



Published in final edited form as:

*Dysphagia*. 2013 June ; 28(2): 178–187. doi:10.1007/s00455-012-9427-8.

## Development, reliability and validation of an infant mammalian penetration-aspiration scale

Shaina Devi Holman, B.S.<sup>1,2,\*</sup>, Regina Campbell-Malone, Ph.D.<sup>1</sup>, Peng Ding, MD, Ph.D.<sup>1</sup>, Estela M. Gierbolini-Norat, B.S.<sup>1</sup>, Anne M. Griffioen, B.A.<sup>1</sup>, Haruhi Inokuchi, M.D.<sup>1</sup>, Stacey L. Lukasik, B.A.<sup>1</sup>, and Rebecca Z. German, PhD<sup>1,2</sup>

<sup>1</sup>Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, 98 N Broadway, Suite 409, Baltimore, MD 21231

<sup>2</sup>Department of Pain and Neural Sciences, University of Maryland School of Dentistry, 650 W Baltimore St, Baltimore, MD 21201

### Abstract

A penetration-aspiration scale exists for assessing airway protection in adult videofluoroscopy and fiberoptic endoscopic swallowing studies, however no such scale exists for animal models. The aim of this study was threefold to 1) develop a Penetration-Aspiration Scale (PAS) for infant mammals, 2) test the scale's intra- and inter-rater reliability, and 3) to validate the use of the scale for distinguishing between abnormal and normal animals. After discussion and reviewing many videos, the result was a 7-Point Infant Mammal PAS. Reliability was tested by having 5 judges score 90 swallows recorded with videofluoroscopy across two time points. In these videos, the frame rate was either 30 or 60 frames per second and the animals were either normal, had a unilateral superior laryngeal nerve (SLN) lesion, or had hard palate local anesthesia. The scale was validated by having one judge score videos of both normal and SLN lesioned pigs and testing the difference using a t-test. Raters had a high intra-rater (average kappa of 0.82, intraclass correlation coefficient (ICC)= 0.92) and high inter-rater reliability (average kappa of 0.68, ICC= 0.66). There was a significant difference in reliability for videos captured at 30 and 60 frames per second for scores of 3 and 7 ( $p < 0.001$ ). The scale was also validated for distinguishing between normal and abnormal pigs ( $p < 0.001$ ). Given the increasing number of animal studies using videofluoroscopy to study dysphagia, this scale provides a valid and reliable measure of airway protection during swallowing in infant pigs that will give these animal models increased translational significance.

### Keywords

Aspiration; Penetration; Airway protection; Videofluoroscopy; Deglutition; Deglutition Disorders; Animal Models

## INTRODUCTION

Penetration and aspiration can occur in the absence of adequate airway protection during swallowing. Penetration occurs when material passes into the airway, but remains at or above the vocal folds in the supraglottic space. Aspiration occurs when material passes below the vocal folds and into the trachea. In clinical studies of swallowing dysfunction, the

\*Corresponding author. swahl001@umaryland.edu, 443-799-7714, 98 N. Broadway, Suite 409, Baltimore MD 21231.

Conflict of Interest

The authors declare that they have no conflicts of interest

incidence of both penetration and aspiration are associated with increased risk of developing pneumonia, which can be life threatening (1). The 8-Point Penetration-Aspiration Scale (PAS) (2), used in clinical research studies, is an established and validated instrument for measuring the airway protection during swallowing (3–7). No such scale, however, currently exists for animal models of swallowing dysfunction.

Animal models are an important tool for understanding both normal and abnormal swallowing. Such models are valuable for collecting data on swallowing dysfunction that is not possible in humans, for both logistic and ethical reasons (8–11). For example, we can study swallow function in animals using videofluoroscopy without the time limitation that is necessary in clinical studies due to limits of radiation exposure. This model also permits multiple and unlimited videofluoroscopy recordings over several days or even months. In animal models, researchers can generate pathological conditions through nerve lesions, anesthesia or other means and then test the effect of that specific condition on function in the same individual (12, 13). The other advantage of animal models is the ability to control extraneous and confounding factors that frequently co-occur in human clinical conditions (14, 15).

A valuable animal model of studying infant swallowing is the infant pig (*Sus scrofa*). The main reason for studying infant pigs in swallowing dysfunction studies is because the anatomy of the infant pig parallels other infant mammals, which enables us to draw clinically relevant conclusions to the treatment of human infants with swallowing dysfunction (16). For example, a model of superior laryngeal nerve (SLN) lesion in the infant pig is currently being used to understand how loss of sensory information carried by this nerve affects the motor control of swallowing (17, 18). In this model, we have documented both penetration and aspiration in the infant pig (17).

The aim of this paper was to 1) develop a novel PAS for the infant pig model of swallowing dysfunction based on the standard clinical research scale, 2) assess the intra- and inter-rater reliability, and 3) test the validity of this scale for differentiating abnormal versus normal animals.

## METHODS

### Summary of methods used for unilateral SLN lesion and palatal local anesthesia

The videofluoroscopic recordings utilized to develop this scale and assess reliability and validity were from ongoing studies in the lab. These methods are summarized below.

The pigs that had a unilateral SLN lesion had undergone two surgeries. The pigs were 10–16 days old and 5–6 kg in weight. During the first surgery they were given general anesthesia (5% Isoflurane) and intubated and then underwent surgery that lasted 3–5 hours. During this surgical procedure, electromyographic (EMG) electrodes were implanted into hyolaryngeal muscles and a radio-opaque marker was sutured to the hyoid bone and thyroid cartilage. These hyolaryngeal muscles include mylohyoid, genioglossus, geniohyoid, digastrics, thyrohyoid, sternothyroid, sternohyoid and cricothyroid. After this surgical procedure was completed, radio-opaque markers were placed into the tongue, gingiva, hard palate and soft palate and a radio-opaque marker was placed with an intraoral approach on the epiglottis. The animal recovered from surgery after 1–5 hours and was then fed swine milk formula (Land O'Lakes Solustart Pig Milk Replacer) containing barium using a bottle with a 'pig nipple' (Nasco, Fort Atkinson, WI) while investigators recorded both lateral videofluoroscopy at 30 or 60 frames per second and EMG signals from the electrodes placed into the hyoid musculature. This first recording was for control data. After 1–2 days the pigs underwent another surgery where the SLN was cut on the right side before it branches into

the internal and external branches that supply the larynx and cricothyroid muscles. After the animal recovered, the animal was again fed using procedures identical to those for control data collection. After all the necessary recordings were completed, the animal was euthanized and the location of the EMG electrodes and markers were confirmed by dissection.

The pigs with palatal local anesthesia underwent one surgery to place EMG electrodes into their hyoid musculature and radio-opaque markers placed into the same structures as with the SLN lesioned pigs. Starting a day after surgery, these pigs were fed while recording lateral videofluoroscopy and from the EMG electrodes at 4 time points, 2 hours apart across a day for control data. Early on the next day, the pig was given a 0.5ml injection of 0.5% bupivacaine, a long-lasting local anesthetic, at 3 locations: the nasopalatine foramen, and left and right greater palatine foramina. The technique used was a standard dental local anesthesia nerve block. Bupivacaine starts being effective 15 minutes post-injection and lasts for 3–5 hours in small dogs, which is considered to be comparable in infant pigs due to their similar size (19). Starting one hour after the nerve blocks the animal was fed using the same procedures as were used for the SLN lesion animals. They were fed 4 total times, every 2 hours. Animals were counterbalanced so that half of the pigs were recorded on day 1 post-surgery with no local anesthesia and on day 2 with local anesthesia, while the other half of pigs were recorded on day 1 post-surgery with local anesthesia and on day 2 without local anesthesia. After all the recordings were completed, the animal was euthanized and the location of the electrodes and markers were confirmed by dissection. All of these procedures were approved by the Johns Hopkins Medical Institute IACUC.

### Development of the Scale

In order to adapt the clinical PAS to the infant mammal, we examined a number of infant pig videofluoroscopic images of normal swallows, as well as swallows with clear penetration and aspiration. The infant pigs studied were all from the previously described studies.

Our infant mammal PAS was based on the current clinical 8-Point PAS (1; Table 1). The result was a 7-Point Infant Mammal PAS (IMPAS) (Table 2). As with the PAS, this scale was multidimensional. It took into account the depth of the bolus into the airway, whether it was above or below the vocal folds, as well as the animal's response to the bolus, whether it was expelled passively, actively or not at all. Similar to the PAS used in clinical research, our scale was ordinal, meaning that lower scores represent less severe conditions (more airway protection during swallowing), and higher scores reflect more severe conditions (less airway protection during swallowing). Below is a description of each score.

**No Penetration**—A score of 1 on the IMPAS is the equivalent to the score of 1 on the clinical PAS (Table 1). During these swallows no material enters the airway. The material, or in this case milk, flows over the epiglottis as it moves caudally to protect the airway (Fig. 1b).

**Penetration**—A score of 2 on the IMPAS is similar to the score 2 on the clinical 8-Point PAS. The score of 2 on the clinical scale is when material enters the airway, remains above the vocal folds and then is ejected. On the clinical scale this material is ejected either passively or by a cough. In the infant pig we observed a similar case where material would enter the airway, remain above the vocal folds and passively leave the airway, often before the epiglottis returned to its upright position (Fig. 1c).

Scores 3 and 4 on the IMPAS are similar to the score of 3 on the clinical PAS. In the clinical scale, material enters the airway and remains above the vocal folds. This was also seen in

infant pigs; however, we observed two distinct conditions where milk remained above the vocal folds. Occasionally, a very small amount of milk was seen on the caudal side of the epiglottis by its base forming a small triangle (Fig. 1d). In other cases, a substantial amount of material remained in the airway, above the vocal folds, however it was traveling towards the vocal folds (Fig. 1e). Thus we divided this category into two scores as having more material closer to the vocal folds was perceived as a more severe condition that would increase the risk of aspiration as it does in human infants (20). The score of 3 is a small amount of material above the vocal folds on the inverse side of the epiglottis forming a small triangle by the base of the epiglottis and a score of 4 is a large amount of material.

It is important to note that for scores 2–4 “material” may be old or new material (Table 2). In clinical videofluoroscopic swallowing studies (VFSS), swallows are often isolated; however in the infant pig videofluoroscopy recordings are made across a feeding session where there are multiple swallows per second with no break. As a result there is often milk residue in the airway from previous swallows. Since any milk in the airway above the vocal folds was due to a failed swallow, it was determined that it should be scored regardless of whether it was from a previous swallow or the current swallow.

**Aspiration**—A score of a 5 and 6 on the ISPAS is similar to a score of 6 and 7 on the clinical 8-Point PAS. A score of 6 on the clinical PAS describes material passing below the vocal folds, and then being ejected into the larynx or out of the airway. A score of 7 was where material passed below the vocal folds and was not ejected despite effort. Although we did not observe a score of 6 or 7 in our videofluoroscopic recordings, past observations of coughing in pigs with SLN lesions meant these two scores are possible. Further, aspiration following sensory or motor nerve lesions could trigger a coughing reflex that may or may not be successful. For this reason a score of 5 on the IMPAS is when material passes below the vocal folds and is ejected into the larynx or out of the airway. A score of 6 is when material passes below the vocal folds and is not ejected despite effort.

Silent aspiration is described as a score of 7 on the IMPAS (Fig. 1f). This score is the equivalent of a score 8 on the clinical 8-Point PAS. This was often seen after nerve lesions in the infant pig model.

With scores of 5, 6 and 7, *new* milk must be visualized moving from the supraglottic space to below the vocal folds. There were some instances where milk, from a previously failed swallow, was visualized in the trachea moving above, and then back below the vocal folds. This was not a score of a 5, 6 or 7 because for those scores new material must be in the supraglottic space then move below the vocal folds. If material is in the supraglottic space and joins milk residue from below the vocal folds, forming a solid stream of milk, then it is aspiration and is scored as a 5, 6 or 7 depending on whether or not there is effort to eject that material.

No equivalent score of 4 and 5 on the clinical PAS was included in the infant pig scale because milk was always clearly either above or below the vocal folds and never just contacted them.

### Guidelines for users of the scale

In order to ensure consistency and higher reliability, specific directions were developed for judges scoring videos. The swallow being scored begins at the start of the epiglottal flipping (posterior/inferior movement) and ends immediately before the next epiglottal posterior/inferior movement. The rationale for this definition was that in some cases the events leading to a score of 3, 4 or 7 did not develop until after the epiglottis had returned to its upright position, but before the next swallow occurred. Judging was also based only on a

score where material was clearly visible in the airway. Sometimes due to the poor quality of the video, judges would describe seeing very small residues of material in the airway. This resulted in variable scoring of the videos. Judges were not permitted to alter the contrast or brightness of the image since that could alter the amount of material visualized or create artifacts. When scoring videos, judges were given as much time to review the swallow as needed and they could view it multiple times and in slow motion.

### **Determining the reliability of the scale**

In order to test the reliability of the IMPAS, five judges scored 90 videofluoroscopy recordings. Images of swallows were randomly selected from 10 infant pigs and 30 total feeding sessions from the previously described ongoing studies. From each feeding session, three swallows were randomly clipped out of the sequence and separated into their own video file.

The judges had various levels of experience and education with respect to analysis of VFSS data, but all had experience with either animal or clinical swallowing research. The judges were given the 90 videofluoroscopic recordings of individual swallows in a randomized order to score according to the scale. They were instructed to score all of the videos within 48 hours and then, as was done with the development of the clinical PAS, to score them all again after a two week hiatus, also within a 48 hour period (2). The judges were given the same set of videos for their second scoring, but in a different, randomized order. All videos were viewed using MaxTRAQ Version 2.2.4.1 (Innovision Systems, Inc.). All judges also were given the same five videos that were examples of scores 1, 2, 3, 4 and 7 to reference during their scoring of the videos.

In order to assess inter- and intra-rater reliability, Cohen's kappa and two-way mixed intraclass correlation coefficient (ICC) with absolute agreement were calculated using SYSTAT 13 (SYSTAT, 2009) and IBM SPSS Statistics 20, respectively. Cohen's kappa was also calculated by scale score in order to determine if there were differences in intra- and inter-rater reliability based on each score. Rosenbek et al 1996 used kappa to calculate reliability, however, we also calculated ICC since the scale is ordinal and that calculation takes into account the degree of difference. Percentage of agreement and the stratification of the scores were also calculated. We also tested for statistically significant differences in the Cohen's kappa score given for videos captured at 30 vs. 60 frames per second using an Analysis of Variance test (ANOVA) and post-hoc Tukey's test. The statistical analysis was carried out by an independent investigator who had not judged the videos.

### **Assessing validity of the scale**

A separate study was conducted to assess if the scale could distinguish normal and abnormal pigs. A blinded judge, who was not one of the five judges used to assess reliability, scored 39 swallows from six intact infant pigs and 39 swallows from three abnormal infant pigs. The abnormal pigs had a unilateral SLN lesion. Each swallow was from a different randomly selected feeding session. The scores were graphed to see if there was a distinct difference in those between normal and abnormal infant pigs. A two sample t-test was calculated to determine if the normal and abnormal pigs were statistically significantly different with an alpha set at 0.05.

## **RESULTS**

### **Reliability of the scale**

In an initial attempt to score the 90 videos using the scale, the intra- and inter-rater reliability was low and was deemed unacceptable. The intra-rater reliability measured using Cohen's

kappa averaged 0.65 and the inter-rater reliability ranged from 0.36–0.67 with an average of 0.58. Following that preliminary analysis, difficulties and problems with the scale were discussed. Ten videos that were a source of significant variation were reviewed in our group in order to better define the seven categories. The subsequent clarification and revision to the scale resulted in the version presented here.

When the scoring was performed again, the results showed much higher inter and intra-rater reliability scores. Intra-rater reliability calculations showed an average Cohen's kappa of 0.82 and an average ICC of 0.92 with 86% agreement (Table 3). The intra-rater reliability by score was calculated using Cohen's kappa (Table 3) and showed higher reliability for scores 1, 2 and 7 (0.90, 0.82, 0.83 respectively) as compared to scores 3 and 4 (0.74, 0.75 respectively). Inter-rater reliability was, as expected, was lower than the intra-rater reliability. For the first scoring, the average Cohen's kappa value was 0.70 and the ICC was 0.89. For the second scoring the average Cohen's kappa value was 0.66 and the ICC was 0.87 (Tables 4 and 5). The Cohen's kappa values ranged from 0.65–0.76 for the first ranking and ranged from 0.58–0.84 for the second ranking. No scores of 5 or 6 were observed by the judges.

The Cohen's kappa for each score showed higher reliability for scores 1, 2, 4 and 7 (0.86, 0.68, 0.67 0.80 respectively) than for 3 (0.50) in the first scoring (Table 6). In the second scoring, reliability was lower for both scores 3 and 4 (0.46, 0.59 respectively) than for scores of 1, 2 and 7 (Table 7). An examination of the distribution and differences of scores between the first and second scoring showed that if there was not agreement, they usually only differed by one score (Table 8).

There was no significant difference in Cohen's kappa for videos captured at 30 vs. 60 frames per second ( $p=0.16$ ). When comparisons were made by score, there were no statistically significant differences in frame rate for scores 1, 2 and 4, however, there was a statistically significant difference for scores 3 and 7 ( $p<0.001$ , Fig. 2). For a score of 3, the videos captured at 60 frames per second had a significantly lower reliability than those captured at 30 frames per second (0.30 and 0.56 respectively). For a score of 7 the reliability of videos captured at 60 frames per second was significantly higher than at 30 frames per second (1.00 and 0.693 respectively).

### Validity of the Scale

When the scale was used to score 39 swallows from intact animals and 39 swallows from abnormal animals, there was a clear difference in the distribution of scores (Fig.3). In intact pigs, 61.6% scored a 1, 33.3% scored a 2, 5.1% scored a 3 and none scored 4's or 7's. Again, there were no scores of 5 or 6. This was a stark contrast to the abnormal pigs where 46.2% scored a 1, 2.6% scored a 2, 7.7% scored a 3, 2.6% scored a 4 and 41.0% scored a 7. In addition, a two sample t-test determined that the normal and abnormal pigs were significantly different ( $t=-4.89$ ,  $p<0.001$ ). The mean for normal pigs was 1.42 with a standard deviation of 0.60 and the mean for abnormal pigs was 3.72 with a standard deviation of 2.86.

## DISCUSSION

The IMPAS will allow researchers to use the infant pig model and other infant mammalian models for assessing the pathophysiology of swallowing dysfunction and outcomes of rehabilitation. The Cohen's kappa result was interpreted as follows: 0–0.20= slight, 0.20–0.40= fair, 0.40–0.60= moderate, 0.60–0.80= substantial, 0.80–1.00= almost perfect strength of agreement (21). The inter-rater reliability assessment demonstrated substantial strength of agreement and the intra-rater reliability demonstrated “almost perfect strength” of



agreement(21). The ICC for intra and inter-rater reliability were also very high (0.92 and 0.88 respectively). The inter-rater reliability by score demonstrated that scores 3 and 4 were harder to score reliably and that reliability decreased after the two week break. This underscores the importance of training of judges, especially when multiple judges are used. Whenever the scale is used, inter- and intra-rater reliability should be calculated and assessed to determine functional significance of results.

A significant difference was found in reliability of the IMPAS for videos captured at 30 vs. 60 frames per second only for scores 3 and 7. For scores of 7, although the difference was statistically significant, it was most likely not functionally significant since the reliability was still high (0.60+). For a score of 3, there was lower reliability for videos at 60 frames per second which was not expected since these videos can capture behaviors with a higher time resolution. A score of 3 may appear differently depending on the capture rate or resolution of the video. While clinical videofluoroscopic swallowing studies utilize 30 frames per second recording, animal studies are able to take advantage of the higher 60 frames per second setting that is an option on most videofluoroscopic units. This suggests that the score of 3 needs to be defined clearly to raters by having a detailed description of the size of the bolus and having examples.

The data also showed that this scale can be used to distinguish normal swallowing from abnormal swallowing. The distribution of scores for normal swallows indicates scores of both 1 and 2 are seen and do not indicate a pathological condition. A score of 2 occurred during approximately 1/3 of all swallows. Very rarely a score of 3 occurred. Although a score of 2 is penetration, it is often seen in normal pigs and may be a result of their developing coordination between sucking, swallowing and breathing although it is not observed in normal infant feeding (22). Both infant humans and pigs feed with an upright posture, so posture, or gravity, should not affect the rate of penetration. The rate of scores of 2 in the normal, healthy infant pigs is a notable difference between infant human and pig swallows. The rate of scores of 2 in infant pigs is actually more comparable to adult humans who have scores of 2 normally in about 20% of swallows and does not indicate abnormal swallowing (5). In the experience of the authors, a 4 and 7 also may occur in a normal, healthy animal, but these are extremely rare conditions. The distribution of scores for infant pigs with a unilateral SLN lesion shows more frequent scores of 4 and many scores of 7 indicating silent aspiration. This difference was statistically significant ( $p < 0.001$ ) when tested using a two-way t-test. As the infant pig model is used to model different causes of swallowing dysfunction, this scale can be used to describe the extent of airway protection.

We found that it is important for all judges to first review the data and discuss it in a group before doing any scoring in order to maximize the reliability of the results. After the judges scored the videos and the inter- and intra-rater reliability scores were low, we had further training based on that data in order to achieve high inter- and intra-rater reliability. After concluding the assessment of inter- and intra-rater reliability, a schematic was developed to help train judges that use this scale in the future (Fig. 4). Because the IMPAS is a multidimensional scale, it may help judges to score the videos.

It is important to note that this scale does not quantify all possible events during swallowing, rather it groups them into categories based on extent of airway protection. Instead of grouping swallows as a) normal, b) having penetration or c) having aspiration, the seven categories described allow for a more precise description of the swallow observed. While this study focused on using videofluoroscopy to assess swallow function, it should be emphasized that videofluoroscopy should be used along with other methods.

Videofluoroscopy is essential for visualizing penetration and aspiration, however, ultrasound and electromyography are other important tools that can allow researchers to understand the

mechanisms of swallowing dysfunction in animal models (23–25). Endoscopy, a valuable research tool in humans, is not possible in many animal models because of the extensive nasal conchae.

There are many opportunities for swallowing research to advance by using the 8-Point PAS for the infant pig model. This scale could be adapted for other infant mammal models such as cats, rabbits and monkeys that are already being used to study feeding and swallowing (26–28). Mammals, both infants and adults, share a common pharyngeal anatomy with an intranasal larynx and a soft palate and epiglottis that contact (16). Despite differences in chewing and oral transport, it is expected that the same scores described in the IMPAS would also be seen in other mammals since they share a common pharyngeal anatomy. Using this scale, new studies can be designed to further understand swallowing neurophysiology by using a pathological animal model. Along with other advanced technology we can further understand what causes penetration and aspiration.

## Acknowledgments

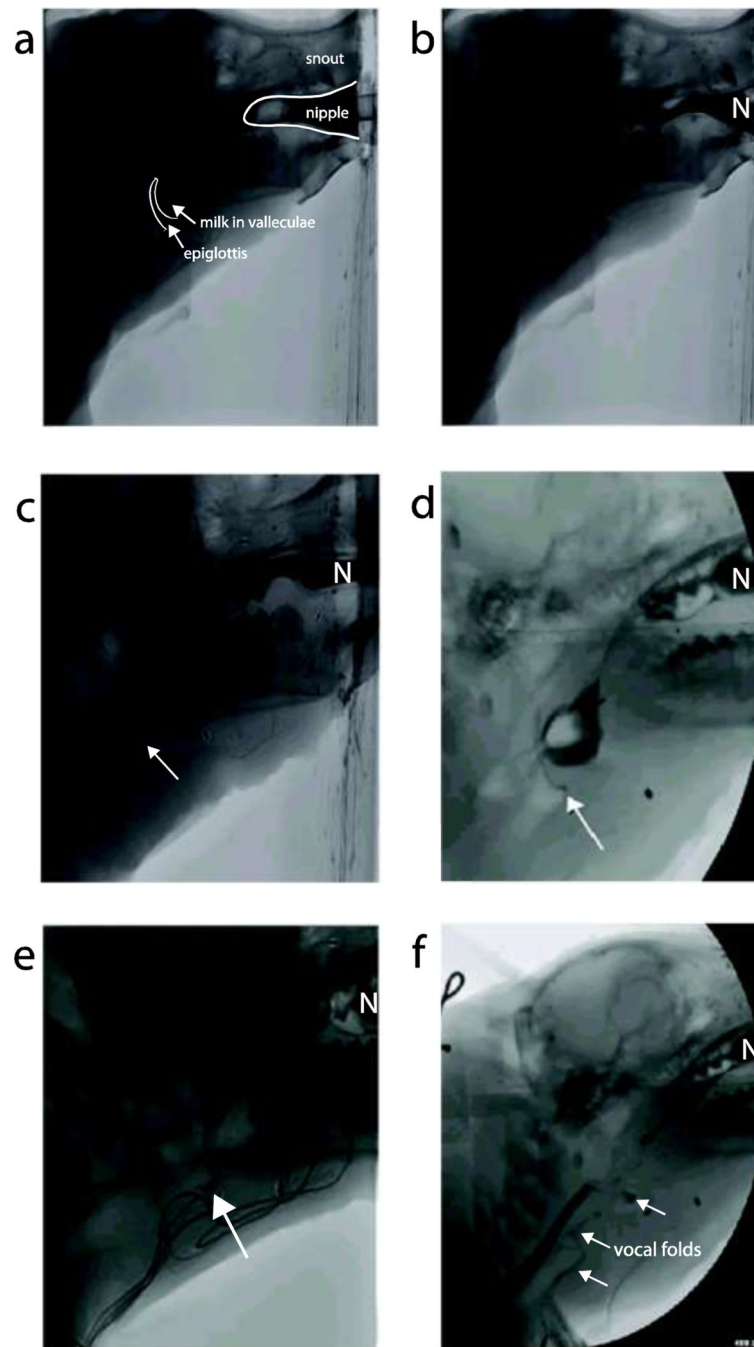
We would like to acknowledge the Johns Hopkins University School of Medicine animal support staff which includes Dr. Dawn Ruben, Dr. Eric Hutchinson, Kristy Koenig and Melanie Albano. We would also like to thank Dr. Ianessa Humbert for her helpful review of this manuscript. This project was funded by a 2012 American Association for Dental Research Student Research Fellowship, T32 DE07309 to the University of Maryland School of Dentistry and National Institutes of Health DC03604 to RZG. The authors of this paper will provide copies of all data upon request.

## Bibliography

1. Pikus L, Levine M, Yang Y-X, Rubesin S, Katzka D, Laufer I, Geftter W. Videofluoroscopic studies of swallowing dysfunction and the relative risk of pneumonia. *AJR Am J Roentgenol*. 2003; 180:1613–1616. [PubMed: 12760930]
2. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996; 11:93–98. [PubMed: 8721066]
3. Rofes L, Arreola V, Almirall J, Cabre M, Campins L, Garcia-Peris P, Speyer R, Clave P. Diagnosis and management of oropharyngeal dysphagia and its nutritional and respiratory complications in the elderly. *Gastroenterol Res Pract*. 2011 2010 Aug 3 [Epub ahead of print].
4. Bulter S, Stuart A, Leng X, Wilhelm E, Rees C, Williamson J, Kritchevsky S. The Relationship of Aspiration Status with Tongue and Handgrip Strength in Healthy Older Adults. *J Gerontol A Biol Sci Med Sci*. 2011; 66A:452–458.
5. Robbins J, Coyle J, Rosenbek J, Roecker E, Wood J. Differentiation of normal and abnormal airway protection during swallowing using the penetration-aspiration scale. *Dysphagia*. 1999; 14:228–232. [PubMed: 10467048]
6. Baijens L, Speyer R, Passos V, Pilz W, Roodenburg N, Clave P. Swallowing in Parkinson Patients versus Healthy Controls: Reliability of Measurements in Videofluoroscopy. *Gastroenterol Res Pract*. 2011 Oct 3. [Epub ahead of print].
7. Hey C, Pluschinski P, Stanschus S, Euler HA, Sader RA, Langmore S, Neumann K. A Documentation System to Save Time and Ensure Proper Application of the Fiberoptic Endoscopic Evaluation of Swallowing (FEES). *Folia Phoniatria et Logopaedica*. 2011; 63:201–208. [PubMed: 20938202]
8. Ciucci M, Russell J, Schaser A, Doll E, Vinney L, Conner N. Tongue force and timing deficits in a rat model of Parkinson disease. *Behavioural Brain Research* Epub ahead of print. 2011
9. Lever T, Simon E, Cox K, Capra N, O'Brien K, Hough M, Murashov A. A Mouse Model of Pharyngeal Dysphagia in Amyotrophic Lateral Sclerosis. *Dysphagia*. 2012; 25:112–126. [PubMed: 19495873]
10. Watrous BJ, Suter PF. Oropharyngeal dysphagias in the dog: A cinefluorographic analysis of experimentally induced and spontaneously occurring swallowing disorders. *Veterinary Radiology*. 1983; 24:11–24.



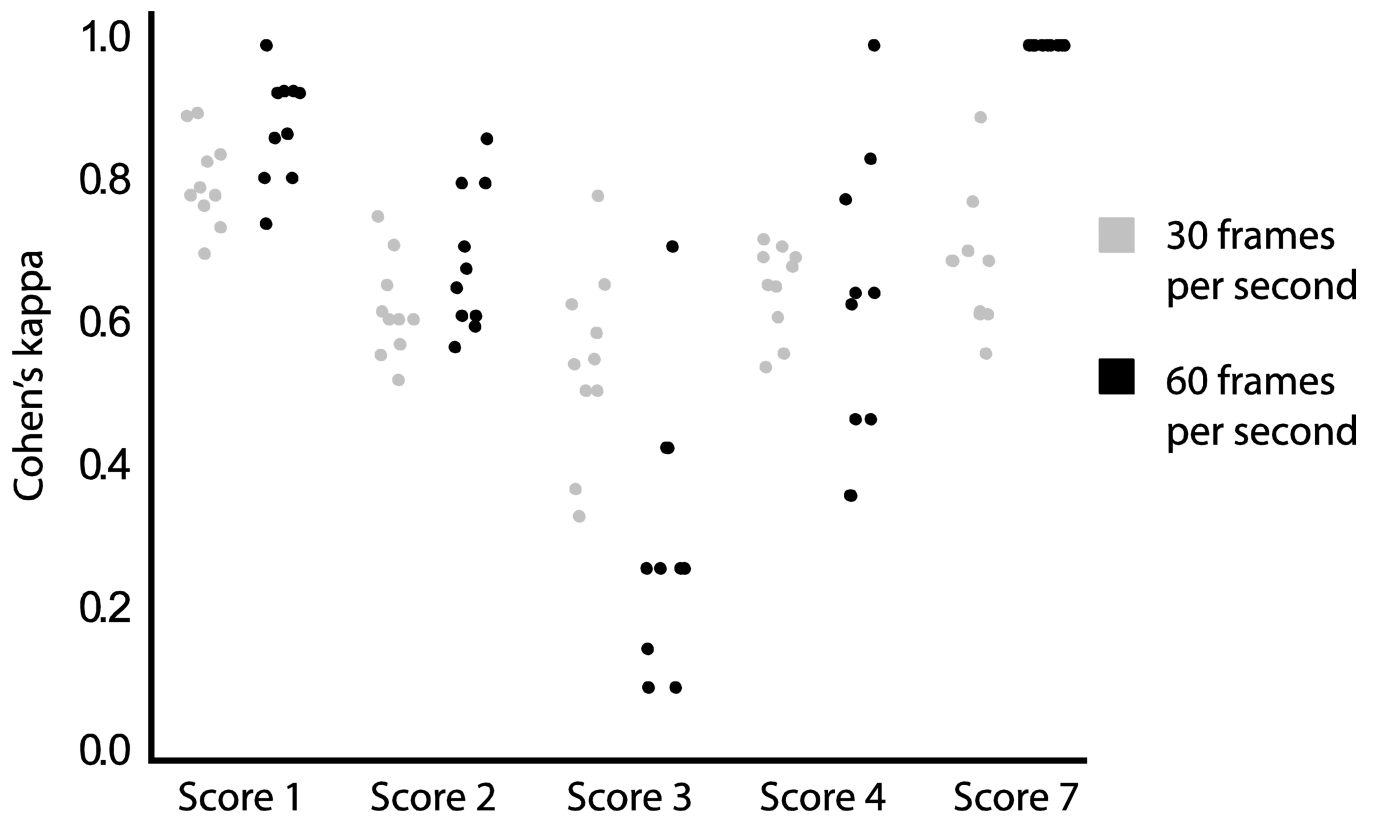
11. German RZ, Crompton AW, Thexton AJ. Integration of the reflex pharyngeal swallow into rhythmic oral activity in a neurologically intact pig model. *J Neurophysiol.* 2009; 102:1017–1025. [PubMed: 19515957]
12. Klein BG, Duffin JR, Kraje B. Neonatal infraorbital nerve damage and the development of eating behavior in the rat. *Behavioural Brain Research.* 1994; 60:25–33. [PubMed: 8185849]
13. Bennett SW, Lanovaz JL, Muir GD. The biomechanics of locomotor compensation after peripheral nerve lesion in the rat. *Behavioural Brain Research.* 2012; 229:391–400. [PubMed: 22309981]
14. Aluffi P, Policarpo M, Cherovac C, Olina M, Dosdegani R, Pia F. Post-thyroidectomy superior laryngeal nerve injury. *European Archives of Oto-Rhino-Laryngology.* 2001; 258:451–454. [PubMed: 11769989]
15. Kark AE, Kissin MW, Wenig BI. Superior laryngeal nerve injury to the editor. *Head & Neck.* 1995; 17:542–543. [PubMed: 8847214]
16. Hiimeae, KM. Feeding in Mammals. In: Schwenk, K., editor. *Feeding: Form, Function and Evolution in Tetrapod Vertebrates.* San Diego: Academic Press; 2000. p. 411–448.
17. Ding, P.; Campbell-Malone, R.; Wahl, S.; Fukuhara, T.; Feng, W.; German, R. *Aspiration Mechanism for Unilateral Superior Laryngeal Nerve Lesion in Animal Model.* Dysphagia Research Society Annual Meeting; San Antonio, TX. 2011.
18. Ding, P.; Chen, Y.; Lukasik, S.; German, R. *Mechanism of Epiglottal Movement during the Infant Pharyngeal Swallow.* Dysphagia Research Society Annual Meeting; Toronto, ON Canada. 2012.
19. Reuss-Lamky H. Administering dental nerve blocks. *J Am Anim Hosp Assoc.* 2007; 43:298–305. [PubMed: 17823481]
20. Friedman B, Frazier JB. Deep laryngeal penetration as a predictor of aspiration. *Dysphagia.* 2000; 15:153–158. [PubMed: 10839829]
21. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
22. Newman L, Cleveland R, Blickman J, Hillman R, Jaramillo D. Videofluoroscopic analysis of the infant swallow. *Invest Radiol.* 1991; 26:870–873. [PubMed: 1960027]
23. Macrae PR, Doeltgen SH, Jones RD, Huckabee M-L. Intra- and inter-rater reliability for analysis of hyoid displacement measured with sonography. *Journal of Clinical Ultrasound.* 2012; 40:74–78. [PubMed: 21953135]
24. Kuo P, Holloway RH, Nguyen NQ. Current and future techniques in the evaluation of dysphagia. *Journal of Gastroenterology and Hepatology.* 2012; 27:873–881. [PubMed: 22369033]
25. Konow N, Thexton AJ, Crompton AW, German RZ. Regional differences in length-change and electromyographic heterogeneity in the sternohyoid muscle during infant mammalian swallowing. *Journal of Applied Physiology.* 2010; 109:439–448. [PubMed: 20538841]
26. Iriarte-Diaz J, Reed DA, Ross CF. Sources of Variance in Temporal and Spatial Aspects of Jaw Kinematics in Two Species of Primates Feeding on Foods of Different Properties. *Integrative and Comparative Biology.* 2011; 51:307–319. [PubMed: 21719431]
27. Tsujimura T, Yamada A, Nakamura Y, Fukuhara T, Yamamura K, Inoue M. The Digastric Muscle is Less Involved in Pharyngeal Swallowing in Rabbits. *Dysphagia.* 2011:1–6.
28. Medda B, Lang I, Dodds W, Christl M, Kern M, Hogan W, Shaker R. Correlation of electrical and contractile activities of the cricopharyngeus muscle in the cat. *Am J Physiol.* 1997; 273:G470–G479. [PubMed: 9277427]



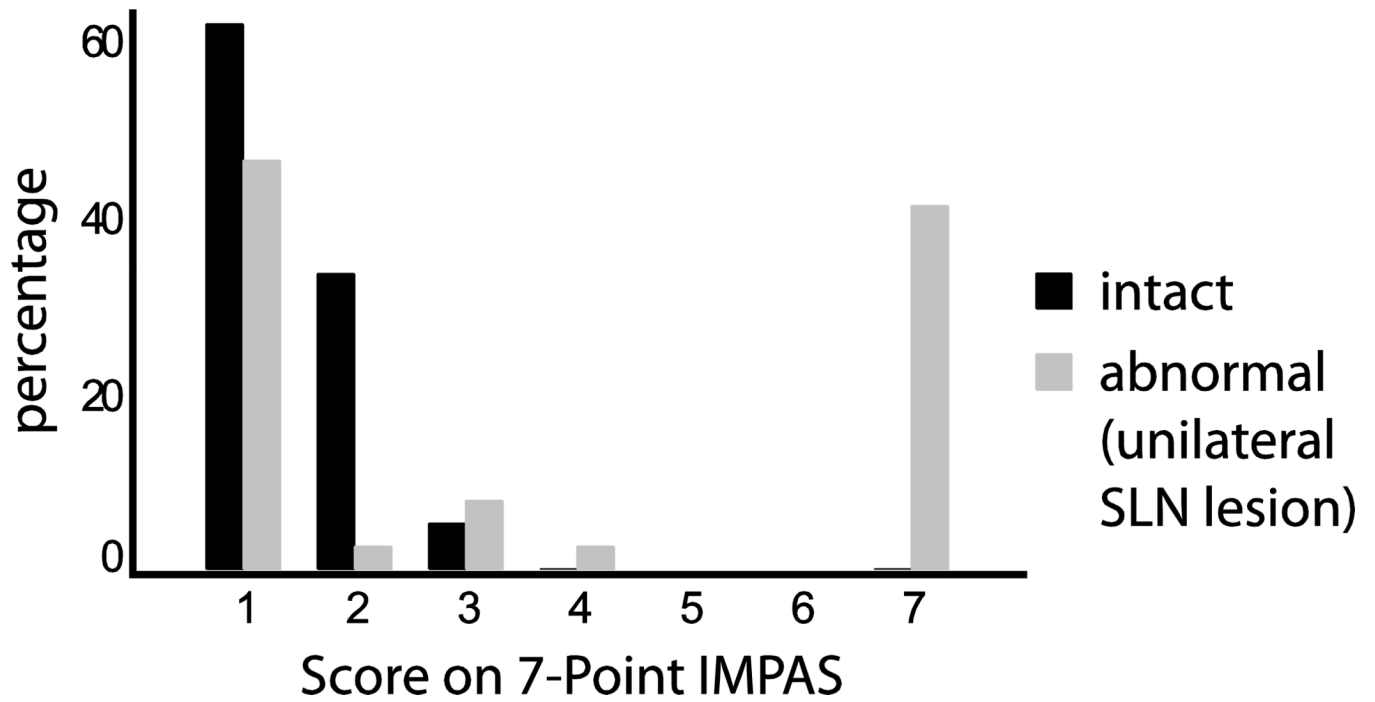
**Fig. 1. Videofluoroscopic images from each score on the 7-Point Infant Mammal PAS**

a) This videofluoroscopic image shows the epiglottis is in the upright position with milk in the valleculae right before a swallow is initiated. The anatomy is clearly outlined and labeled for orientation. b) In the image, milk is being emptied out of the valleculae, the epiglottis is fully flipped to protect the airway, and milk continues into the esophagus. This is a score of 1. c) The epiglottis has flipped fully to protect the airway, however, milk penetrates the airway as identified by the arrow. When the epiglottis returns to its rest position, there is no milk left in the supraglottic space. This is a score of 2. d) The epiglottis is in its upright position after a swallow. The arrow points to a small amount of milk visible on the caudal side of the epiglottis forming a triangle at the base. When this occurs it is scored as a 3. e)

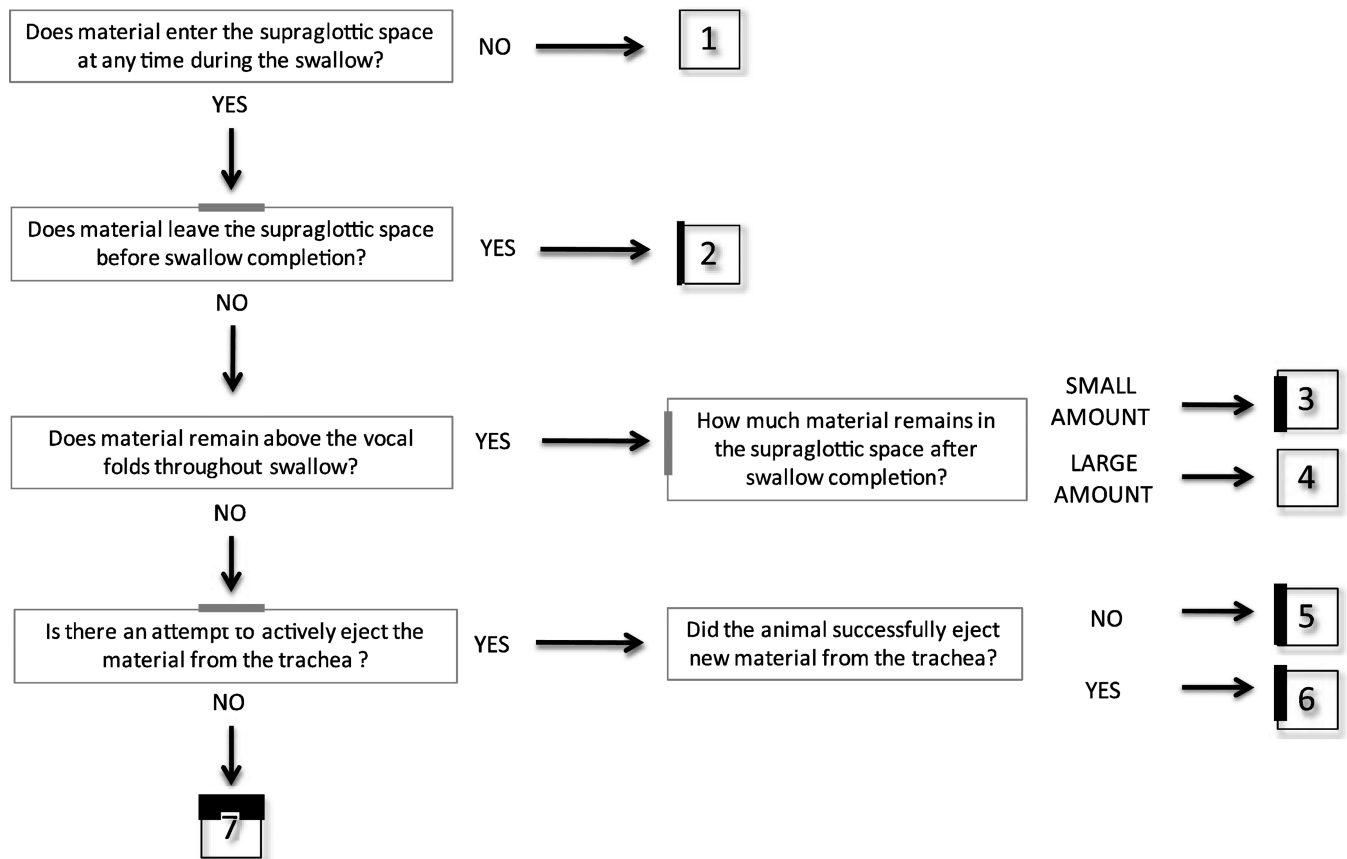
The epiglottis has returned to its rest position after a swallow. The arrow points at a large amount of milk residue left on the caudal side of the epiglottis, but above the vocal folds. This is scored as a 4. f) In the image, the epiglottis is flipped to protect the airway during the swallow. Milk is clearly visible crossing the vocal folds and entering the larynx and is identified by the arrows. The nipple is in the far right of all the images and is labeled with the letter “N”.



**Fig. 2. Distribution of intraclass correlation coefficients (ICCs) for videos captured at 30 and 60 frames per second by 7-point infant mammalian penetration-aspiration scale score**  
 The figure shows the distribution of ICCs, by score and by video capture rate. There is no significant difference in ICC between videos captured at 30 and 60 frames per second for scores 1, 2 and 4. There is a significant difference for scores 3 and 7 ( $p < 0.001$ ).



**Fig. 3. Distribution of scores given to control animals and animals with a unilateral superior laryngeal nerve lesion**  
 The graph shows a clear distinction between scores given to normal and abnormal animals.



**Fig. 4. Schematic representation of the 7-Point Infant Mammal Penetration-Aspiration Scale**  
 This schematic can be used by the judges as a systematic way to approach scoring the videos.



**Table 1**  
**8-Point Penetration-Aspiration Scale from Rosenbek et al 1996**

Score	Description
1	Material does not enter the airway
2	Material enters the airway, remains above the vocal folds, and is ejected from the airway
3	Material enters the airway, remains above the vocal folds, and is not ejected from the airway
4	Material enters the airway, contacts the vocal folds, and is ejected from the airway
5	Material enters the airway, contacts the vocal folds, and is not ejected from the airway
6	Material enters the airway, passes below the vocal folds and is ejected into the larynx or out of the airway
7	Material enters the airway, passes below the vocal folds, and is not ejected from the trachea despite effort
8	Material enters the airway, passes below the vocal folds, and no effort is made to eject

**Table 2**  
**The 7-Point Infant Mammal Penetration-Aspiration Scale**

Score	Description
1	Material does not enter the airway
2	Material is in the supraglottic space, remains above the vocal folds, and passively leaves the airway before the epiglottis returns to rest position
3	Material is in the supraglottic space, a small amount remains above the vocal folds after epiglottis in rest position
4	Material is in the supraglottic space, a larger amount remains above the vocal folds after epiglottis in rest position
5	New material is in the supraglottic space, then passes below the vocal folds, and is actively ejected, above the vocal folds
6	New material is in the supraglottic space, then passes below the vocal folds and is not ejected from the trachea despite effort
7	New material is in the supraglottic space, then passes below the vocal folds, and no effort is made to eject (silent aspiration)

**Table 3****Intra-rater reliability by score**

Percentage of agreement is shown for each judge between the first time they did the scoring and when they did scoring two weeks later. Cohen's kappa is also calculated overall and by score for each judge.

Judge	Agreement		ICC	Cohen's kappa	Cohen's kappa by scale score						
	n	%			1	2	3	4	5	6	7
1	75	83.33	0.87	0.78	0.90	0.82	0.68	0.65	-	-	0.71
2	79	87.78	0.97	0.84	0.80	0.80	0.84	0.85	-	-	0.96
3	82	91.11	0.96	0.89	0.92	0.87	0.86	0.87	-	-	0.88
4	79	87.78	0.95	0.84	0.91	0.85	0.73	0.82	-	-	0.86
5	72	80.00	0.87	0.74	0.94	0.77	0.57	0.58	-	-	0.73
AVERAGE	77.4	86.00	0.92	0.82	0.90	0.82	0.74	0.75	-	-	0.83

**Table 4**

**Inter-rater reliability for each pair of judges at first scoring**

This round of scoring was completed within one week of a training session for the judges where examples were reviewed. The average Cohen's kappa was 0.70 and the average percentage of agreement was 76.7%. Most scores that did not agree differed by only one score.

Judge pair	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5	Average
Two scores agree											
n	66	69	70	71	67	67	67	70	73	70	69.0
%	73.3	76.7	77.8	78.9	74.4	74.4	74.4	77.8	81.1	77.8	76.7
Number of scores that differ by											
1	20	17	18	19	18	22	21	20	16	19	19.0
2	3	4	1	0	5	1	1	0	1	0	1.6
3	1	0	1	0	0	0	1	0	0	1	0.4
>3	0	0	0	0	0	0	0	0	0	0	0.0
Cohen's Kappa	0.65	0.70	0.71	0.73	0.67	0.67	0.67	0.71	0.76	0.71	0.70

**Table 5**

**Inter-rater reliability for each pair of judges at second scoring**

This scoring was completed two weeks after the first scoring. The average Cohen's kappa is lower, but still acceptable at 0.66 with a percentage of agreement of 74.2%. As before, most scores that were not in agreement differed by one score.

Judge pair	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5	average
Two scores agree											
n	66	69	68	71	62	63	61	78	65	65	66.8
%	73.3	76.7	75.6	78.9	68.9	70.0	67.8	86.7	72.2	72.2	74.2
Number of scores that differ by											
1	20	19	22	14	22	24	24	11	20	21	19.7
2	4	2	0	3	5	3	3	1	4	3	2.8
3	0	0	0	2	1	0	2	0	1	1	0.7
>3	0	0	0	0	0	0	0	0	0	0	0.0
Cohen's Kappa	0.65	0.70	0.68	0.72	0.60	0.61	0.58	0.83	0.64	0.64	0.66

Table 6

**Interjudge Cohen's kappa by scale score for first scoring**

The interjudge kappa by score shows high reliability (0.60+) for scores, except for a score of 3 which is more variable and has a lower average Cohen's kappa value. There were no scores of 5 or 6 observed.

Judge Pair	Scale score						
	1	2	3	4	5	6	7
1-2	0.94	0.77	0.57	0.58	-	-	0.73
1-3	0.90	0.64	0.46	0.63	-	-	0.81
1-4	0.83	0.67	0.46	0.75	-	-	0.81
1-5	0.86	0.70	0.58	0.70	-	-	0.75
2-3	0.84	0.64	0.49	0.60	-	-	0.73
2-4	0.88	0.62	0.31	0.71	-	-	0.73
2-5	0.85	0.59	0.44	0.66	-	-	0.79
3-4	0.84	0.72	0.49	0.65	-	-	0.86
3-5	0.81	0.69	0.68	0.75	-	-	0.94
4-5	0.85	0.72	0.52	0.65	-	-	0.81
average	0.86	0.68	0.50	0.67	-	-	0.80



Table 7

**Interjudge Cohen's kappa by scale score for second scoring**

The interjudge Cohen's kappa by score shows high reliability (0.60+) for scores, except for a score of 3 and 4 which are more variable and have lower average Cohen's kappa value. This indicates that the training may not be retained over two weeks and is important for scoring 3's and 4's. There were no scores of 5 or 6 observed.

Judge Pair	Scale score						
	1	2	3	4	5	6	7
1-2	0.77	0.61	0.45	0.52	-	-	0.81
1-3	0.92	0.65	0.33	0.58	-	-	0.89
1-4	0.82	0.62	0.40	0.68	-	-	0.88
1-5	0.77	0.64	0.75	0.71	-	-	0.81
2-3	0.79	0.49	0.30	0.53	-	-	0.81
2-4	0.78	0.56	0.38	0.56	-	-	0.69
2-5	0.72	0.48	0.49	0.55	-	-	0.63
3-4	0.89	0.77	0.75	0.84	-	-	0.88
3-5	0.84	0.63	0.37	0.45	-	-	0.68
4-5	0.89	0.65	0.35	0.47	-	-	0.78
Average	0.82	0.61	0.46	0.59	-	-	0.78

**Table 8**  
**Distribution of first and second grading scores for all raters collectively**

The table shows the frequency of scores at the first scoring and two weeks later at the second scoring. No scores of 5 or 6 were observed.

	first grading score							Total	% of total	% agree
	1	2	3	4	5	6	7			
1	75	4						79	29.3	94.9
2	9	68	1				1	79	29.3	86.1
3	1	3	29	4				37	13.7	78.4
4		1	3	36			5	45	16.7	80.0
5					0			0	0	0
6						0		0	0	0
7			1	1			28	30	11.1	93.3
Total	85	76	34	41	0	0	34	270		
% of total	31.5	28.1	12.6	15.2	0	0	12.6			
% agree	88.2	89.5	85.3	87.8	0	0	82.35			