



Universiteit
Leiden
The Netherlands

Using classical test theory to determine the psychometric properties of the deglutition handicap index

Speyer, R.; Cordier, R.; Bouix, C.; Gallois, Y.; Woisard, V.

Citation

Speyer, R., Cordier, R., Bouix, C., Gallois, Y., & Woisard, V. (2021). Using classical test theory to determine the psychometric properties of the deglutition handicap index. *Dysphagia*, 37(1), 65-73. doi:10.1007/s00455-021-10250-2

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3589624>

Note: To cite this publication please use the final published version (if applicable).



Using Classical Test Theory to Determine the Psychometric Properties of the Deglutition Handicap Index

Renée Speyer^{1,2,3} · Reinie Cordier^{2,4} · Clara Bouix⁵ · Yohan Gallois⁶ · Virginie Woisard^{5,7,8}

Received: 23 June 2020 / Accepted: 13 January 2021 / Published online: 30 January 2021
© The Author(s) 2021

Abstract

The Deglutition Handicap Index (DHI) is a self-report measure for patients at risk of oropharyngeal dysphagia on deglutition-related aspects of functional health status (FHS) and health-related quality of life (HR-QoL). The DHI consists of 30 items which are subsumed within the Symptom, Functional and Emotional subscales. The purpose of this study was to evaluate the psychometric properties of the DHI using Classic Test Theory according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria. A total of 453 patients with dysphagia with different aetiologies were recruited concurrently at two academic hospitals. Dysphagia was confirmed by fiberoptic endoscopic and/or videofluoroscopic evaluation of swallowing. In addition, a healthy control group of 132 participants were recruited. Structural validity was determined using exploratory and confirmatory factor analyses and internal consistency by calculating Cronbach's alpha coefficients. Hypothesis testing was evaluated using Mann–Whitney *U*-tests, linear regression analysis and correlations analysis. Diagnostic performance and receiver operating characteristic curves analysis were calculated. Factor analyses indicated that the DHI is a unidimensional measure. The DHI has good internal consistency with some indication of item redundancy, weak to moderate structural validity and strong hypothesis testing for construct validity. The DHI shows high diagnostic performance as part of criterion validity. These findings support that the DHI is an appropriate choice as a patient self-report measure to evaluate FHS and HR-QoL in dysphagia. Ongoing validation to assess the measure for possible item redundancy and to examine the dimensionality of the DHI using item response theory is recommended.

Keywords DHI · Psychometrics · Reliability · Validity · Diagnostic performance · Classical test theory

Introduction

Dysphagia is commonly defined as difficulties or abnormalities in swallowing [1]. Dysphagia is considered to be a multifaceted condition that requires a combination

of outcome measures to fully comprehend and assess, for example, anatomy and cranial nerve integrity, oral motor skills, nutritional and oral intake status, functional health status (FHS) and health-related quality of life (HR-QoL) [1]. Patient self-reported measures are commonly used in

✉ Renée Speyer
renee.speyer@isp.uio.no

¹ Department Special Needs Education, University of Oslo, Oslo, Norway

² School of Occupational Therapy, Social Work and Speech Pathology, Faculty of Health Sciences, Curtin University, Perth, Australia

³ Department of Otorhinolaryngology and Head and Neck Surgery, Leiden University Medical Centre, Leiden, The Netherlands

⁴ Department Social Work, Education and Community Wellbeing, Northumbria University, Newcastle upon Tyne, UK

⁵ Logopedics Training Center, Medical Department, Paul Sabatier University Toulouse III, Toulouse, France

⁶ Department of Otoneurology and Pediatric ENT, Pierre Paul Riquet Hospital, University Hospital of Toulouse, Toulouse, France

⁷ Voice and Deglutition Unit, Department of Otorhinolaryngology and Head and Neck Surgery, Larrey Hospital, University Hospital of Toulouse, Toulouse, France

⁸ Oncorehabilitation Unit, Toulouse University Cancer Institute, Oncopole Hospital, Toulouse, France

the assessment of FHS and HR-QoL as a component of the multidimensional assessment of dysphagia [2]. FHS refers to the ability to perform tasks across multiple domains and measures function in relation to disease and/or treatment and the effects thereof on activities of daily life [3]. In turn, HR-QoL reflects the unique personal perception of one's health, taking into account social, functional and psychological factors [4].

The use of a measure in clinics or research can only be justified by its robust psychometric properties across all psychometric domains [1]. The COSMIN group (COnsensus-based Standards for the selection of health Measurement INstruments) established an international consensus-based taxonomy, terminology and definitions of measurement properties for health-related patient-reported outcome measures [5–8]. The framework comprises nine measurement properties subsumed within three domains: reliability, validity and responsiveness.

To select appropriate measures from available patient self-report questionnaires, the psychometric properties of each questionnaire must be determined and compared. Recent systematic reviews have evaluated the psychometric properties of currently available patient self-report questionnaires on FHS and HR-QoL in dysphagia [3, 9]. Both reviews identified poor and incomplete psychometric data for most of the included questionnaires and prompted the need for ongoing validation of self-reported measure of FHS and HR-QoL.

The Deglutition Handicap Index (DHI) is a patient self-report measure in dysphagia (comprising constructs related to both FHS and HR-QoL) and was developed by Woisard and Andrieux [10]. The DHI is a 30-item questionnaire of deglutition-related aspects of daily life using a 5-point scale per item (0–4: never, almost never, sometimes, almost always, always). Higher scores indicate higher degrees of swallowing problems affecting daily functioning. The DHI is subdivided into the following three domains each consisting of ten items: emotional domain (psychosocial), functional domain (nutritional and respiratory), and physical domain (symptoms related to swallowing). The total DHI score ranges from zero (indicating no impact) to 120 (indicating maximum impact).

The DHI was developed based on literature and the content expertise of two phoniaticians and two speech and language pathologists. A preliminary version of the DHI was trialled in a diverse patient population with dysphagia, including patients with head and neck cancer and neurological disorders. Thirty patients were asked about the comprehensibility of items. Patients were also asked whether any important concepts were missing (comprehensiveness) and whether all items were relevant. Next, a revised version of the DHI based on patients' comments was re-trialled in a new group of ten patients. No further revisions were needed.

After meeting the requirements for content validity [7], the DHI was first published in 2006 [10].

Some other psychometric properties of the DHI have been evaluated in the literature using Classical Test Theory (CTT): internal consistency [10, 11], reliability over time or test–retest reliability [12], structural validity [11], and responsiveness or sensitivity to change [13]. However, due to relatively small sample sizes, only preliminary conclusions could be drawn. Also, no information was available about possible DHI cut-off scores to distinguish between patient populations and healthy participants.

The purpose of this study is to evaluate the psychometric properties of the DHI using CTT according to the COSMIN framework, thus meeting current standards in psychometrics. We aim to determine internal consistency, structural validity, hypothesis testing for construct validity and interpretability. We will also calculate diagnostic performance as part of criterion validity.

Methods

Participants

To determine the psychometric properties of the DHI, two different populations were recruited: a patient group diagnosed with dysphagia and a control group of healthy participants. Patients were recruited from outpatient clinics at two academic hospitals in France between February 2015 and December 2019: the Oncology-rehabilitation unit, Toulouse University Cancer Institute, Oncopole Hospital, and the Voice and deglutition unit, department of Otorhinolaryngology and Head and Neck Surgery, University Hospital of Toulouse, Larrey Hospital. Patients were included if they were diagnosed as having dysphagia following instrumental assessment (fiberoptic endoscopic evaluation of swallowing [FEES] and/or videofluoroscopic recording of swallowing [VFS]) by one of three phoniaticians. Patients were diagnosed with dysphagia if swallowing problems were detected from either FEES or VFS recordings. People with severe cognitive impairments were excluded. A control group was recruited from relatives and caretakers of included patients at the Larrey Hospital location between March 2016 and January 2020. Control participants were excluded if they showed signs of swallowing problems, chronic cough or non-specific respiratory diseases.

A total of 453 patients were recruited: 56.4% men and 43.6% women. Patients ranged in age from 19 to 100 years of age with a mean age of 59.9 years ($SD \pm 17.5$). Patients with various diagnoses were included: neurological diseases ($n = 196$; 43.3%), head and neck cancer ($n = 136$; 30.0%), oesophageal disorders ($n = 61$; 13.5%) and other diseases or disorders including aerodigestive tract disorders, cervical

spine injuries and structural abnormalities of the larynx ($n = 60$; 13.2%). The control group consisted of 132 participants (32.8% men; 67.2% women) and ranged in age from 15 to 96 years with a mean age of 54.4 years ($SD \pm 19.8$).

Protocol

All patients received either a FEES or a VFS or both instrumental assessments using a standardised protocol that involved repeated swallow trials of different viscosities and volumes. Next, patients completed the DHI. Based on evaluations of the FEES and VFS recordings, the phoniatician scored severity of swallowing difficulty using a 3-point rating scale: mild (some residue and/or penetration), moderate (significant residue and penetration/aspiration) and severe (significant residue and aspiration with complications resulting in for example, pulmonary or nutritional problems). The scores represent an overall clinical expert judgement of dysphagia severity based on instrumental assessment (FEES and/or VFS).

Statistical Analysis

Measurement properties of the DHI were determined according to the COSMIN taxonomy of psychometric properties and definitions for health-related outcomes [6, 7]. Patient data were used to evaluate internal consistency, structural validity and hypothesis testing for construct validity. Interpretability is not considered a psychometric property, but is regarded as an important characteristic of a measure to assign qualitative meaning to quantitative data [6]. Data from both clinical and control groups were used to calculate diagnostic performance.

Structural validity Both an exploratory Principal Component factor analysis and a confirmatory Maximum Likelihood (ML) factor analysis were performed to determine structural validity (i.e. the degree to which scores reflect the dimensionality of the construct to be measured).

Internal consistency Internal consistency provides information on the degree of interrelatedness among items of a questionnaire. The internal consistency reliability was examined by calculating Cronbach's alpha coefficients for the whole questionnaire, as well as for each subscale separately. A low Cronbach's alpha ($\alpha < 0.70$) indicates inadequate internal consistency, whereas a high Cronbach's alpha ($\alpha > 0.90$) suggests redundancy of items [14].

Hypothesis testing for construct validity Hypothesis testing for construct validity is defined as the degree to which scores of a measure are consistent with hypotheses, for instance, with regard to internal relationships, relationships to scores of other measures, or differences between relevant groups, based on the assumption that the measure truly measures the construct under investigation [6]. The

following hypotheses were tested: (1) The DHI Total score will be positively associated with the 3-point severity rating by the phoniaticians (Pearson or Spearman correlations); (2) patients with dysphagia will have significantly higher DHI Total score compared to healthy control participants (Mann–Whitney U test); (3) no significant differences on DHI Total scores are expected between genders; and (4) age is not positively associated with DHI Total scores.

Criterion validity (diagnostic performance) Criterion validity refers to the degree to which the scores of a measure are an adequate reflection of a 'gold standard' [6]. In dysphagia assessment, instrumental assessments (VFS and FEES) are usually considered the best standard available even though no international consensus exists on what measure to use to interpret recordings of swallowing [15]. As such, the comparison of the DHI scores to VFS and FEES can be considered criterion validity.

Data from healthy participants were used to determine diagnostic cut-off points between healthy participants and patients with dysphagia. Diagnostic performance was determined using results from both FEES and VFS as reference tests to calculate the following diagnostic parameters: sensitivity, specificity, positive and negative predictive value, and positive and negative likelihood ratio. Receiver Operating Characteristic (ROC) curves analysis using data from both clinical and control groups further investigated the diagnostic test performance of the DHI. Greater areas under the curve indicate at better diagnostic performance [16].

Interpretability The degree to which one can assign qualitative meaning to quantitative scores or change in quantitative scores is referred to as interpretability [6]. Interpretability was evaluated by comparing means and standard deviations of scores of patients with healthy controls. The DHI was also investigated for floor and ceiling effects which are considered to be present if more than 15% of all participants achieve the lowest or highest possible score, respectively [17]. The presence of floor and ceiling effects indicates that extreme items in the lower or upper end of the scale are missing, which can result in inadequate content validity and reliability [18]. As it relates to the DHI, both the comparison of distributions of clinical versus healthy controls as well as the assessment of floor and ceiling effects provide qualitative meaning about how to interpret raw DHI scores (quantitative scores).

Results

Structural Validity

An exploratory Principal Component factor analysis was performed for all DHI items. Five factors were revealed, explaining 55.9% of the total variance (goodness-of-fit

test, $p < 0.001$). All items loaded mainly on the first factor explaining 33.6% of all variance. Unidimensionality can be assumed for the following two reasons: (1) the first factor accounts for at least 20% of the variability, and (2) the ratio of the variance explained by the first to the second factor is greater than 4 ($\sigma^2_{\text{Factor 1}} / \sigma^2_{\text{Factor 2}} = 33.6/7.2 = 4.7$) [19].

When performing a confirmatory factor analysis using a three-factor model (representing the three underlying constructs or subscales), 41.6% of the total variance was explained (goodness-of-fit test, $p < 0.001$). The explained variance to account for the model is below the minimum expected values of between 50 [20] and 60% [21]. The first factor explained 31.6%, the second factor 4.7% and the third factor 5.2% of the variance. Again, all items loaded mainly on the first factor (Table 1). Table 2 presents the heat map of the absolute values of the correlation matrix including all DHI items. The determinant value of 6.225E-7 indicates low correlations between items (conventional cut-off > 0.00001 [22]). No multicollinearity was present as all correlations ≤ 0.730 (conventional cut-off < 0.08 [22, 23]). The findings of the factor analyses strongly suggest that the DHI represents a unidimensional construct. Therefore, it was decided to only consider the DHI Total score for further analyses of interpretability and diagnostic performance.

Internal Consistency

The data were not normally distributed, as such, non-parametric correlations were calculated. Spearman correlations between items within the DHI varied greatly ($0.004 \leq r_s \leq 0.734$). The strength of r -values can be interpreted based on Cohen’s classification: 0.10 as weak, 0.30 as moderate, and 0.50 as strong in terms of magnitude [24]. The mean correlation for items within the DHI Total score was weak to moderate ($r_s = 0.294$; $SD \pm 0.13$). Cronbach’s alpha was calculated for the total DHI ($\alpha = 0.927$), indicating at good internal consistency with possible item redundancy.

Hypothesis Testing for Construct Validity

Hypothesis testing for construct validity was assessed for the following four hypotheses:

- (1) The hypothesis that the DHI Total score will be positively associated with the 3-point severity rating was supported ($r_s = 0.421$).
- (2) The hypothesis that patients with dysphagia will have significantly higher DHI Total scores compared with healthy control participants was supported. A Mann–Whitney U test indicated that patients scored significantly higher than control participants on the DHI Total score: Mean Rank_{Patient} = 325.25 (Sum of

Table 1 Confirmatory maximum likelihood factor analysis of the DHI [physical domain (items P1–P10), functional domain (items F1–F10), emotional domain (items E1–E10)]

Item	Factor 1	Factor 2	Factor 3	Communalities
P1	0.520	0.067	0.308	0.369
P2	0.516	−0.049	0.289	0.353
P3	0.414	−0.080	0.359	0.307
P4	0.441	−0.061	0.274	0.273
P5	0.386	0.079	0.421	0.333
P6	0.266	0.143	0.308	0.186
P7	0.439	−0.399	−0.050	0.354
P8	0.403	0.022	0.172	0.193
P9	0.283	−0.156	−0.093	0.113
P10	0.427	0.041	0.422	0.362
F1	0.651	− 0.490	−0.085	0.671
F2	0.641	− 0.549	−0.179	0.744
F3	0.606	−0.177	−0.002	0.399
F4	0.693	−0.173	0.068	0.515
F5	0.275	0.046	0.264	0.147
F6	0.669	−0.032	0.191	0.485
F7	0.587	−0.089	0.128	0.370
F8	0.634	0.058	0.238	0.461
F9	0.345	−0.072	0.285	0.205
F10	0.322	0.089	0.282	0.191
E1	0.623	0.098	−0.287	0.481
E2	0.696	0.094	−0.235	0.548
E3	0.770	0.190	−0.166	0.657
E4	0.773	0.083	0.104	0.615
E5	0.728	0.116	0.006	0.544
E6	0.561	0.297	−0.099	0.412
E7	0.496	0.378	0.024	0.389
E8	0.735	0.284	−0.193	0.659
E9	0.605	0.369	−0.218	0.550
E10	0.712	0.192	−0.187	0.579
Eigenvalue	100.070	20.1536	10.3567	
% of Total variance	310.608	40.733	50.212	
Total variance	410.554%			

As all items loaded mainly on the first factor, the assumption that the DHI is a multidimensional measure was not supported. As such, the factors do not represent different domains

Factor loadings over 0.40 appear in bold

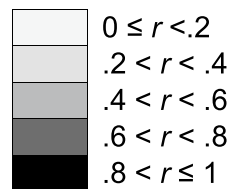
- Ranks = 10,127); Mean Rank_{Control} = 77.31 (Sum of Ranks = 128,474); $U = 1481$; $p < 0.001$, two-tailed.
- (3) The hypothesis that no significant differences on DHI scores between genders are expected was supported by a Mann–Whitney U test. No significant differences between male and female patients were identified on the DHI Total score: Mean Rank_{Male} = 191.44 (Sum of Ranks = 44,413); Mean Rank_{Female} = 206.19 (Sum of Ranks = 33,402); $U = 17,385$; $p = 0.206$, two-tailed.

Table 2 Confirmatory maximum likelihood factor analysis: heat map of the absolute values of the correlation matrix including all DHI items [physical domain (items P1–P10), functional domain (items F1–F10), emotional domain (items E1–E10)]

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	
P1	1.000	.509	.294	.340	.344	.202	.156	.233	.020	.408	.309	.238	.321	.345	.213	.356	.276	.405	.177	.227	.227	.247	.350	.406	.495	.282	.200	.389	.281	.341	P1
P2	-	1.000	.256	.354	.364	.243	.170	.316	.035	.375	.354	.319	.334	.386	.179	.303	.295	.436	.175	.133	.297	.294	.358	.425	.362	.219	.219	.338	.185	.273	P2
P3	-	-	1.000	.418	.421	.190	.172	.326	.143	.274	.276	.282	.236	.254	.148	.333	.261	.336	.262	.298	.134	.238	.255	.323	.286	.206	.215	.190	.159	.193	P3
P4	-	-	-	1.000	.397	.284	.213	.293	.053	.200	.298	.302	.249	.255	.115	.372	.136	.248	.271	.274	.200	.241	.284	.322	.326	.158	.154	.293	.211	.301	P4
P5	-	-	-	-	1.000	.261	.154	.310	.089	.326	.157	.129	.224	.270	.188	.291	.180	.307	.204	.233	.144	.155	.254	.303	.328	.148	.179	.251	.209	.227	P5
P6	-	-	-	-	-	1.000	.078	.199	.085	.287	.131	.009	.129	.159	.197	.201	.162	.184	.189	.245	.116	.157	.155	.235	.156	.150	.227	.161	.233	.146	P6
P7	-	-	-	-	-	-	1.000	.197	.358	.173	.461	.524	.295	.383	.127	.345	.323	.266	.189	.028	.258	.274	.248	.265	.209	.144	.099	.184	.164	.300	P7
P8	-	-	-	-	-	-	-	1.000	.079	.268	.247	.210	.210	.221	.060	.186	.224	.256	.320	.187	.279	.306	.335	.287	.254	.189	.183	.301	.204	.236	P8
P9	-	-	-	-	-	-	-	-	1.000	.083	.283	.255	.157	.232	.130	.195	.170	.091	.181	.050	.180	.208	.202	.172	.176	.178	.158	.121	.146	.257	P9
P10	-	-	-	-	-	-	-	-	-	1.000	.230	.177	.264	.291	.269	.368	.326	.402	.249	.257	.180	.221	.230	.387	.283	.236	.255	.246	.195	.188	P10
F1	-	-	-	-	-	-	-	-	-	-	1.000	.730	.465	.515	.060	.375	.383	.372	.293	.199	.387	.436	.370	.442	.430	.265	.129	.388	.252	.361	F1
F2	-	-	-	-	-	-	-	-	-	-	-	1.000	.504	.491	.141	.407	.363	.296	.168	.118	.372	.396	.430	.414	.426	.227	.131	.365	.244	.378	F2
F3	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.492	.111	.450	.389	.332	.216	.153	.307	.374	.476	.463	.412	.313	.230	.384	.287	.419	F3
F4	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.255	.523	.609	.508	.248	.145	.389	.466	.485	.559	.500	.339	.266	.424	.289	.450	F4
F5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.357	.270	.279	.136	.204	.121	.100	.166	.220	.181	.147	.198	.124	.177	.134	F5
F6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.535	.460	.300	.287	.338	.433	.485	.593	.476	.315	.359	.397	.339	.465	F6
F7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.484	.284	.149	.333	.429	.361	.519	.322	.297	.315	.337	.295	.358	F7
F8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.283	.198	.341	.388	.441	.577	.457	.350	.331	.470	.304	.413	F8
F9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.368	.128	.185	.188	.262	.212	.153	.228	.165	.086	.215	F9
F10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.118	.182	.206	.282	.263	.216	.236	.207	.166	.216	F10
E1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.698	.554	.464	.356	.361	.239	.583	.467	.462	E1
E2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.623	.518	.457	.399	.343	.547	.485	.540	E2
E3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.671	.631	.464	.401	.613	.563	.623	E3
E4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.599	.391	.403	.550	.432	.515	E4
E5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.448	.366	.580	.487	.593	E5
E6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.615	.545	.492	.437	E6
E7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.472	.489	.409	E7
E8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.611	.637	E8
E9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	.567	E9
E10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.000	E10

Note. Determinant = 6.225E-7

Legend:



(4) The hypothesis that age is not positively associated with the DHI Total scores was assessed using linear regression analysis. A linear regression was calculated to predict the DHI Total score based on age. No significant regression equation was found: $\beta = 0.142$; $t = 1.917$; $p = 0.56$. A non-significant regression equation was found: $R^2 = 0.009$; $F(1389) = 3.673$; $p = 0.056$.

Criterion Validity (Diagnostic Performance)

Data from the healthy control group were used to calculate cut-off points to distinguish between people with and without

dysphagia. Different cut-off points were calculated using 1 SD, 1.5 SD, 2 SD and 2.5 SD, above the mean score of the DHI Total scores (Mean = 2.48; SD 2.67). A cut-off point of 2 SD above the mean score is common practice in the literature [25] and showed most optimal diagnostic performance for the DHI when balancing between optimal sensitivity and specificity scores. Using a cut-off of 8 [MN + 2SD = 2.48 + 2(2.67) = 7.82], the diagnostic performance of the DHI can be summarised as follows: sensitivity = 92.7%; specificity = 94.7%; positive predictive value = 98.1%; negative predictive value = 81.0%; likelihood ratio positive = 17.49; likelihood ratio negative = 0.08. The ROC analysis revealed

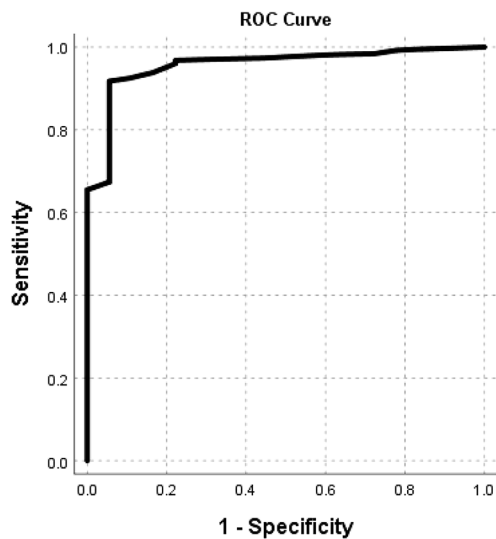


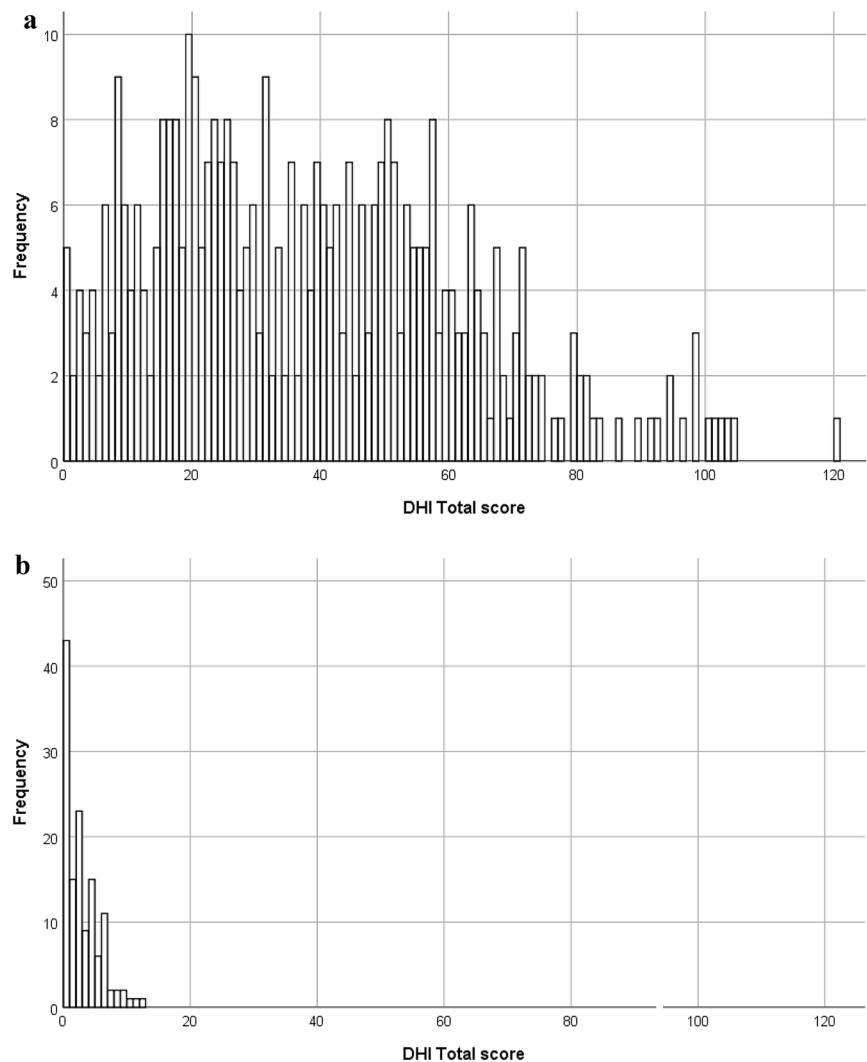
Fig. 1 ROC curve of the DHI Total score

that the DHI Total score covers 0.957 (95.7%) of the area under the curve ($p < 0.001$; lower bound = 0.920; upper bound = 0.993; Fig. 1).

Interpretability

Means and standard deviations for the DHI Total score were determined for both clinical and healthy controls (see hypothesis testing for construct validity). In addition, floor and ceiling effects were assessed (see Fig. 2). The DHI Total score can range between 0 and 12. Findings demonstrated that the percentage of participants achieving the lowest possible score for the DHI Total score was 1.3% ($n = 5/395$). The percentage of participants achieving the highest possible score for the DHI Total score was 0.2% ($n = 1/395$). The DHI Total score did not show any floor or ceiling effects.

Fig. 2 **a** Data distribution for the DHI Total score for the patient group (no floor and ceiling effects). **b** Data distribution for the DHI Total score for the control group



Discussion

In this study, the psychometric properties of the DHI, a patient self-report measure of FHS and HR-QoL in dysphagia, were evaluated using CTT. The COSMIN taxonomy, terminology and definitions of measurement properties for health-related patient-reported outcome measures were used to guide the psychometric evaluation of the DHI.

The internal consistency of the DHI was good with possible item redundancy. Both factor analyses and the high Cronbach's alpha of the DHI total score supported the assumption that the DHI is a unidimensional measure. All items target the same underlying construct. Consequently, we do not recommend the use of subscale scores as these do not reflect the underlying unidimensional structure of the DHI. Items representing different deglutition-related aspects of daily life and referring to either psychosocial aspects of dysphagia (Emotional subscale), functional aspects (Functional subscale) or physical aspects (Physical subscale) were interrelated and could not be disentangled from each other. The different aspects as defined and represented by the 30 items of the DHI proved to be intertwined making it difficult to break them up into different subscales. This means that even though the three aspects that make up the theoretical constructs of FHS and HR-QoL as operationalised by the DHI, these aspects cannot be regarded as independent factors, rendering the use of the subscales superfluous. The relatively low percentage of explained variance indicates that the current DHI items do not sufficiently explain the model.

Hypothesis testing for construct validity confirmed all four hypotheses. DHI Total scores were positively correlated with severity ratings by the phoniatrician. Patients with swallowing problems could be distinguished from healthy controls as demonstrated by significantly higher DHI Total scores. Also, as expected, no gender or age differences were associated with the DHI Total scores. DHI scores of healthy control participants were used to calculate cut-off scores to distinguish between people with and without dysphagia. The diagnostic performance of the DHI was determined using FEES and/or VFS as reference tests. Diagnostic performance proved excellent as confirmed by high sensitivity and specificity values, and a high value for the area under the curve as determined by ROC analysis. Histograms confirmed positive skewed data distribution for healthy controls. The DHI showed no floor and ceiling effects which indicated that no items were missing in the lower or upper end of the scale, thus not eroding content validity or reliability [18].

A psychometric review of FHS questionnaires in dysphagia identified poor methodological quality scores for most measurement properties of current published questionnaires [9]. Another psychometric review of HR-QoL questionnaires identified similar shortcomings, but made some preliminary

recommendations about questionnaires showing the strongest ratings on psychometric criteria [3]. However, more recent studies applying Item Response Theory (IRT) using Rasch analysis [26–28] contradicted these recommendations and, instead, recommended redevelopment of some of the most commonly used self-report measures like the Eating Assessment Tool (EAT-10) [29] and the Swallowing Quality of Life Questionnaire (SWAL-QoL) [30] due to significant psychometric shortcomings. IRT is a contemporary methodology to evaluate the psychometric quality of measures. In IRT, the unit of analysis is the item and results are not bound by the test population. By contrast, the CTT framework evaluates the performance of the measure as a whole and psychometric findings are specific to the sample population used to evaluate the measure [31, 32]. Conversely, even though CTT has obvious limitations, the procedures and interpretation of CTT are relatively straight forward compared with IRT and are considered a useful first step when exploring the psychometric properties of a measure.

Recent psychometric reviews and studies identified and reaffirmed the urgent need to develop new, robust self-report measures. When comparing the psychometric characteristics of the DHI with the results of existing psychometric reviews and studies on self-report measures in dysphagia (e.g. [3, 9, 27, 33]), the DHI seems to be one of the most robust self-report measures on dysphagia-related FHS and HR-QoL published to date. The DHI presents with promising psychometric properties, whereas many comparable measures (e.g. the EAT-10 and the SWAL-QoL) lack sufficient psychometric evidence and do not meet the psychometric criteria for measurement properties [26–28, 33–35].

To support clinicians and researchers, Prinsen et al. [19] published a consensus-based practical guideline on methods for selecting outcome measures, involving four main steps: (1) clarifying conceptual considerations prior to searching for measures about the underlying construct and target population; (2) identifying existing outcome measures using recent systematic reviews, literature searches and other information sources; (3) conducting quality assessment of outcome measures by means of evaluation of measurement properties and feasibility aspects of the identified measures; and (4) following generic recommendations on the selection of outcome measures (e.g. minimum requirements for content validity and internal consistency).

Limitations and Future Research

As no repeated measurements were available, no measurement error or reliability over time or test–retest reliability could be determined. Also, responsiveness was outside the scope of this study and should be determined in future studies. In addition, only CTT was used to determine the psychometric properties of the DHI. The use of CTT

should be considered as the first step in the psychometric evaluation of the DHI. Future research should extend analyses combining CTT with Item Response Theory (IRT) [6, 31, 36], thus adding to the current psychometric evidence of the DHI.

Conclusion

Based on all CTT analyses, we conclude that the DHI is a promising self-report measure of FHS and HR-QoL in dysphagia. The DHI is a unidimensional measure with weak to moderate structural validity and has good internal consistency and strong hypothesis testing for construct validity. In addition, the DHI demonstrates excellent diagnostic performance as part of criterion validity. The reported analyses indicate that the DHI is an appropriate choice as a patient self-report measure to evaluate FHS and HR-QoL in dysphagia. We recommend ongoing validation to assess the measure for possible item redundancy and to assess the dimensionality of the DHI (structural validity) using item response theory.

Acknowledgements We would like to thank Dr Sabine Crestani and Dr Pascale Fichaux Bourin (ENT phoniatricians) for assisting in patient recruitment at the Voice and Deglutition Unit, Department of Otorhinolaryngology and Head and Neck Surgery, Larrey Hospital, University Hospital of Toulouse, Toulouse, France.

Funding Open Access funding provided by University of Oslo (incl Oslo University Hospital).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Speyer R. Oropharyngeal dysphagia: screening and assessment. *Otolaryngol Clin North Am* . 2013;46(6):989–1008.
- Baijens LWJ, et al. ESSD-EUGMS white paper: oropharyngeal dysphagia as a geriatric syndrome. *Clin Interv Aging*. 2016;11:1403–28.
- Timmerman AA, et al. Psychometric characteristics of health-related quality-of-life questionnaires in oropharyngeal dysphagia. *Dysphagia*. 2014;29(2):183–98.
- Ferrans CE, et al. Conceptual model of health-related quality of life. *J Nurs Scholarsh* . 2005;37(4):336–42.
- Mokkink LB, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* . 2010;63(7):737–45.
- Mokkink LB, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* . 2018;27(5):1171–9.
- Terwee CB, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* . 2018;27(5):1159–70.
- Prinsen CAC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* . 2018;27(5):1147–57.
- Speyer R, et al. Psychometric properties of questionnaires on Functional Health Status in oropharyngeal dysphagia: a systematic literature review. *Biomed Res Int* . 2014;2014:11. <https://doi.org/10.1155/2014/458678>.
- Woisard V, Andrieux MP, Puech M. Validation of a self-assessment questionnaire for swallowing disorders (Deglutition Handicap Index). *Revue de Laryngol Otol Rhinol* . 2006;127(5):315–25.
- Speyer R, et al. Quality of life in oncological patients with oropharyngeal dysphagia: validity and reliability of the Dutch version of the MD Anderson Dysphagia Inventory and the Deglutition Handicap Index. *Dysphagia*. 2011;26(4):407–14.
- Woisard V, Lepage B. The “Deglutition Handicap Index” a self-administrated dysphagia-specific quality of life questionnaire: temporal reliability. *Revue de Laryngol Otol Rhinol* . 2010;131(1):19–22.
- Crestani S, Moerman M, Woisard V. The “Deglutition Handicap Index” a self-administrated dysphagia-specific quality of life questionnaire: sensibility to change. *Revue De Laryngol Otol Rhinol* . 2011;132:3–7.
- Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2003.
- Swan K, et al. Psychometric properties of visuoperceptual measures of videofluoroscopic and fibre-endoscopic evaluations of swallowing: a systematic review. *Dysphagia*. 2019;34(1):2–33.
- Bouter LM, Zielhuis GA, Zeegers MPA. Diagnostic and prognostic research. *Textbook of epidemiology*. Houten: Bohn Stafleu van Loghum; 2007. p. 171–201.
- Terwee CB, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
- Mokkink LB, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs)—user manual (Version 1.0). https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf (2018). Accessed 6 Apr 2020.
- Prinsen CAC, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—a practical guideline. *Trials*. 2016;17:10. <https://doi.org/10.1186/s13063-016-1555-2>.
- Sarstedt M, Mooi E. A concise guide to market research: the process, data, and methods using IBM SPSS Statistics. Berlin: Springer; 2014. p. 346.
- Hair JF, et al. Multivariate data analysis. Cengage Learning EMEA: Andover; 2018. p. 832.
- Field A. Exploratory factor analysis. *Discovering statistics using IBM SPSS statistics*. Los Angeles: SAGE; 2018. p. 777–833.
- Vatcheva KP, et al. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale)*. 2016;6(2):20.

24. Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Routledge; 1988. p. 567.
25. Indrayan A. *Medical biostatistics*. Boca Raton: Chapman and Hall/CRC; 2013. p. 1024.
26. Cordier R, et al. Evaluating the psychometric properties of the Eating Assessment Tool (EAT-10) using Rasch analysis. *Dysphagia*. 2017;32(2):250–60.
27. Kean J, et al. An introduction to item response theory and Rasch analysis: Application using the eating assessment tool (EAT-10). *Brain Impair* . 2018;19(1):91–102.
28. Wilmskoetter J, et al. Construct validity of the eating assessment tool (EAT-10). *Disabil Rehabil* . 2019;41(5):549–59.
29. Belafsky PC, et al. Validity and reliability of the Eating Assessment Tool (EAT-10). *Ann Otol Rhinol Laryngol* . 2008;117(12):919–24.
30. McHorney CA, et al. The SWAL–QOL and SWAL–CARE outcomes tool for oropharyngeal dysphagia in adults: III. Documentation of reliability and validity. *Dysphagia*. 2002;17(2):97–114.
31. Linacre J. *A user’s guide to Winsteps Rasch-model computer programs: program manual 3.92.0*. Chicago: Mesa-Press II; 2016.
32. Aryadoust V, Tan HAH, Ng LY. A Scientometric review of Rasch measurement: the rise and progress of a specialty. *Front Psychol* . 2019;10:16. <https://doi.org/10.3389/fpsyg.2019.02197>.
33. Cordier R, et al. Using Rasch analysis to evaluate the reliability and validity of the Swallowing Quality of Life questionnaire: an item response theory approach. *Dysphagia*. 2018;33(4):441–56.
34. Simpelaere IS, et al. Validation to the Dutch version of the Swallowing Quality-of-Life Questionnaire (SWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health Qual Life Outcomes*. 2017;15:16. <https://doi.org/10.1186/s12955-017-0639-3>.
35. Hansen T, Kjaersgaard A. Item analysis of the Eating Assessment Tool (EAT-10) by the Rasch model: a secondary analysis of cross-sectional survey data obtained among community-dwelling elders. *Health Qual Life Outcomes*. 2020;18(1):139–139.
36. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res* . 2007;16(Suppl 1):5–18.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Renée Speyer PhD

Reinie Cordier PhD

Clara Bouix MSc

Yohan Gallois MSc

Virginie Woisard PhD