla Brown, Melissa Cooke, Abigail Nellis, and Elizabeth Platt. The authors also wish to acknowledge that this study was supported by the National Institute on Deafness and Other Communication Disorders at the National Institutes of Health (NIH/NIDCD 2K24DC012801-0).

References

- Adkins, C., Takakura, W., Spiegel, B. M. R., Lu, M., Vera-Llonch, M., Williams, J., & Almario, C. V. (2020). Prevalence and characteristics of dysphagia based on a population-based survey. *Clinical Gastroenterology and Hepatology*, 18(9), 1970–1979.e2. <https://doi.org/10.1016/j.cgh.2019.10.029>
- Arrese, L. C., Carrau, R., & Plowman, E. K. (2017). Relationship between the eating assessment Tool-10 and objective clinical ratings of swallowing function in individuals with head and neck cancer. *Dysphagia*, 32(1), 83–89. <https://doi.org/10.1007/s00455-016-9741-7>
- Arrese, L. C., Schieve, H. J., Graham, J. M., Stephens, J. A., Carrau, R. L., & Plowman, E. K. (2019). Relationship between oral intake, patient perceived swallowing impairment, and objective videofluoroscopic measures of swallowing in patients with head and neck cancer. *Head & Neck*, 41(4), 1016–1023. <https://doi.org/10.1002/hed.25542>
- Bryant, K. N., Finnegan, E., & Berbaum, K. (2012). VFS Interjudge reliability using a free and directed search. *Dysphagia*, 27(1), 53–63. <https://doi.org/10.1007/s00455-011-9337-1>
- Clark, H. M., Stierwalt, J. A. G., Tosakulwong, N., Botha, H., Ali, F., Whitwell, J. L., & Josephs, K. A. (2020). Dysphagia in progressive supranuclear palsy. *Dysphagia*, 35(4), 667–676. <https://doi.org/10.1007/s00455-019-10073-2>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Cudeck, R. (2000). 10—Exploratory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 265–296). Academic Press. <https://doi.org/10.1016/B978-012691360-6/50011-2>
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613–619. <https://doi.org/10.1177/001316447303300309>
- Frowen, J. J., Cotton, S. M., & Perry, A. R. (2008). The stability, reliability, and validity of videofluoroscopy measures for

- patients with head and neck cancer. *Dysphagia*, 23(4), 348–363. <https://doi.org/10.1007/s00455-008-9148-1>
- Gibson, E., Phyland, D., & Marschner, I.** (1995). Rater reliability of the modified barium swallow. *Australian Journal of Human Communication Disorders*, 23(2), 54–60. <https://doi.org/10.3109/asl2.1995.23.issue-2.05>
- Gullung, J. L., Hill, E. G., Castell, D. O., & Martin-Harris, B.** (2012). Oropharyngeal and esophageal swallowing impairments: Their association and the predictive value of the modified barium swallow impairment profile and combined multi-channel intraluminal impedance—Esophageal manometry. *Annals of Otolaryngology, Rhinology & Laryngology*, 121(11), 738–745. <https://doi.org/10.1177/000348941212101107>
- Hind, J. A., Gensler, G., Brandt, D. K., Miller Gardner, P. J., Blumenthal, L., Gramigna, G. D., Kosek, S., Lundy, D., McGravey-Toler, S., Rockafellow, S., Sullivan, P. A., Villa, M., Gill, G. D., Lindblad, A. S., Logemann, J. A., & Robbins, J.** (2008). Comparison of trained clinician ratings with expert ratings of aspiration on Videofluoroscopic images from a randomized clinical trial. *Dysphagia*, 24(2), 211. <https://doi.org/10.1007/s00455-008-9196-6>
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E.** (2008). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153. <https://doi.org/10.1007/s11135-008-9190-y>
- Hutcheson, K. A., Barrow, M. P., Barringer, D. A., Knott, J. K., Lin, H. Y., Weber, R. S., Fuller, C. D., Lai, S. Y., Alvarez, C. P., Raut, J., Lazarus, C. L., May, A., Patterson, J., Roe, J. W. G., Starmer, H. M., & Lewin, J. S.** (2017). Dynamic imaging grade of swallowing toxicity (DIGEST): Scale development and validation. *Cancer*, 123(1), 62–70. <https://doi.org/10.1002/cncr.30283>
- Hutcheson, K. A., Lewin, J. S., Barringer, D. A., Lisec, A., Gunn, G. B., Moore, M. W. S., & Holsinger, F. C.** (2012). Late dysphagia after radiotherapy-based treatment of head and neck cancer. *Cancer*, 118(23), 5793–5799. <https://doi.org/10.1002/cncr.27631>
- Im, I., Jun, J.-P., Crary, M. A., Carnaby, G. D., & Hong, K. H.** (2019). Longitudinal kinematic evaluation of pharyngeal swallowing impairment in thyroidectomy patients. *Dysphagia*, 34(2), 161–169. <https://doi.org/10.1007/s00455-018-9949-9>
- Im, I., & Ko, M.-H.** (2020). Relationships between temporal measurements and swallowing impairment in unilateral stroke patients. *Communication Sciences & Disorders*, 25(4), 966–975. <https://doi.org/10.12963/csd.20778>
- Karnell, M. P., & Rogus, N. M.** (2005). Comparison of clinician judgments and measurements of swallow response time: A preliminary report. *Journal of Speech, Language, and Hearing Research*, 48(6), 1269–1279. [https://doi.org/10.1044/1092-4388\(2005/088\)](https://doi.org/10.1044/1092-4388(2005/088))
- Kelly, A. M., Leslie, P., Beale, T., Payten, C., & Drinnan, M. J.** (2006). Fiberoptic endoscopic evaluation of swallowing and videofluoroscopy: Does examination type influence perception of pharyngeal residue severity? *Clinical Otolaryngology*, 31(5), 425–432. <https://doi.org/10.1111/j.1749-4486.2006.01292.x>
- Kim, D. H., Choi, K. H., Kim, H. M., Koo, J. H., Kim, B. R., Kim, T. W., Ryu, J. S., Im, S., Choi, I. S., Pyun, J. S., Park, J. W., Kang, J. K., & Yang, H. S.** (2012). Inter-rater reliability of Videofluoroscopic Dysphagia Scale. *Annals of Rehabilitation Medicine*, 36(6), 791–796. <https://doi.org/10.5535/arm.2012.36.6.791>
- Koo, T. K., & Li, M. Y.** (2016). A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lachin, J. M.** (2004). The role of measurement reliability in clinical trials. *Clinical Trials*, 1(6), 553–566. <https://doi.org/10.1191/1740774504cn057oa>
- Lambert, H. C., Gisel, E. G., & Wood-Dauphinee, S.** (2002). The functional assessment of dysphagia. *Physical & Occupational Therapy in Geriatrics*, 19(3), 1–14. https://doi.org/10.1080/J148v19n03_01
- Lee, J. W., Randall, D. R., Evangelista, L. M., Kuhn, M. A., & Belafsky, P. C.** (2017). Subjective assessment of videofluoroscopic swallow studies. *Otolaryngology—Head and Neck Surgery*, 156(5), 901–905. <https://doi.org/10.1177/0194599817691276>
- Mann, G., Hankey, G. J., & Cameron, D.** (2000). Swallowing disorders following acute stroke: Prevalence and diagnostic accuracy. *Cerebrovascular Diseases*, 10(5), 380–386. <https://doi.org/10.1159/000016094>
- Martin-Harris, B., Brodsky, M. B., Michel, Y., Castell, D. O., Schleicher, M., Sandidge, J., Maxwell, R., & Blair, J.** (2008). MBS measurement tool for swallow impairment—MBSImp: Establishing a standard. *Dysphagia*, 23(4), 392–405. <https://doi.org/10.1007/s00455-008-9185-9>
- Martin-Harris, B., McFarland, D., Hill, E. G., Strange, C. B., Focht, K. L., Wan, Z., Blair, J., & McGrattan, K.** (2015). Respiratory–swallow training in patients with head and neck cancer. *Archives of Physical Medicine and Rehabilitation*, 96(5), 885–893. <https://doi.org/10.1016/j.apmr.2014.11.022>
- Martino, R., Pron, G., & Diamant, N. E.** (2004). Oropharyngeal dysphagia: Surveying practice patterns of the speech-language pathologist. *Dysphagia*, 19(3), 165–176. <https://doi.org/10.1007/s00455-004-0004-7>
- McCullough, G. H., Wertz, R. T., Rosenbek, J. C., Mills, R. H., Webb, W. G., & Ross, K. B.** (2001). Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. *Dysphagia*, 16(2), 110–118. <https://doi.org/10.1007/PL00021291>
- Northern Speech Services.** (2016). *MBSImp—Registered clinicians*. <https://www.mbsimp.com/clinicians.cfm>
- O'Rourke, A., Humphries, K., Lazar, A., & Martin-Harris, B.** (2017). The pharyngeal contractile integral is a useful indicator of pharyngeal swallowing impairment. *Neurogastroenterology & Motility*, 29(12), Article e13144. <https://doi.org/10.1111/nmo.13144>
- Pettigrew, C. M., & O'Toole, C.** (2007). Dysphagia evaluation practices of speech and language therapists in Ireland: Clinical assessment and instrumental examination decision-making. *Dysphagia*, 22(3), 235–244. <https://doi.org/10.1007/s00455-007-9079-2>
- R Core Team.** (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W.** (2021). *psych: Procedures for personality and psychological research*. Northwestern University. <https://CRAN.R-project.org/package=psychVersion=2.1.6>
- Rommel, N., Borgers, C., Van Beckevoort, D., Goeleven, A., Dejaeger, E., & Omari, T. I.** (2015). Bolus Residue Scale: An easy-to-use and reliable videofluoroscopic analysis tool to score bolus residue in patients with dysphagia. *International Journal of Otolaryngology*, 2015, 780197. <https://doi.org/10.1155/2015/780197>
- Rosenbek, J. C., Robbins, J. A., Roecker, E. B., Coyle, J. L., & Wood, J. L.** (1996). A Penetration–Aspiration Scale. *Dysphagia*, 11(2), 93–98. <https://doi.org/10.1007/BF00417897>
- Rumbach, A., Coombes, C., & Doeltgen, S.** (2018). A survey of Australian dysphagia practice patterns. *Dysphagia*, 33(2), 216–226. <https://doi.org/10.1007/s00455-017-9849-4>
- Scott, A., Perry, A., & Bench, J.** (1998). A study of interrater reliability when using videofluoroscopy as an assessment of

-
- swallowing. *Dysphagia*, 13(4), 223–227. <https://doi.org/10.1007/PL00009576>
- Souza, A. C., Alexandre, N. M. C., & de Brito Guirardello, E.** (2017). Psychometric properties in the evaluation of instruments: Evaluation of reliability and validity. *Epidemiology and Health Services*, 26(3), 649–59. <https://doi.org/10.5123/s1679-49742017000300022>
- Stoeckli, S. J., Huisman, T. A. G. M., Seifert, B. A. G. M., & Martin-Harris, B. J. W.** (2003). Interrater reliability of videofluoroscopic swallow evaluation. *Dysphagia*, 18(1), 53–57. <https://doi.org/10.1007/s00455-002-0085-0>
- Swan, K., Cordier, R., Brown, T., & Speyer, R.** (2019). Psychometric properties of visuoperceptual measures of videofluoroscopic and fibre-endoscopic evaluations of swallowing: A systematic review. *Dysphagia*, 34(1), 2–33. <https://doi.org/10.1007/s00455-018-9918-3>
- United States Census Bureau.** (2021). *U.S. Census Bureau QuickFacts: United States*. <https://www.census.gov/quickfacts/fact/table/US/PST045219>
- Webb, N., & Shavelson, R.** (2005). *Generalizability theory: Overview*. <https://doi.org/10.1002/0470013192.bsa703>
- Wilmskoetter, J., Martin-Harris, B., Pearson, W. G., Bonilha, L., Elm, J. J., Horn, J., & Bonilha, H. S.** (2018). Differences in swallow physiology in patients with left and right hemispheric strokes. *Physiology & Behavior*, 194, 144–152. <https://doi.org/10.1016/j.physbeh.2018.05.010>
- Xinou, E., Chrysogonidis, I. A., Kalogera-Fountzila, A., Panagiotopoulou-Mpoukla, D., & Printza, A.** (2018). Longitudinal evaluation of swallowing with videofluoroscopy in patients with locally advanced head and neck cancer after chemoradiation. *Dysphagia*, 33(5), 691–706. <https://doi.org/10.1007/s00455-018-9889-4>