



Published in final edited form as:

Disabil Rehabil. 2019 March ; 41(5): 549–559. doi:10.1080/09638288.2017.1398787.

Construct Validity of the Eating Assessment Tool

Janina Wilmskoetter¹, Heather Bonilha^{1,2}, Ickpyo Hong³, Jordan Hazelwood⁴, Bonnie Martin-Harris⁵, and Craig Velozo⁶

¹Department of Health Sciences and Research, College of Health Professions, Medical University of South Carolina, Charleston (SC), USA

²Department of Otolaryngology – Head and Neck Surgery, Medical University of South Carolina, Charleston (SC), USA

³Division of Rehabilitation Sciences, School of Health Professions, University of Texas, Medical Branch, Galveston (TX), USA

⁴Department of Communication Sciences and Disorders, Beaver College of Health Sciences, Appalachian State University, Boone (NC), USA

⁵Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, School of Communication, Northwestern University, Evanston (IL), USA

⁶Division of Occupational Therapy, Department of Health Professions, College of Health Professions, Medical University of South Carolina, Charleston (SC), USA

Abstract

Purpose—We aimed to evaluate the construct validity of the Eating Assessment Tool by determining its dimensionality, rating scale integrity, item-person match, precision, and relationship with the degree of airway invasion and functional oral intake.

Methods—We conducted a retrospective analysis of patients' Eating Assessment Tool scores. We used the Rasch rating scale model. We investigated correlations between the Eating Assessment Tool and scores on the Penetration-Aspiration Scale and Functional Oral Intake Scale.

Results—The median score of the Eating Assessment Tool from 127 patients was 16 of 40 (range 0 to 40). Confirmatory factor analysis supported unidimensionality. The 5-point rating scale categories met published criteria. Two items misfit the Rasch model and two other items displayed differential item functioning. Rasch person reliability was 0.79. Our patient cohort was divided into three person-strata. Correlations between the Eating Assessment Tool and the Penetration-Aspiration Scale and Functional Oral Intake Scale were weak to moderate in strength (respectively: $r = 0.26$, $p = 0.0036$; $r = -0.27$, $p = 0.0027$).

Correspondence information: Janina Wilmskoetter, Department of Health Sciences & Research, College of Health Professions, Medical University of South Carolina, 77 President St. MSC 700, Charleston, SC 29425, USA, Phone: (+1) 843-452-0670, Fax: (+1) 843-792-1358, wilmskoe@musc.edu.

Declaration of interest:

All authors report they have no other declaration of interest.

Conclusions—Our analyses identified deficits in the construct validity of the Eating Assessment Tool suggestive of a need to improve the Eating Assessment Tool to support its frequent use in clinical practice and research.

Keywords

Rasch model; deglutition disorder; outcome assessment (health care); surveys and questionnaires; reliability; validity

Introduction

Dysphagia, or trouble with swallowing, has a prevalence between 2.3% and 16% in the general population [1] and an even higher prevalence of 51% in institutionalized older people [2]. Dysphagia is associated with mortality [3], aspiration pneumonia [4, 5], malnutrition [6], dehydration [7], and significantly increased medical expenses [8]. Moreover, this potentially life-threatening disorder represents a social and psychological burden for patients with adverse effects on self-esteem, socialization, and enjoyment of life [9].

The Eating Assessment Tool (EAT-10) was developed to evaluate patients' self-perceived symptoms of dysphagia to estimate initial severity and monitor change in dysphagia symptoms as a response to treatment [10]. The instrument is frequently used in clinical practice and research for patients with suspected swallowing problems stemming from various etiologies as it is quick to administer and easy to understand. The EAT-10 consists of 10 items. Each item includes a statement describing a scenario that may be a problem for patients experiencing swallowing difficulty (e.g. "Swallowing liquids takes extra effort" or "The pleasure of eating is affected by my swallowing problem"). The EAT-10 has been translated into several languages, including Portuguese, Italian, Swedish, and Arabic [11, 12, 13, 14]. Its psychometric properties have been tested using classical test theory (CTT) to determine the quality of the instrument [10]. In addition to having an excellent internal consistency (Cronbach's alpha = 0.96), test-retest reproducibility ($r = 0.72 - 0.91$), and criterion-based validity (all items on the EAT-10 showed a highly significant improvement with intervention) [10], the EAT-10 also showed a positive linear correlation with aspiration events in patients with dysphagia ($r = 0.273$, $p < 0.001$) [15] as measured with the Penetration-Aspiration Scale (PAS) [16]. This led to the conclusion that the EAT-10 is a suitable instrument for predicting aspiration risk [15].

However, using CTT has disadvantages when assessing an instrument's psychometric properties at the item-level and person-level. For example, CTT assesses an instrument as a whole by using an instrument's total score based on the summation of the individual item scores and, therefore, CTT cannot determine the fit of items or persons to a measurement model [17]. Thus, CTT cannot identify erratic items or persons, nor can it readily detect differential item functioning across subgroups (e.g., gender or medical diagnosis). Therefore, the consequence of not detecting if an instrument shows DIF by using CTT instead of the Rasch model leads to a false assumption that the instrument is invariant, objective and stable regarding its measurements. This may lead to a biased measurement

estimation of the instrument's quality and incorrect interpretations of its results. The EAT-10 has been criticized for limitations in its psychometric evaluation [18] that in part stem from its development based on CTT. To the best of our knowledge, there have been no psychometric studies of the English version of the EAT-10 on properties of construct validity, such as: dimensionality, rating scale integrity, precision, or relationship to instruments used to assess swallowing disorders other than the PAS [10, 18]. As we describe in the following, an understanding of these characteristics of the EAT-10 is crucial to justify its use in research and clinical practice.

Unidimensionality of an instrument is important in order to interpret a (total) test score. If items of an instrument do not represent the same underlying construct, the meaning of a (total) test score is distorted [19]. Therefore, without unidimensionality, a total score, which is the summation of scores of several distinct constructs introduces theoretical imprecision [19]. For clinical applications, unidimensionality of a measurement instrument is important, because it ensures that only one measurement construct is measured at a time, thus, clinicians can make valid inferences about the measurement score and the underlying construct that has contributed to the specific score. This is especially important in instances where clinicians seek to compare scores between patients or for the same patient at different time points.

Similar to unidimensionality, integrity of an instrument's rating scale is a second important property to make valid inferences from patients' scores for each item and for the total score. The EAT-10 uses the same 5-point rating scale for each of its 10 items, but it remains unknown whether all five possible ratings are actually used by patients. Further, it is not known if the ratings are used in an expected manner, that is patients with less perceived consequences of dysphagia are more likely use lower ratings on an item and patients with more perceived consequences are more likely use higher ratings [20, 21].

A third property of the EAT-10 important for interpreting test scores that has not been assessed is precision. In the Rasch model, precision is defined as the degree to which measures can be replicated, thus, how reliable measures are. Precision of an instrument is especially important in order to compare total test scores. The EAT-10 was developed to estimate the initial severity of patients' self-perceived dysphagia symptoms and monitor symptom changes as a response to treatment [10]. Thus, understanding the precision of the EAT-10 can help clinicians to estimate how well the instrument can differentiate between patients with different severity levels of self-perceived symptoms and how well the instrument can detect change over time within a patient assessed at different time points.

In addition to understanding an instrument's dimensionality, rating scale integrity and precision, knowing the relationship of an instrument with other outcome scales is foundational in order to draw conclusions from the test score. The relationship of the EAT-10 to the PAS and Functional Oral Intake Scale (FOIS) [22] could provide a better understanding of how patients perceive the consequences of having dysphagia, the degree of airway invasion they experience, and the impact of their recommended oral intake status. By understanding the relationship of the EAT-10 and PAS, and EAT-10 and FOIS, clinician

better interpret the contributions and limitations of the EAT-10 as compared to other dysphagia assessments.

The combined knowledge of the dimensionality, precision, rating scale integrity and relation of the EAT-10 to other swallow-specific scales will help clinicians to better estimate the informative value of the EAT-10 when it is used as part of an initial assessment of dysphagia or for monitoring treatment responses. A psychometric evaluation targeting the aforementioned properties of the instrument, is needed to improve our understanding of the usefulness of the EAT-10 in clinical and research settings.

A method that allows for the evaluation of an instrument's item-level psychometric properties, rating scale integrity, and precision, is the Rasch model. The Rasch model is a one-parameter logistic model within the tradition of the item response theory. The Rasch model is being increasingly applied to assess instruments used in rehabilitation [23, 24, 25, 26]. One of its advantages is that it transforms ordinal counts into linear objective measures. According to the Institute of Objective Measurement (2001) objective measurement can be defined as "the repetition of a unit amount that maintains its size, within an allowable range of error, no matter which instrument, intended to measure the variable of interest, is used and no matter who or what relevant person or thing is measured". Thus, measures are the basis to compare, exchange and interpret quantitative values [27]. The Rasch model provides its users (e.g. clinicians) with the opportunity to estimate patients' ability levels based on true measurement properties. The Rasch model places persons as well as items on a single linear continuum and, thereby, allows the determination of what items on the instrument are likely or unlikely to be challenging for a person [20, 28]. Thus, the Rasch model considers measurement aspects that CTT is unable to assess, such as the hierarchy of an instrument's items regarding their difficulty (item difficulty hierarchy) and model fit [20, 29]. Also, extensions of the Rasch model have been developed that allow users to assess rating scales, for example Likert scales, and determine their integrity. These model extensions are the rating scale model developed by Andrich [30] and the partial credit model developed by Masters [31]. The rating scale model by Andrich was developed to assess the integrity of a rating scale across all items of a tool, thus, assuming that each item uses the same rating scale [30] – which is the case for the EAT-10. The partial credit model has been described as an extension to Andrich's rating scale model. The partial credit model assesses the rating scale of each item separately, and, thus, can be applied to tools that include items with different rating scales [31]. Furthermore, the Rasch model evaluates an instrument in terms of its ability to measure a person. This means that the Rasch model evaluates if the measure or the total score are precise enough to compare different persons or the same person at different time points.

Through its unique capability to construct linear measures from counts of qualitatively-ordered observations – provided the structure of quantity is present in the data –, the Rasch model provides the opportunity to assess the relationship of an instrument with other instruments based on linear measures and not only raw scores. Therefore, we can use the measures constructed by the Rasch model to determine the relationship of the EAT-10 with other swallow-specific scales, such as the PAS and FOIS.

Besides the psychometric evaluation of an instrument's integrity, the Rasch model has several immediate benefits from a clinical perspective, for example: 1) it provides information on both the validity of scores of individual items and the total test measure, 2) it provides a hierarchy for items based on their difficulty (or for the EAT-10 based on the likelihood that items will be reported as a problem by the patient), and 3) it analyses how many person/patient strata can be separated by using the instrument, thus, providing information on an instrument's precision to differentiate severity levels.

To our knowledge, there have been no studies using the Rasch model to assess the psychometric properties of the English version of the EAT-10, but there has been a recent study applying the Rasch model to the Turkish, Spanish, Italian and Swedish translations of the EAT-10 to assess its cross-cultural validity [32]. The study revealed significant weaknesses of the translated versions of the EAT-10 in terms of item redundancy, floor effects, disordered rating scale categories, and DIF. DIF was found for various items for patients who differed by: gender, confirmed oropharyngeal dysphagia, instrumental swallow assessments, language (Italian, Swedish, Spanish, Turkish), and comorbid diagnosis.

The goal of our study is to provide clinicians and researchers with information on the psychometric quality and informative value of the English version of the EAT-10 [10]. We utilized the Rasch model to assess the construct validity of the EAT-10 by determining: dimensionality, rating scale integrity, item-level psychometrics (item fit statistics, person fit statistics, item hierarchy, item-person match), precision and reliability, and the relationship of the EAT-10 with the degree of airway invasion, measured by the PAS, and the functional oral intake status, measured by the FOIS. Based on findings from previous studies [15, 32], we hypothesized that the EAT-10 shows weaknesses in its construct validity.

Methods

Instruments

We used three instruments in the study presented here. The EAT-10 was the primary instrument we assessed using the Rasch model. Additionally, we used the PAS and the FOIS to assess DIF and associations with the EAT-10.

The EAT-10 is a self-report instrument consisting of 10 items. The patient is instructed to circle the appropriate rating that matches the extent of their swallowing problem on a 5-point rating from "0" (no problem) to "4" (severe problem). The scores "1", "2", and "3" are not further defined. Item ratings are summed to a total score. A total score of "3" or higher is deemed to be abnormal [8].

The PAS is an 8-point scale used to describe the presence and depths of penetration and aspiration events, as well as whether penetrated or aspirated material is expelled from the airway [16]. PAS scores range from "1" (Material does not enter the airway) to "8" (Material enters the airway, passes below the vocal folds and no effort is made to eject). We employed the PAS to identify the worst event of airway invasion (highest score on the PAS) across all swallows performed during a modified barium swallow study.

The FOIS is a 7-point scale used to describe a patient's functional oral intake level with scores ranging from "1" (Nothing by mouth) to "7" (Total oral diet with no restrictions) [22].

Patients

We utilized patient enrollment from an ongoing NIH-funded study. At the time of our study enrollment, we retrospectively enrolled all patients currently enrolled in the NIH-funded study who had records for a modified barium swallow study, had a complete EAT-10, with every item having a response documented, and were at least 18 years old. We did not apply any other inclusion or exclusion criteria and thus, enrolled patients with various etiologies and levels of dysphagia.

Procedures

As part of standard of care procedures, all patients in this study received a modified barium swallow study and were assigned a diet (FOIS). Patients scored the EAT-10 as either an inpatient or outpatient at the Medical University of South Carolina in Charleston, South Carolina. Modified barium swallow studies are considered the gold-standard for assessing swallow physiology and are part of the standard of care for patients with swallowing disorders. Modified barium swallow studies in our study were conducted following a standardized protocol presenting barium of different consistencies (thin liquid, nectar thick liquid, honey thick liquid, pudding, solid) and volumes (depending on consistency: 5ml, cup sip, sequential cup sips, ½ cookie) [33]. Patients were asked to complete the EAT-10 instrument at the time of the modified barium swallow study. When patients were not able to complete the EAT-10 independently, a speech-language pathologist assisted the patient, or a caregiver provided responses. The speech-language pathologist conducting the modified barium swallow study, also assigned a FOIS score reflecting their diet recommendation following the completed modified barium swallow study [22]. We excluded patients with missing PAS or FOIS data from our analyses of the relationship between these scales and the EAT-10. The Institutional Review Board of the Medical University of South Carolina approved this study.

Analyses

Dimensionality—Since the Rasch model assumes unidimensionality, we determined, prior to applying the Rasch model, whether the EAT-10 instrument is unidimensional, meaning all its instrument items are considered to be of the same measurement construct. We conducted confirmatory factor analysis with a one-factor solution, determined local independence and confirmed these findings using Rasch principal component analysis of residual and item and person fit statistics.

Confirmatory Factor Analysis: For confirmatory factor analysis, we used four model fit indices: normed chi-square (chi-square (χ^2)/degrees of freedom (df) test (<3.84 indicating good fit)) that accounts for sample size, comparative fit index (> 0.95), Tucker-Lewis Index (> 0.95), and root mean square error of approximation (< 0.08 adequate fit, < 0.06 excellent fit) [34, 35, 36].

Local Independence: We assessed the local independence of each item, to ensure that all EAT-10 items are independent of each other and only correlate with each other by the one latent trait that they are measuring. Local independence is an assumption of the Rasch model because it can affect the unidimensionality of the instrument [37]. Residual correlations derived from the confirmatory factor analysis greater than ± 0.2 were deemed as violating local independence [35]. We interpreted high positive residual correlations as indicative of local item dependency due to response dependency and high negative residual correlations as local item dependency caused by multidimensionality. We used Mplus version 7.31 (Los Angeles, CA 2012) to conduct the confirmatory factor analysis.

Rasch Principal Component Analysis of Residual: For the Rasch principal component analysis of residual, we used two criteria for determining unidimensionality: 1) >50% variance of the data is explained by the Rasch dimension and, 2) the first contrast in the standardized residuals has a size (Eigenvalue) of ≥ 2.0 [38, 39]. We used WINSTEPS version 3.81.0 for Rasch analyses [39].

Rating Scale Analysis—After confirmation of the unidimensionality, we assessed the rating scale of the EAT-10 with the Rasch model rating scale analysis. We used the following three essential criteria for rating scale assessment [40]: 1) at least ten observations per rating scale category, 2) the category measures advance (monotonicity, which is the average measure for the EAT-10 rating scale category 1 is greater than for the category 0, etc.), and 3) the outfit mean square value for each rating scale category is < 2.0 . We used WINSTEPS version 3.81.0 for Rasch analyses [39] with the rating scale model derived by Andrich [30]. We chose the rating scale model over the partial credit model, because all items of the EAT-10 are scored on the same rating scale. WINSTEPS software uses the joint maximum likelihood estimation method.

Item Fit Statistics and Person Fit Statistics—We assessed item and person fit statistics following previously reported guidelines [29]. Fit statistics differentiate between infit and outfit mean squares. Item infit statistics are more sensitive to the pattern of responses from individuals targeted at the item, and item outfit statistics are more sensitive to the pattern of responses from individuals far from the item. For example, for an item commonly reported as a problem, infit refers to the pattern of responses reported by individuals with mildly perceived swallowing disorders. For that same item, outfit would refer to the pattern of responses reported by individuals with seriously perceived swallowing disorders. Further, according to the Rasch model, items that are least commonly reported as a problem are rated as a problem only by individuals with seriously perceived swallowing disorders while items that are most commonly reported as a problem are rated as a problem by both individuals with seriously perceived swallowing disorders and individuals with mildly perceived swallowing disorders. Items or persons that do not fit this pattern are considered to misfit the Rasch model.

Person fit statistics describe how well the responses from a person fit to the predicted responses based on the Rasch model [20]. Responses can be too predictable (overfit) or too unpredictable, for example when a person who rates most items as no problem (a person with no or mild self-reported swallowing problems), rates an item as a problem that is

generally not rated as a problem by persons with no or mild swallowing problems. The Rasch model would have predicted that this person would have rated that item as no problem.

The Rasch model expects the infit and outfit mean square value to be 1.0. Higher mean square values suggest higher variation (unexplained) than expected by the Rasch model and lower mean square values suggest less variation (overly predictable ratings) than expected by the Rasch model. Wright and Linacre suggest for rating scales and surveys that reasonable item and person mean square values range from 0.6 to 1.4 for infit and outfit statistics with standardized z values of < 2.0 [41].

Item Hierarchy—Item hierarchy is the order of items based on their difficulty (likelihood to be rated as a problem). We estimated item difficulty with the Rasch model on a logit scale that provided the linear measures for each item. The lower the measure (higher the total score) of an item, the more patients scored this item as a “severe problem” and fewer patients scored it as “no problem”. In contrast to the linear measure, the total score is an ordinal count. Thus, the order/hierarchy of items is the same using the total score or measure. However, only the measure can be used to understand the distance between two items, which reflects the difference in the likelihood of items being rated as a problem.

To assess invariance and stability of the item hierarchy across groups, we performed DIF. The Rasch model assumes that patients with the same ability level, should rate items with the same probability independent of their group. DIF determines whether patients from different groups, who have the same perceived degree of swallowing problems, respond differently to any of the ten items across subgroup characteristics [42]. We hypothesized that patient gender (female vs. male), oral intake (normal vs. abnormal) and aspiration status (aspiration vs. no aspiration) will have no effect on the responses to the ten items of the EAT-10. We determined that an item displays DIF if the DIF contrast was $|DIF| \geq 0.43$ logits and the probability was ≤ 0.05 . Patients with a FOIS score of 1–6 were classified as patients with oral intake restrictions (abnormal oral intake), patients with a FOIS score of 7 as patients with normal oral intake, patients with a PAS score of 1–5 as patients who do not aspirate, and patients with a PAS score of 6–8 as patients who aspirate.

Item-Person Match—The Rasch model produces item calibrations (difficulties) and person calibrations (abilities) on the same interval scale using log equivalent units (logits). A logit scale value of 0 is arbitrarily set as the mean of the item difficulty estimate for the sample under study. Therefore, a logit scale value close to 0 is near the average item difficulty. For our EAT-10 analysis, patients with negative logit scale values have a higher ability, rating EAT-10 items, on average, as less of a problem as compared to the average item difficulty. Patients with positive logit scale values have a lower ability, rating EAT-10 items, on average, as more of a problem, as compared to the average item difficulty [16]. The Rasch analytic program produces an item-person map, displaying the difficulty (likelihood to be rated as a problem) of each EAT-10 item relative to the abilities (likelihood to rate EAT-10 items as a problem) of each person in the sample. This display indicates how well the likelihood of the EAT-10 items to be rated as a problem match the person likelihood to rate EAT-10 items as a problem. Ceiling and floor effects were considered when more

than 15% of the patients showed the maximum or minimum extreme score, respectively [43].

Precision and Reliability—We determined precision by person-separation reliability ratios and person-strata calculations. Acceptable person-separation ratios (G) in Rasch are determined at >2.0 which represents a Rasch reliability of 0.80 [44, 45]. Rasch reliability is similar to Cronbach's alpha and represents the ratio of the true variance to observed variance, where observed variance is the true variance adjusted for measurement error [46]. Error estimates depend on the precision of the instrument to measure the ability of targeted persons with the targeted items. Rasch reliability is an index for the repeatability of an instrument's measure. This is different than, Cronbach's alpha, here Kuder-Richardson 20, which represents the repeatability of raw scores. Because of this difference, it has been argued that Rasch reliability is more conservative, and therefore possibly less misleading, than Cronbach's alpha for reliability estimates of generalizable measures that are test-independent [47].

We also determined precision using person-strata calculations, which are based on the person-separation ratio, and define the instrument's precision to detect statistically distinguishable groups of patients. Person-strata is the number of center points, three standard errors apart that can be placed in a distribution; the more center points, the greater the precision. The person-strata calculations were computed using the discernable Strata formula, $(4G+1)/3$; where G=separation index [24, 31].

Relationship of the Eating Assessment Tool with the Penetration-Aspiration Scale and Functional Oral Intake Scale—To investigate the relationship between patients' ability level on the EAT-10 and the patients' degree of airway invasion and oral intake status, we conducted two correlation analyses: 1) the patients' corresponding linear measure of the EAT-10 with the patients' scores on the PAS, and 2) the patients' corresponding linear measure of the EAT-10 with the patients' scores on the FOIS. We applied Spearman correlations for ordinal variables with statistical significance considered to be $p < 0.05$. We interpreted the size of the correlation with $|r| = 0.1$ as weak, $|r| = 0.3$ as moderate, and $|r| = 0.5$ as strong [48].

Results

Patients

We included 127 patients in our study (female, $n = 79$). The average age was 59.9 years (SD 15.0, range 19 – 89). Almost 60% of patients had otolaryngological or neurological etiologies (Table 1).

The score distributions for the EAT-10, PAS and FOIS across all patients are shown in Table 2. The PAS was not reported for seven patients, while FOIS scores were not reported for two patients. The median EAT-10 total score was 16 out of 40 (range 0 to 40). The median PAS score was 2 (“Material enters the airway, remains above the vocal folds, and is ejected from the airway”) (range 1 to 8). The median FOIS score was 7 (“Total oral diet with no restrictions”) (range 1 to 7).

Dimensionality

Confirmatory Factor Analysis—Overall, confirmatory factor analysis supported a one-factor model (single measurement construct) for the EAT-10. Three of the four model fit indices for confirmatory factor analysis assessment were fulfilled: normed Chi-square was 2.26 (< 3.84), comparative fit index was 0.98 (> 0.95) and Tucker-Lewis Index was 0.97 (> 0.95). Only root mean square error of approximation marginally violated the criterion of < 0.08 with a value of 0.1. Factor loadings of the EAT-10 items supported a one-factor solution with loadings > 0.55 for all items.

Local Independence—Two items of the EAT-10 (“My swallowing problem interferes with my ability to go out for meals” and “Swallowing is painful”) demonstrated a high residual correlation, -0.269 (> 0.2), violating local independence and indicating multidimensionality.

Rasch Principal Component Analysis of Residual—The Rasch principal component analysis of residual of the EAT-10 fulfilled both criteria for unidimensionality: 53.9% of the data variance was explained by the Rasch dimension and the Eigenvalue of the first contrast (the secondary dimension) was 1.96.

Overall, the confirmatory factor analysis and Rasch principal component analysis of residual supported a single measurement construct of the EAT-10. While the root mean square error of approximation did not meet the criteria in the confirmatory factor analysis, and two items showed local dependence indicative of multidimensionality, all other indices met the fit criteria. Additionally, the Rasch confirmatory factor analysis demonstrated a single measurement construct. Since unidimensionality is an assumption of the Rasch model and our analyses supported that the EAT-10 is unidimensional, we investigated the EAT-10 with the Rasch model.

Rating Scale Analysis

The 5-point EAT-10 rating scale fulfilled all three essential criteria for use of the Rasch model. All five rating scale categories showed more than 10 observations, ranging from 146 observations for EAT-10 rating category “3” to 459 observations for EAT-10 rating category “0”. Rating score measures (category measures) advanced from the lowest to highest rating score. The outfit mean square was < 2.0 for all rating categories.

Item Fit Statistics and Person Fit Statistics

By assessing Rasch Item Fit, we found that two items of the EAT-10 did not fit the Rasch measurement model. The items “I cough when I eat” and “My swallowing problem has caused me to lose weight” had mean square values for infit and outfit of more than 1.4 and a z-score of more than 2.0 (Table 3).

Assessing Rasch Person Fit, we found that 8 of 127 patients (6%) did not fit the Rasch model and showed mean square values for infit and outfit of more than 1.4 and a z-score of more than 2.0.

Item Hierarchy

Table 4 shows the item hierarchy of the EAT-10, the order of items based on their difficulty (likelihood to be rated as a problem). Also shown in Table 4 is the total score of a single item that is the sum of scores from all patients (N=127) for this item. The EAT-10 item with the lowest measure (highest total score) was “Swallowing solids takes extra effort”, representing the most difficult item in Rasch terminology; meaning the majority of patients scored this item as a severe problem. The two easiest items were “Swallowing is painful” and “My swallowing problem has caused me to lose weight”; meaning that the majority of patients scored these items as no problem.

Several items had virtually identical average item difficulty measures. For example, the items “I cough when I eat” and “The pleasure of eating is affected by my swallowing” had identical Rasch item measures (-0.24) and the items “My swallowing problem interferes with my ability to go out for meals” and “Swallowing liquids takes extra effort” had very similar measures (0.18 and 0.23, respectively) (Table 4 and Figure 1).

Results of the DIF analyses are displayed in Table 5. No item displayed DIF for female compared to male patients. Item 6 (“swallowing is painful”) showed DIF for patients who aspirated compared with patients who did not aspirate during the modified barium swallow (DIF=-0.87, p=0.0445). Item 2 (“my swallowing problem interferes with my ability to go out for meals”) showed DIF for patients on a normal oral diet compared with patients who had oral diet restrictions (DIF=0.60, p=0.0445). Other items (“swallowing solids takes extra effort”, “swallowing pills takes extra effort”, “the pleasure of eating is affected by my swallowing”) fulfilled one but not both criteria (|DIF contrast| 0.43 and probability 0.05) for DIF.

Item-Person Match

In accordance with the Rasch model, average item difficulty is anchored at 0 logits. The patient cohort of this study had an average person ability of -0.45 logits. Considering that higher scores on the EAT-10 indicate patients’ perception of more severe swallowing problems, the average patient ability was slightly above the average item difficulty of the EAT-10 but within one standard deviation. Overall, the items and scale scores targeted the severity range of patients’ perception of swallowing problems fairly well, except for patients with the lowest and highest measures (Figure 1). Patients were least likely to score the item “Swallowing is painful” with a high scale score. Figure 1 shows the location of the item “Swallowing is painful” labeled as .35 in relation to patients’ ability of the change in scoring this item as “4” instead of “3”. Conversely, patients were most likely to score the item “Swallowing solids takes extra effort” as “1” instead of “0” (labeled as .05 in Figure 1).

Two of 127 (1.6%) patients showed a maximum extreme score by scoring all 10 items as “severe problem” (total test score of 40 for the EAT-10). Seven of 127 (5.5%) patients showed a minimum extreme score by scoring all 10 items as “no problem” (total test score of 0 for the EAT-10). Thus, floor and ceiling effects were not observed, as they were <15%.

Precision and Reliability

The Rasch person separation value of the EAT-10 was 1.94 and marginally failed the criterion of being >2.0 . The person-strata value was 2.92 and indicated that the instrument was able to divide our patient cohort into approximately three significantly different person-strata based on how much of a problem patients perceived their swallowing to be (such as no problem, moderate problem, and severe problem). The Rasch person reliability of the EAT-10 was 0.79. Cronbach's alpha for the EAT-10 was 0.90.

Relationship of the Eating Assessment Tool with the Penetration-Aspiration Scale and Functional Oral Intake Scale

Eating Assessment Tool and Penetration-Aspiration Scale—The correlations between the EAT-10 and PAS was statistically significant but weak in strength for the EAT-10 linear person measure ($r = 0.26, p = 0.0036$). The higher (worse) the EAT-10 linear person measure, the higher (worse) the PAS score.

Eating Assessment Tool and Functional Oral Intake Scale—The correlation between the EAT-10 and FOIS was statistically significant but weak in strength for the EAT-10 linear person measure ($r = -0.27, p = 0.0027$). The higher (worse) the EAT-10 linear person measure, the lower (worse) the FOIS score.

Discussion

The purpose of our study was to evaluate the construct validity of the EAT-10 in order to determine its quality and informative value. Hereby, we aimed to contribute to the psychometric evaluation of the EAT-10 that was previously criticized as limited [18]. Our study identified opportunities to improve the psychometric properties of the EAT-10.

Dimensionality

Our assessment of the EAT-10 supported a unidimensional construct with two exceptions. First, one of four criteria was not fulfilled in the confirmatory factor analysis to support unidimensionality of the EAT-10. We found a root mean square error of approximation > 0.08 . It has been discussed elsewhere [49] that the failure to achieve the root mean square error of approximation criterion can be related to a small sample size. Therefore, this confirmatory factor analysis criterion might be reached with a larger sample size. Second, we found that two EAT-10 items “My swallowing problem interferes with my ability to go out for meals” and “Swallowing is painful”, showed high negative residual correlations indicating local dependence. Violations of local independence suggest a second dimension in the instrument, because the interdependence of these items acts like another trait that is not reflected by the Rasch single dimension [37]. Despite the violation of local independence seen in our confirmatory factor analysis results, we decided to analyze the EAT-10 with the Rasch model as our results of the Rasch principal component analysis of residual were in favor of a unidimensional construct. Also, Cordier et al. (2017) support the unidimensionality of the EAT-10 [32].

Rating Scale Analysis

The Rasch rating scale analysis of the EAT-10 revealed that Linacre's three essential criterion were met [40]. Thus, patients used the full spectrum of the rating scale and used it in an expected manner with higher scores indicating more severe perceived problems.

Item Fit Statistics and Person Fit Statistics

Following the Rasch item fit analysis the EAT-10 had two misfitting items ("I cough when I eat" and "My swallowing problem has caused me to lose weight") for both infit and outfit statistics. Patients did not respond to these two items in a pattern that was predicted by the Rasch model as based on the patients' average perceived swallowing problems and based on the item's average likelihood to be rated as a problem. These two items were also indicative of misfitting the Rasch model in the study by Cordier et al. (2017) [32]. There can be various reasons for misfitting items. It is possible that some patients show but do not attribute symptoms of coughing and weight loss to their swallowing problem and, thus, respond to these items unexpectedly. Further, it is possible that the specific characteristics of our patient cohort contributed to the misfit of these items because our sample consisted mostly of patients with mild swallowing problems, as seen from low PAS and high FOIS scores. A patient cohort with a higher proportion of patients with severe swallowing problems would likely increase the range and variability of responses to the items. It is not known whether Cordier et al. (2017) included a higher proportion of patients with severe swallowing problems, as severity was not reported [32]. Therefore, re-evaluation of the EAT-10 utilizing a sample of patients with more severe dysphagia is warranted before any recommendations are made to remove the misfitting items ("I cough when I eat" and "My swallowing problem has caused me to lose weight") from the EAT-10.

Based on the Rasch person fit analysis, we found that only a few patients (6%) did not respond to the items of the EAT-10 in the pattern that was predicted by the Rasch model. There are different possible explanations for this observation. It is possible that some patients (or the person who filled out the EAT-10) did not understand instructions or items correctly and, thus, provided inadequate responses. Further, it is possible that some patients experience their dysphagia symptoms in a pattern that is very different to the majority of patients we included in our study. These variations and possible reasons should be addressed in future studies.

Item Hierarchy

Our clinical experience supports the identified item difficulty hierarchy of the EAT-10. The items "Swallowing solids takes extra effort" and "Swallowing pills takes extra effort" were most often rated as a problem. Swallowing solids or pills are known to be common clinical complaints for people with dysphagia [50, 51]. The items "My swallowing problem has caused me to lose weight" and "Swallowing is painful" were most often rated as no problem. While patients with dysphagia can experience weight loss or pain as associated complaints with dysphagia these are not as common as other reported problems.

The DIF analyses revealed four items that showed variance in the item hierarchy for the groups tested. As expected, female and male patients with the same degree of perceived

swallowing problems scored all ten items similarly. However, the items “My swallowing problem interferes with my ability to go out for meals” and “Swallowing is painful” showed DIF for patients who aspirated compared to patients who did not aspirate and for those with normal oral intake compared to those with oral intake restrictions. Further, the items “Swallowing pills take extra effort” and “The pleasure of eating is affected by my swallowing” displayed DIF for patients without aspiration compared to patients with aspiration. These four items were also found by Cordier et al. (2017) to display DIF against gender, confirmed oropharyngeal dysphagia, and language [32]. These findings suggest that, despite equally perceived swallowing problems among patients, the item hierarchy for these four EAT-10 items varies based on the group to which patients belong. Specifically, in our current study, we suspect that DIF was shown for patients with different aspiration status and different oral intake levels because some patients score items as no problem due to decreased occurrence. For example, patients who are known to aspirate and/or who have been recommended to restrict their oral intake might not experience going out for meals or swallowing pills. Thus, even though these patients have more severe swallowing problems, they might score these EAT-10 items as less of a problem because the treatment of their dysphagia alters their participation in eating and drinking tasks.

Item-Person Match

The average instrument difficulty of the EAT-10 matched the average perceived swallowing problems of our patient cohort fairly well. Also, the EAT-10 showed no floor effect for patients reporting severe swallowing problems and showed no notable ceiling effect for patients reporting no swallowing problems. Thus, our results suggest that the difficulty level of the EAT-10 is appropriate for patients with a wide range of severity of perceived swallowing problems. Because our patient cohort was skewed towards patients with rather mild swallowing problems (based on PAS and FOIS scores), a re-evaluation with a patient cohort with more severe dysphagia should be applied to validate this observation.

Precision and Reliability

The analysis of the precision showed that the EAT-10 divided the patient cohort into three significantly different person-strata (such as no problem, moderate problem, and severe problem). In terms of reliability, the EAT-10 had a Rasch person reliability value of 0.79, closely approximating a value >0.8 , classifying it as acceptable with a strong value [20]. Thus, we can assume that the reproducibility based on the measurement error and, consequently, the precision level of the results of the EAT-10 are within satisfactory limits for our sample. It is important to note that strata calculations are influenced by the range of abilities of the sample under study. With the inclusion of more individuals with more severe deficits, we would expect the strata to increase. Since one intended use of the EAT-10 is to document initial self-perceived dysphagia severity, a three person-strata separation seems to be sufficiently precise for an initial classification to inform the need for further evaluation. When documenting differences in patients' perceived consequences of dysphagia, clinicians can be confident in the precision of EAT-10 results, because differences in the total score (e.g. between two patients, or between two assessment sessions for the same patient) are most likely true differences rather than differences due to measurement error. Interestingly, Cordier et al. (2017) found poor person reliability (0.55) and a low person separation index

of less than 2 [32]. The difference in our findings and the findings by Cordier et al. (2017) might reflect differences in the patient cohorts or differences in how patients scored the translated versions of the EAT-10.

However, the EAT-10 is also intended to track changes before and after treatment. A person separation into three strata (or even less based on the study by Cordier et al. (2017) [32]) may not provide the needed precision to detect slight or moderate changes and, therefore, might provide false negative results (no significant change after treatment when a change truly occurred). The precision of the EAT-10 may be improved by adding more items that measure patients with a broader range of perceived swallowing problems. Further, in order to determine whether the person separation of the EAT-10 is adequate for measuring change, a (treatment/intervention) study with a longitudinal design is warranted, as opposed to a cross-sectional design used in the present study.

Relationship of the Eating Assessment Tool with the Penetration-Aspiration Scale and Functional Oral Intake Scale

Last, we investigated the relationship between the EAT-10 with the PAS and FOIS. Our analyses revealed significant, but weak to moderate, correlations of the EAT-10 and PAS and the EAT-10 and FOIS. Patients with higher EAT-10 scores were more likely to penetrate or aspirate (higher PAS scores) and patients with higher EAT-10 scores tended to have more restrictions in their oral intake (higher FOIS scores). Our findings for the PAS align with previous findings of weak correlations between the EAT-10 and the PAS [12]. Our results of the correlation analyses indicate that the extent of patients' self-perception of their swallowing problem (based on the EAT-10 rating) has only a minor association with their underlying swallowing disorder in terms of penetration and aspiration events and oral intake status documented after a modified barium swallow study. Our findings support previous findings that indices of patients' quality of life and health status, and diet recommendations provide distinct, but complimentary information for dysphagia management [33, 52]. Thus, clinicians should be aware that the EAT-10 is solely an instrument to assess patients' self-perception of dysphagia symptoms and cannot replace assessments of functional swallowing limitations. The weak to moderate correlations with the PAS and FOIS demonstrate that scores on the EAT-10 do not necessarily reflect what the patient can eat and drink. Nevertheless, the EAT-10 offers complimentary diagnostic value.

Additionally, the EAT-10 is likely to be useful in another, seldom discussed way. We have found in our clinical experience that instruments like the EAT-10 help clinicians and patients reach a common understanding of the patient's swallowing complaints. This allowing for opportunities for targeted treatment options that address individualized goals. Further investigation of the EAT-10 to classify self-perceived symptom severity, track longitudinal changes as well as to identify and have shared development of individualized treatment goals is warranted.

Study Limitations

The EAT-10 was developed to assess patients with dysphagia stemming from a broad range of underlying diseases. Therefore, we recruited patients with heterogeneous diagnostic

categories. Nevertheless, different diagnostic categories might represent different latent states of our sample and, thus, might have influenced the results. Further, even though our patient cohort consisted of patients representing the whole severity spectrum of swallowing disorders as measured with airway invasion and oral intake status, overall, our patient sample had rather mild swallowing problems (median PAS score = 2; median FOIS score = 7). This may be because we have mainly recruited patients who were able to complete the EAT-10 on their own and, thus, were generally more medically stable. Repeating our study with a cohort of patients with more severe dysphagia may support the generalizability of our results to patients with varying levels of swallowing ability.

Conclusion

Our analysis indicates that the EAT-10 is unidimensional, the rating scale is appropriately used, the EAT-10 matched our patient cohort fairly well and the EAT-10 showed significant but low correlation with oral intake status and airway invasion. Nevertheless, our analysis identified significant weaknesses of the psychometric properties of the EAT-10. Patients did not respond in a predicted pattern to two of ten items and, thus, these two items misfitted the Rasch model, and four other items showed DIF. Further, the EAT-10 with its current collection of items may lack precision to document changes in patients' self-perceived swallowing problems. Due to the deficits seen in the construct validity of the EAT-10 based on the results of our study and the findings of Cordier et al. (2017) [32], we suggest the need to improve the EAT-10 to support its frequent use in clinical practice and research, we recommend further analyses with larger and more severe patient cohorts to decide whether items should be exchanged or deleted from the EAT-10 and if precision can be improved.

Acknowledgments

This work was supported by the National Institute of Deafness and Other Communication Disorders at the National Institute of Health (Grants 1K24DC12801 and T32DC0014435) and an unrestricted educational grant from Bracco Diagnostics, Inc.

References

1. Kertscher B, Speyer R, Fong E, Georgiou AM, Smith M. Prevalence of oropharyngeal dysphagia in the Netherlands: a telephone survey. *Dysphagia*. 2015; 30:114–20. [PubMed: 25432669]
2. Lin LC, Wu SC, Chen HS, Wang TG, Chen MY. Prevalence of impaired swallowing in institutionalized older people in taiwan. *J Am Geriatr Soc*. 2002; 50:1118–23. [PubMed: 12110075]
3. Sharma JC, Fletcher S, Vassallo M, Ross I. What influences outcome of stroke–pyrexia or dysphagia? *Int J Clin Pract*. 2001; 55:17–20. [PubMed: 11219313]
4. Martino R, Foley N, Bhogal S, Diamant N, Speechley M, Teasell R. Dysphagia after stroke: incidence, diagnosis, and pulmonary complications. *Stroke*. 2005; 36:2756–63. [PubMed: 16269630]
5. Marik PE, Kaplan D. Aspiration pneumonia and dysphagia in the elderly. *Chest*. 2003; 124:328–36. [PubMed: 12853541]
6. Foley NC, Martin RE, Salter KL, Teasell RW. A review of the relationship between dysphagia and malnutrition following stroke. *J Rehabil Med*. 2009; 41:707–13. [PubMed: 19774302]
7. Crary MA, Humphrey JL, Carnaby-Mann G, Sambandam R, Miller L, Silliman S. Dysphagia, nutrition, and hydration in ischemic stroke patients at admission and discharge from acute care. *Dysphagia*. 2013; 28:69–76. [PubMed: 22684924]

8. Bonilha HS, Simpson AN, Ellis C, Mauldin P, Martin-Harris B, Simpson K. The one-year attributable cost of post-stroke dysphagia. *Dysphagia*. 2014; 29:545–52. [PubMed: 24948438]
9. Ekberg O, Hamdy S, Woisard V, Wuttge-Hannig A, Ortega P. Social and psychological burden of dysphagia: its impact on diagnosis and treatment. *Dysphagia*. 2002; 17:139–46. [PubMed: 11956839]
10. Belafsky PC, Mouadeb DA, Rees CJ, Pryor JC, Postma GN, Allen J, Leonard RJ. Validity and reliability of the Eating Assessment Tool (EAT-10). *Annals of Otolaryngology, Rhinology & Laryngology*. 2008; 117:919–24.
11. Goncalves MI, Remaili CB, Behlau M. Cross-cultural adaptation of the Brazilian version of the Eating Assessment Tool – EAT-10. *Codas*. 2013; 25:601–4. [PubMed: 24626972]
12. Schindler A, Mozzanica F, Monzani A, Ceriani E, Atac M, Jukic-Peladic N, Venturini C, Orlandoni P. Reliability and validity of the Italian Eating Assessment Tool. *Ann Otol Rhinol Laryngol*. 2013; 122:717–24. [PubMed: 24358633]
13. Farahat M, Mesallam TA. Validation and Cultural Adaptation of the Arabic Version of the Eating Assessment Tool (EAT-10). *Folia Phoniatri Logop*. 2016; 67:231–7.
14. Moller R, Safa S, Ostberg P. Validation of the Swedish translation of eating assessment tool (S-EAT-10). *Acta Otolaryngol*. 2016:1–5.
15. Cheney DM, Siddiqui MT, Litts JK, Kuhn MA, Belafsky PC. The Ability of the 10-Item Eating Assessment Tool (EAT-10) to Predict Aspiration Risk in Persons With Dysphagia. *Ann Otol Rhinol Laryngol*. 2014
16. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996; 11:93–8. [PubMed: 8721066]
17. Hambleton RK, Jones RW. An NCME Instructional Module on “Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development”. *Educational Measurement: Issues and Practice*. 1993; 12:38–47.
18. Speyer R, Cordier R, Kertscher B, Heijnen BJ. Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: a systematic literature review. *Biomed Res Int*. 2014; 2014:458678. [PubMed: 24877095]
19. Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Annu Rev Clin Psychol*. 2009; 5:1–25. [PubMed: 19086835]
20. Bond, TG, Fox, CM. *Applying the Rasch model: fundamental measurement in the human sciences*. 2. Mahwah, N.J: Lawrence Erlbaum Associates Publishers; 2007.
21. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med*. 2003; 35:105–15. [PubMed: 12809192]
22. Crary MA, Mann GD, Groher ME. Initial psychometric assessment of a functional oral intake scale for dysphagia in stroke patients. *Arch Phys Med Rehabil*. 2005; 86:1516–20. [PubMed: 16084801]
23. Velozo CA, Seel RT, Magasi S, Heinemann AW, Romero S. Improving measurement methods in rehabilitation: core concepts and recommendations for scale development. *Arch Phys Med Rehabil*. 2012; 93:S154–63. [PubMed: 22840881]
24. Sabari JS, Woodbury M, Velozo CA. Rasch analysis of a new hierarchical scoring system for evaluating hand function on the motor assessment scale for stroke. *Stroke Res Treat*. 2014; 2014:730298. [PubMed: 25177513]
25. Velozo CA, Woodbury ML. Translating measurement findings into rehabilitation practice: an example using Fugl-Meyer Assessment-Upper Extremity with patients following stroke. *J Rehabil Res Dev*. 2011; 48:1211–22. [PubMed: 22234665]
26. Baylor C, Hula W, Donovan NJ, Doyle PJ, Kendall D, Yorkston K. An introduction to item response theory and Rasch models for speech-language pathologists. *American journal of speech-language pathology*. 2011; 20:243–59. [PubMed: 21622595]
27. Program Committee of the Institute for Objective Measurement. Definition of Objective Measurement 2000. Dec 17, 2016. Available from: <http://www.rasch.org/define.htm>.
28. White LJ, Velozo CA. The use of Rasch measurement to improve the Oswestry classification scheme. *Archives of Physical Medicine and Rehabilitation*. 2002; 83:822–31. [PubMed: 12048662]

29. Jafari P, Bagheri Z, Ayatollahi SMT, Soltani Z. Using Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQL(TM)4.0 Generic Core Scales in school children. *Health and Quality of Life Outcomes*. 2012; 10:27. [PubMed: 22414135]
30. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978; 43:561–73.
31. Masters GN. A rasch model for partial credit scoring. *Psychometrika*. 1982; 47:149–74.
32. Cordier R, Joosten A, Clave P, Schindler A, Bulow M, Demir N, Arslan SS, Speyer R. Evaluating the Psychometric Properties of the Eating Assessment Tool (EAT-10) Using Rasch Analysis. *Dysphagia*. 2017; 32:250–60. [PubMed: 27873090]
33. Martin-Harris B, Brodsky MB, Michel Y, Castell DO, Schleicher M, Sandidge J, Maxwell R, Blair J. MBS measurement tool for swallow impairment–MBSImp: establishing a standard. *Dysphagia*. 2008; 23:392–405. [PubMed: 18855050]
34. Hu, Lt; Bentler, PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55.
35. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007; 45:S22–31. [PubMed: 17443115]
36. Brown, TA. *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications; 2006.
37. Baghaei P. Local Dependency and Rasch Measures. *Rasch Measurement Transactions*. 2008; 21:1105–6.
38. Linacre JM. Data Variance Explained by Measures. *Rasch Measurement Transactions*. 2006; 20:1045–54.
39. Author. *Winsteps Rasch measurement [Computer software]*. Chicago, IL: Winsteps.com; 2015.
40. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002; 3:85–106. [PubMed: 11997586]
41. Wright BD, Linacre JM. Reasonable Mean-square Fit Values, *Rasch Measurement Transactions*. *Rasch Measurement Transactions*. 1994; 8:370.
42. Tennant A, Pallant JF. DIF matters: a practical DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions*. 2007; 20:1082–4.
43. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995; 4:293–307. [PubMed: 7550178]
44. Fisher W. Reliability, Separation, Strata Statistics. *Rasch Measurement Transactions*. 1992; 6:238.
45. Wright, BD, Masters, G. *Rating scale analysis*. Chicago: Mesa Press; 1982.
46. Linacre JM. Reliability, Separation and Strata: Percentage of Sample in Each Level. *Rasch Measurement Transactions*. 2013; 24:1399.
47. Linacre JM. KR-20/Cronbach Alpha or Rasch Person Reliability: Which Tells the “Truth”? *Rasch Measurement Transactions*. 1997; 11:580–1.
48. Cohen, J. *Statistical power analysis for the behavioral sciences 2nd ed ed*. Hillsdale, NJ: Lawrence Erlbaum; 1988.
49. Kenny DA, Kaniskan B, McCoach DB. The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*. 2014; 44:486–507.
50. Kocdor P, Siegel ER, Giese R, Tulunay-Ugur OE. Characteristics of dysphagia in older patients evaluated at a tertiary center. *Laryngoscope*. 2015; 125:400–5. [PubMed: 25196400]
51. Carnaby-Mann G, Crary M. Pill swallowing by adults with dysphagia. *Arch Otolaryngol Head Neck Surg*. 2005; 131:970–5. [PubMed: 16301368]
52. McHorney CA, Martin-Harris B, Robbins J, Rosenbek J. Clinical validity of the SWAL-QOL and SWAL-CARE outcome tools with respect to bolus flow measures. *Dysphagia*. 2006; 21:141–8. [PubMed: 16715210]

Implications for Rehabilitation

- Swallowing disorders are associated with severe complications, such as pneumonia and malnutrition, and impose both social and psychological burdens on patients.
- The Eating Assessment Tool is a self-report instrument developed to estimate initial dysphagia severity and monitor change in patient-reported dysphagia symptoms as a response to treatment.
- This study shows that the Eating Assessment Tool has deficits in its construct validity and a need to improve the instrument to support its frequent use in clinical practice and research.

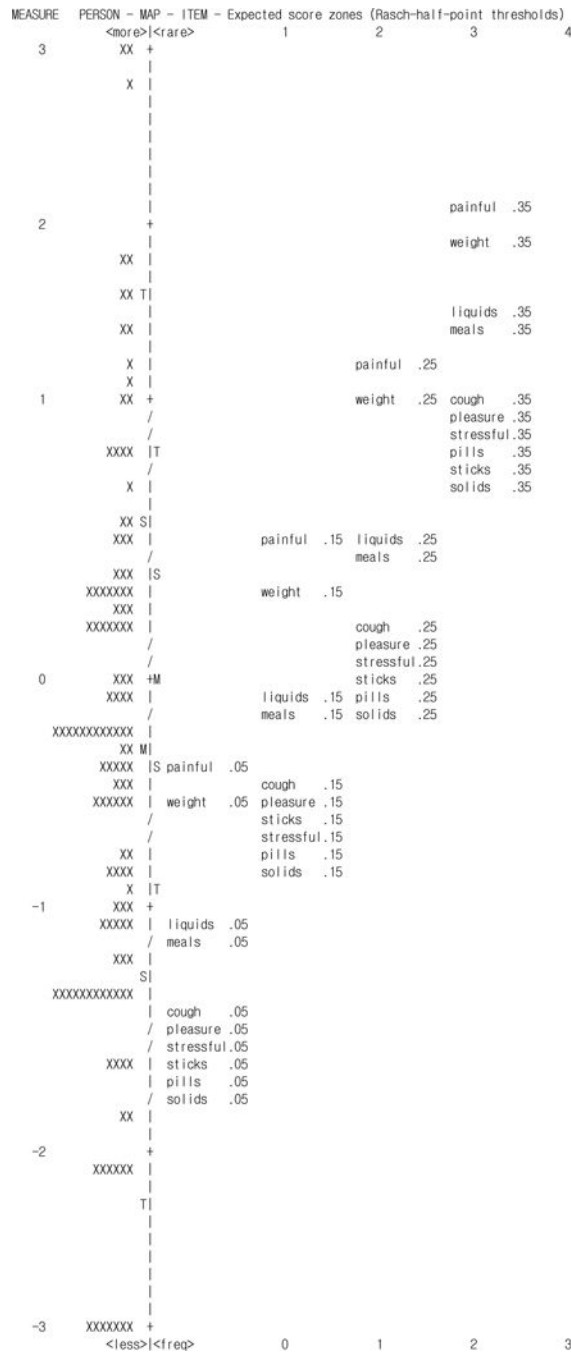


Figure 1. Rasch Person-item map with Rasch-half-point thresholds for the Eating Assessment Tool

Rasch person-item map with Rasch-half-point thresholds between adjacent response categories for the Eating Assessment Tool. Weight = “My swallowing problem has caused me to lose weight”; meals = “My swallowing problem interferes with my ability to go out for meals”; liquids = “Swallowing liquids takes extra effort”; solids = “Swallowing solids takes extra effort”; pills = “Swallowing pills takes extra effort”; painful = “Swallowing is painful”; pleasure = “The pleasure of eating is affected by my swallowing”; sticks = “When

I swallow food sticks in my throat”; cough = “I cough when I eat”; stressful = “Swallowing is stressful”;/= break in the vertical scale due to carriage returns

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Diagnostic Categories

Diagnostic Category	N (%)
Otolaryngology (e.g. Thyroidectomy)	55 (43.31)
Neurological (e.g. Stroke, ALS, Parkinson)	23 (18.11)
Pulmonology (e.g. COPD, Lung Cancer)	14 (11.02)
Head and Neck Cancer	10 (7.87)
General Medicine	9 (7.09)
Gastroenterology	4 (3.15)
Cardiothoracic Surgery	4 (3.15)
Others (e.g. Cervical Spine Surgery, idiopathic)	8 (6.30)
Total	127 (100)

N= number of patients; ALS= amyotrophic lateral sclerosis; COPD= chronic obstructive pulmonary disease

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Patient characteristics and swallowing assessments

	Number of patients	Mean (SD)	Median	Range
Age	127	59.93 (14.95)	62	19–89
Female	79 (62.2%)	–	–	–
Eating Assessment Tool (min 0, max 40)	127	16.02 (10.79)	16	0–40
Penetration-Aspiration Scale (worst score across all swallows) (min 1, max 8)	120	2.83 (2.41)	2	1–8
Functional Oral Intake Scale (min 1, max 7)	125	6.01 (1.64)	7	1–7

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3


Item fit statistics for the Eating Assessment Tool

Eating Assessment Tool Item	Measure	Error	Infit MnSq	Infit Zstd	Outfit MnSq	Outfit Zstd
1. My swallowing problem has caused me to lose weight.	0.67	0.10	1.45*	2.80*	1.61*	2.70*
2. My swallowing problem interferes with my ability to go out for meals.	0.18	0.09	1.14	1.10	1.00	0.10
3. Swallowing liquids takes extra effort.	0.23	0.09	1.07	0.60	0.99	0.00
4. Swallowing solids takes extra effort.	-0.45	0.09	0.61	-3.6	0.64	-2.70
5. Swallowing pills takes extra effort.	-0.38	0.09	0.99	0.00	0.97	-0.10
6. Swallowing is painful.	0.83	0.11	1.13	0.80	0.92	-0.30
7. The pleasure of eating is affected by my swallowing.	-0.24	0.09	0.71	-2.6	0.63	-2.70
8. When I swallow food sticks in my throat.	-0.33	0.09	1.01	0.10	1.30	1.90
9. I cough when I eat.	-0.24	0.09	1.60*	4.20*	1.80*	4.30*
10. Swallowing is stressful.	-0.26	0.09	0.58	-4.00	0.53	-3.70

* = misfit item to the measurement model; MnSq=mean square; Zstd=standardized z value

Table 4

Items in measure order (Item Difficulty Hierarchy)

Item Difficulty	Eating Assessment Tool Item	Total Score	Measure	Error
Least commonly reported as a problem 	6. Swallowing is painful.	111	0.83	0.11
	1. My swallowing problem has caused me to lose weight.	126	0.67	0.10
	3. Swallowing liquids takes extra effort.	173	0.23	0.09
	2. My swallowing problem interferes with my ability to go out for meals.	179	0.18	0.09
	9. I cough when I eat.	231	-0.24	0.09
	7. The pleasure of eating is affected by my swallowing.	232	-0.24	0.09
	10. Swallowing is stressful.	234	-0.26	0.09
	8. When I swallow food sticks in my throat.	243	-0.33	0.09
	5. Swallowing pills takes extra effort.	250	-0.38	0.09
	Most commonly reported as a problem	4. Swallowing solids takes extra effort.	258	-0.45

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Results of the differential item functioning analyses.

Eating Assessment Tool Item	Gender Females (n=79) vs Males (n=48)			Aspiration (n=21) vs Aspiration (n=106)			Oral intake Normal (n=76) vs Abnormal (n=49)		
	DIF contrast	Rasch-Welch probability	Rasch-Welch probability	DIF contrast	Rasch-Welch probability	Rasch-Welch probability	DIF contrast	Rasch-Welch probability	Rasch-Welch probability
1. My swallowing problem has caused me to lose weight.	0.00	1.000	1.000	0.00	1.000	1.000	0.28	0.1751	0.1751
2. My swallowing problem interferes with my ability to go out for meals.	0.25	0.1855	0.0449*	0.48*	0.0449*	0.0449*	0.60*	0.0018*	0.0018*
3. Swallowing liquids takes extra effort.	0.18	0.3496	0.1609	0.33	0.1609	0.1609	-0.11	0.5427	0.5427
4. Swallowing solids takes extra effort.	0.00	1.000	0.8954	0.03	0.8954	0.8954	0.31	0.0972	0.0972
5. Swallowing pills takes extra effort.	-0.21	0.2412	0.0280*	-0.52*	0.0280*	0.0280*	-0.03	0.8842	0.8842
6. Swallowing is painful.	-0.14	0.5417	0.0047*	-0.87*	0.0047*	0.0047*	-0.48*	0.0303*	0.0303*
7. The pleasure of eating is affected by my swallowing.	0.15	0.4007	0.0433*	0.53*	0.0433*	0.0433*	0.35	0.0606	0.0606
8. When I swallow food sticks in my throat.	0.19	0.2884	0.7870	-0.06	0.7870	0.7870	-0.30	0.0945	0.0945
9. I cough when I eat.	-0.07	0.6867	0.8691	0.04	0.8691	0.8691	-0.40	0.0263*	0.0263*
10. Swallowing is stressful.	-0.32	0.0810	0.5210	-0.15	0.5210	0.5210	-0.23	0.2006	0.2006

DIF=differential item functioning;

* = |DIF contrast| 0.43 or $p < 0.05$.