

REVIEW

A scoping review of Rasch analysis and item response theory in otolaryngology: Implications and future possibilities

David T. Liu MD, PhD¹  | Christian A. Mueller MD¹  | Ahmad R. Sedaghat MD, PhD² 

¹Department of Otorhinolaryngology, Head and Neck Surgery, Medical University of Vienna, Vienna, Austria

²Department of Otolaryngology—Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

Correspondence

David T. Liu, Department of Otolaryngology—Head and Neck Surgery, Medical University of Vienna/Vienna General Hospital, Waehringer Guertel 18-20, 1090 Vienna, Austria.
Email: david.liu@meduniwien.ac.at

Abstract

Objective: Item response theory (IRT) is a methodological approach to studying the psychometric performance of outcome measures. This study aims to determine and summarize the use of IRT in otolaryngological scientific literature.

Methods: A systematic search of the Medline, Embase, and the Cochrane Library databases was performed for original English-language published studies indexed up to January 28, 2023, per the following search strategy: (“item response theory” OR “irt” OR “rasch” OR “latent trait theory” OR “modern mental test theory”) AND (“ent” OR “otorhinolaryngology” OR “ear” OR “nose” OR “throat” OR “otology” OR “audiology” OR “rhinology” OR “laryngology” OR “neurotology” OR “facial plastic surgery”).

Results: Fifty-five studies were included in this review. IRT was used across all subspecialties in otolaryngology, and most studies utilizing IRT methodology were published within the last decade. Most studies analyzed polytomous response data, and the most commonly used IRT models were the partial credit and the rating scale model. There was considerable heterogeneity in reporting the main assumptions and results of IRT.

Conclusion: IRT is increasingly being used in the otolaryngological scientific literature. In the otolaryngology literature, IRT is most frequently used in the study of patient-reported outcome measures and many different IRT-based methods have been used. Future IRT-based outcome studies, using standardized reporting guidelines, might improve otolaryngology-outcome research sustainably by improving response rates and reducing patient response burden.

Level of evidence: 2.

KEY WORDS

ENT, IRT, item response theory, otolaryngology, Rasch, systematic review

1 | INTRODUCTION

Over the last few decades, patient-reported outcomes (PROs) have become fundamental not only in research but also during clinical practice, as they directly involve patients in the care they receive.^{1,2} Although there is an expanding prominence and growing use of PROs in clinical practice globally, there is evidence of performance-related issues during the development of many measurement tools.^{3,4} The use of PROs of high quality is also a vital pillar of any clinical outcome research.^{5,6} The quality of these instruments is largely determined by their psychometric/measurement properties, such as reliability (i.e., consistency between different measurements/observations) and validity (i.e., measuring the construct that the instrument is intended to measure). The importance of robust psychometric properties of existing PROs is further highlighted, as studies have shown that those with very poor or unknown measurement properties might overestimate treatment effects, significantly negatively affecting clinical outcome research.⁴

Traditionally, the development and refinement of PROs and the evaluation of their psychometric properties have been done using methods derived from a psychometric approach called “Classical Test Theory” (CTT).⁷ CTT assumes that a test subject's total score equals the true score plus a random measurement error, with the true score being immeasurable.⁸ The results of CTT analytical methods depend upon on total test scores and the group (e.g., cohort of patients/test subjects) in which those test scores were determined.⁹ Well-known measures of CTT include Cronbach's alpha, Pearson Correlation Coefficient for reliability measures, stepwise regression, or factor analysis.¹⁰ Although methods of CTT are established and remain the dominant paradigm, a new group of mathematical models, including Item Response Theory (IRT, also called modern or latent test theory)¹¹⁻¹⁶ and Rasch analysis (RA),¹⁷ have gained increasing popularity in clinical outcome research due to their methodological advantages over CTT. Because IRT and RA assess the relationship of individual items of a PRO to the underlying construct being measured (e.g., ability or health-related quality of life), both models offer great detail about the information each item contributes. This information is critical during PRO refinement. Another advantage of IRT over CTT is that IRT-derived PRO performance metrics are constant from one population to the next, whereas in CTT they are sample-dependent.¹⁸

Despite the increasingly recognized importance of IRT and RA in PRO validation throughout the biomedical literature, the utilization of IRT and RA as complementary approaches to CTT is unclear as it relates to PROs described in the otolaryngology literature.¹⁸ In this study, we sought to identify and summarize the studies in the field of otolaryngology that used IRT and RA methodology using a systematic approach. We used this information to gain a clear and comprehensive overview of available evidence on novel test methods in otolaryngology outcome research. We believe the results of this study will serve in several important capacities: (1) to inform the field of otolaryngology about IRT, (2) to provide a contemporary accounting of how IRT has been used in otolaryngology PROMs to date, and (3) to inform future studies and improve methodological approaches to otolaryngology outcome research.

2 | METHODS

2.1 | Search strategy

We performed a systematic review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.¹⁹ Two independent reviewers (D.T.L. and A.R.S.) searched Medline, Embase, and the Cochrane Library on item response theory in otolaryngology literature. The search strategy included a combination of keywords for the following two concepts: item response theory and otolaryngology. We searched databases mentioned above from inception to January 28, 2023, for the following search terms: (item response theory OR irt OR rasch OR latent trait theory OR modern mental test theory) AND (ent OR otorhinolaryngology OR ear OR nose OR throat OR otology OR audiology OR rhinology OR laryngology OR neurotology OR facial plastic surgery).

2.2 | Selection criteria

Studies should have the following criteria to be eligible for this systematic review: (i) original article, (ii) English language, (iii) use of IRT methodology, and (iv) content related to otolaryngology literature. We excluded short reports and communications from the final qualitative synthesis.

2.3 | Data extraction

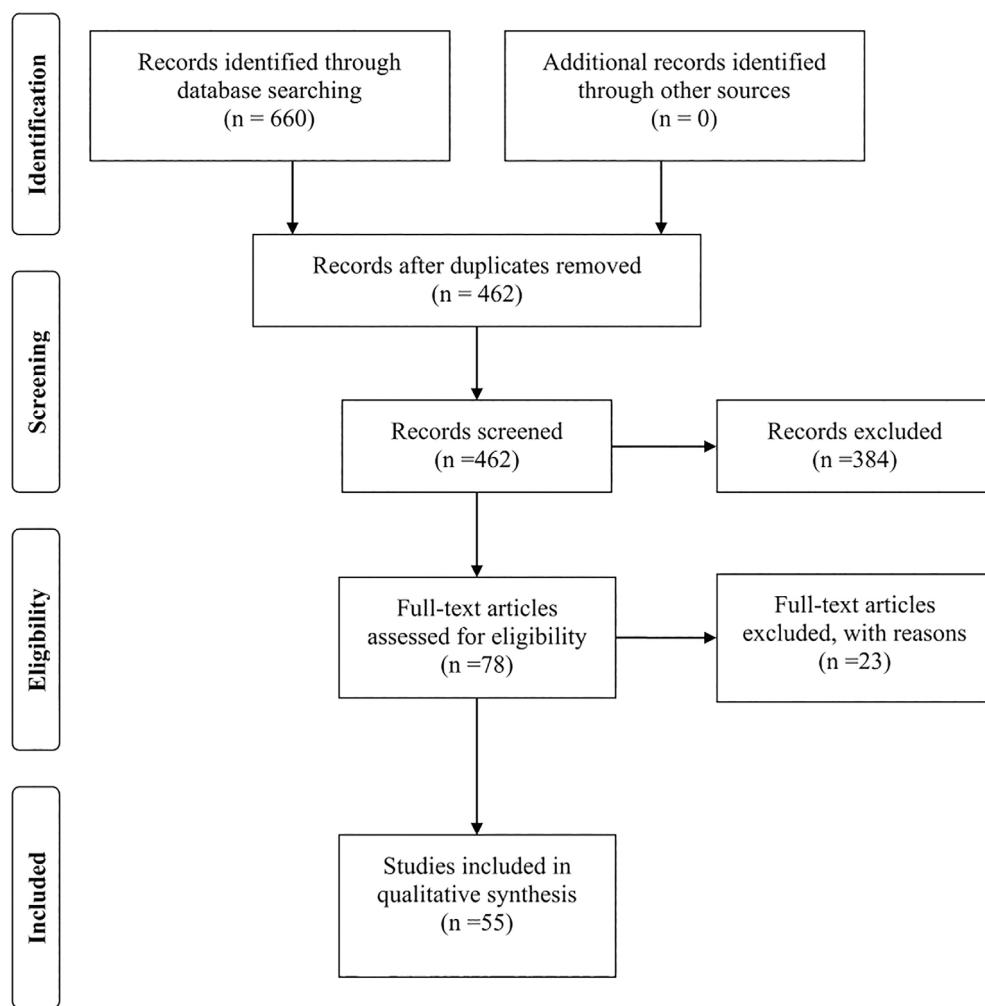
Following inclusion and full-text review, two reviewers (D.T.L. and A.R.S.) independently extracted the following information when available: author, year of publication, IRT model used, software used for IRT analysis, study design (i.e., quantitative or mixed methods study), reporting goodness of fit testing, reporting individual item parameters, assessment of local independence, assessment of unidimensionality, confirmatory factors analysis (CFA) used on data, sample size, response type (i.e., dichotomous or polytomous), type of data (i.e., ability, attitude, appearance, performance, QoL), rational for using IRT, content area, sample location, and scales used.

3 | RESULTS

3.1 | Search and study characteristics

The initial screening process returned 660 citations, from which we removed 198 due to duplicate results. We screened titles and abstracts of 462 articles and included 76 articles for full-text review (Figure 1). Finally, 55 articles met our inclusion criteria and were included in the review (Table 1). All studies were original articles and used IRT methodology in otolaryngology literature. Included studies were published between 1988 and 2022.

FIGURE 1 PRISMA flowchart.



3.2 | Sample size and participant locations

The sample size of participants included in all studies ranged from 40 to 13,273 (median: 309, IQR = 383.3). IRT methodology was used on patients and participants recruited from almost all continents (Figure 2).

3.3 | Which otolaryngology subspecialties are using IRT methods?

IRT models were used across all subspecialties in otolaryngology. Eighteen studies were related to otology/neurotology/audiology (32.7%), followed by 13 in facial plastics (23.6%), 12 in the field of rhinology (21.8%), five studies in laryngology (9.1%), four in head and neck surgery (7.3%), and three in pediatric otolaryngology (5.5%).

3.4 | Increasing popularity of IRT methods in otolaryngology

Although IRT methodology was first used in otolaryngology literature in 1988, it was not until 2012 that IRT was used across all

subspecialties. Within the last 5 years (2017–2022), 35 publications were using IRT methodology, and in the previous 10 years (2012–2022), there were 50 articles, which gives an average of 5 articles per year, respectively, although there appears to be increasing utilization of IRT in recent years (Figure 3).

3.5 | Popular models and theoretical approaches

Eighteen studies (32.7%) mentioned “Rasch analysis/model” or “Rasch measurement theory”, followed by the partial credit model ($n = 11$, 20%), the rating scale model ($n = 10$, 18.2%), the graded response model/GRM ($n = 9$, 16.4%), and the two-parameter logistic model (2PL), which was used in four studies (7.3%). The Bayesian IRT model, the Mokken scale analysis, and the dichotomous logistic model were used once each (1.8% each) (Table 2).

3.6 | Popular software used for IRT methodology

The statistical software used for IRT analysis was reported in all studies. The most popular software was RUMM2030[®],²⁰ used in

TABLE 1 Table of evidence.

Authors	Year	Model used	Tool/technique used	Study design	Testing goodness of fit	Reporting individual item parameters	Assessing fit for multiple models	Assessing local independence	Assessing unidimensionality
Bement et al. ³⁷	1988	Rasch measurement model	BiCAL	Mixed methods	nr	Yes	nr	nr	nr
Bochner et al. ³⁸	1992	Rasch measurement model	BiCAL	Quantitative study	nr	Yes	nr	nr	nr
Klassen et al. ³⁹	2021	Rasch measurement theory	RUMM2030	Quantitative study	nr	Yes	nr	nr	nr
La Padula et al. ⁴⁰	2022	Rasch measurement theory	RUMM2030	Mixed methods	Yes	Yes	nr	nr	nr
Lauritsen et al. ⁴¹	2015	Rasch model	DIGRAM	Quantitative study	Yes	nr	nr	nr	nr
Pusic et al. ⁴²	2013	Rasch model	RUMM2030	Mixed methods approach	Yes	Yes	nr	nr	nr
Leong et al. ⁴³	2005	Rasch model	WINSTEPS	Quantitative study	nr	Yes	nr	nr	nr
Abdullah et al. ⁴⁴	2020	PL	R software	Mixed methods	nr	nr	nr	nr	nr
Sucharew et al. ⁴⁵	2012	PL	SAS and R Software	Quantitative study	Yes	Yes	Yes	nr	nr
Leijon et al. ⁴⁶	2021	Bayesian IRT model	Python	Quantitative study	nr	nr	nr	nr	nr
Liu et al. ⁴⁷	2022	CRM	R Software	Quantitative methods	Yes	Yes	Yes	nr	nr
Crump et al. ⁴⁸	2015	GRM	SAS and Mplus	Quantitative study	nr	Yes	nr	nr	nr
Apon et al. ⁴⁹	2020	Partial credit model	RUMM2030	Quantitative study	Yes	nr	nr	nr	nr
Dehqan et al. ⁵⁰	2016	Rating scale model	WINSTEPS	Mixed methods	Yes	nr	nr	nr	nr
Atallah et al. ⁵¹	2019	Rating scale model	STATA	Mixed methods	Yes	Yes	nr	nr	nr
Wulff et al. ⁵²	2020	Rasch analysis	DIGRAM	Mixed methods	nr	nr	nr	Yes	nr
Klassen et al. ⁵³	2016	Rasch measurement theory	RUMM2030	Mixed methods (qualitative followed by quantitative methods)	Yes	Yes	nr	Yes	nr
Lee et al. ⁵⁴	2018	Rasch measurement theory	RUMM2030	Quantitative study	Yes	nr	nr	Yes	nr
Klassen et al. ⁵⁵	2013	Rasch measurement theory	RUMM2030	Mixed methods	Yes	Yes	nr	Yes	nr
Klassen et al. ⁵⁶	2016	Rasch measurement theory	RUMM2030	Mixed methods approach	Yes	Yes	nr	Yes	nr
Klassen et al. ⁵⁷	2015	Rasch measurement theory	RUMM2030	Mixed methods (qualitative followed by quantitative methods)	Yes	Yes	nr	Yes	nr
Klassen et al. ⁵⁸	2016	Rasch measurement theory analysis	RUMM2030	Quantitative	Yes	Yes	nr	Yes	nr
Klassen et al. ⁵⁹	2021	Partial credit model	RUMM2030	Mixed methods	Yes	Yes	nr	Yes	nr

TABLE 1 (Continued)

Authors	Year	Model used	Tool/technique used	Study design	Testing goodness of fit	Reporting individual item parameters	Assessing fit for multiple models	Assessing local independence	Assessing unidimensionality
Klassen et al. ⁶⁰	2021	Partial credit model	RUMM2030	Mixed methods (qualitative followed by quantitative methods)	Yes	Yes	nr	Yes	nr
Klassen et al. ⁶⁰	2022	Partial credit model	RUMM2030	Mixed methods	Yes	Yes	nr	Yes	nr
Dobbs et al. ⁶¹	2020	Rasch model	R Software	Quantitative	Yes	nr	nr	Yes	Yes
Vitoratou et al. ⁶²	2021	PL	R Software	Quantitative	Yes	Yes	nr	nr	Yes
Waalboer-Spuij et al. ⁶³	2018	PL	R Software	Mixed methods	Yes	Yes	nr	nr	Yes
Bogaardt et al. ⁶⁴	2007	Dichotomous (logistic) Rasch model	OPLM	Quantitative study	Yes	Yes	nr	nr	Yes
Killan et al. ⁶⁵	2019	GRM	STATA	Quantitative study	Yes	nr	nr	nr	Yes
Coeih et al. ⁶⁶	2021	GRM	R Software	Mixed methods	Yes	Yes	nr	nr	Yes
Liu et al. ⁶⁷	2021	GRM	R Software	Quantitative methods	Yes	Yes	Yes	nr	Yes
Liu et al. ⁶⁸	2021	GRM	R Software	Quantitative methods	Yes	Yes	Yes	nr	Yes
Zhao et al. ⁶⁹	2022	Partial credit model	ACER ConQuest	Quantitative study	Yes	Yes	nr	nr	Yes
Dye et al. ⁷⁰	2013	Rating scale model	WINSTEPS	Quantitative study	Yes	Yes	nr	nr	Yes
Cordier et al. ⁷¹	2017	Rating scale model	WINSTEPS	Quantitative study	Yes	nr	nr	nr	Yes
Stewart et al. ⁷²	2015	Rating scale model	Jmetrik	Quantitative study	Yes	Yes	nr	nr	Yes
Niklaassen et al. ⁷³	2022	Rasch analysis	DIGRAM and SAS	Mixed methods	Yes	nr	nr	Yes	Yes
Lee et al. ⁷⁴	2017	Rasch model	ACER ConQuest	Mixed methods	Yes	Yes	nr	Yes	Yes
Simpelaere et al. ⁷⁵	2017	Rasch model	RUMM2030	Quantitative study	Yes	nr	nr	Yes	Yes
Chenault et al. ⁷⁶	2016	GRM	IRTPRO	Quantitative methods	Yes	Yes	nr	Yes	Yes
Yang et al. ⁷⁷	2020	GRM	R Software	Quantitative study	Yes	Yes	nr	Yes	Yes
Jessen et al. ⁷⁸	2018	GRM	STATA	Quantitative study	Yes	Yes	nr	Yes	Yes
Priester et al. ⁷⁹	2013	Mokken analysis	MSP—Mokken Scale program (STATA module)	Quantitative study	Yes	Yes	nr	Yes	Yes
Mueller et al. ⁸⁰	2015	Partial credit model	RUMM2030	Quantitative study	Yes	nr	nr	Yes	Yes
Franchignoni et al. ⁸¹	2010	Partial credit model	WINSTEPS	Quantitative study	Yes	Yes	nr	Yes	Yes
Hughes et al. ⁸²	2021	Partial credit model	WINSTEPS JMLE	Quantitative study	Yes	Yes	nr	Yes	Yes
Heffernan et al. ⁸³	2020	Partial credit model	RUMM2030	Quantitative study	Yes	Yes	Yes	Yes	Yes

(Continues)

TABLE 1 (Continued)

Authors	Year	Model used	Tool/technique used	Study design	Testing goodness of fit	Reporting individual item parameters	Assessing fit for multiple models	Assessing local independence	Assessing unidimensionality
Molinengo et al. ⁸⁴	2017	Partial credit model and rating scale model	WINSTEPS	Quantitative study	Yes	Yes	Yes	Yes	Yes
Rodriguez et al. ⁸⁵	2019	Partial credit Rasch model	RUMM2030	Mixed methods approach	Yes	nr	nr	Yes	Yes
Declau et al. ⁸⁶	2021	Rating scale model	WINSTEPS	Mixed methods approach	Yes	Yes	nr	Yes	Yes
McRackan et al. ⁸⁷	2022	Rating scale model	WINSTEPS	Mixed methods approach	Yes	nr	nr	Yes	Yes
Graboyes et al. ⁸⁸	2020	Rating scale model	WINSTEPS	Mixed methods approach	Yes	nr	nr	Yes	Yes
McRacken et al. ⁸⁹	2019	Rating scale model	WINSTEPS	Quantitative study	Yes	nr	nr	Yes	Yes
McRacken et al. ⁹⁰	2019	Rating scale model	WINSTEPS	Quantitative study	Yes	Yes	nr	Yes	Yes
CFA used	Sample size (n) ^a	Response format	Type of data	Item focus	Rationale for using IRT programs	ENT content	Sample location	Author type	Scale used
nr	46	Dichotomous	Performance	Item focus	To analyze the Discrimination Test responses	Otology/Neurology/Audiology	USA	Clinical	The Lipreading Discrimination Test
nr	40	Dichotomous	Performance	Item focus	To explore construct validity	Pediatric otolaryngology	USA	Clinical	Speech Sound Pattern Discrimination Test (SSPDT)
nr	1495	Polytomous	Appearance, adverse effects, QoL, and performance	Item focus	To examine reliability and validity	Facial plastics	Australia, Brazil, Canada, Chile, France, Ireland, Sweden, UK, and USA	Clinical	FACE-Q Craniofacial module for children and young adults with facial conditions
nr	1000	Polytomous	Performance	Item focus	To analyze the Face-and Neck-Lift Objective Photo-Numerical Assessment Scale	Facial plastics	France	Clinical	The Face-and Neck-Lift Objective Photo-Numerical Assessment Scale
nr	339	Dichotomous	Performance	Item focus	To evaluate the reliability and internal construct validity and to utilize the Rasch model as an analytical approach for further development of the scale	Otology/Neurology/Audiology	Denmark	Clinical	The Galkin test
nr	100	Polytomous	QoL	Item focus	To develop and evaluate the psychometric properties of the core FACE-Q scale	Facial plastics	France, Germany, UK, and Israel	Clinical	FACE-Q Satisfaction with Appearance Scale

TABLE 1 (Continued)

CFA used on data	Sample size (n) ^a	Response format	Type of data	Person versus item focus	Rationale for using IRT programs	ENT content	Sample location	Author type	Scale used
nr	43	Polytomous	QoL	Item focus	To investigate the psychometric properties	Rhinology	Singapore	Clinical	Rhinoconjunctivitis Quality-of-Life Questionnaire (RQLQ) and the Medical Outcome Short-Form 36 (SF-36)
nr	302	Dichotomous and polytomous	Attitude	Item focus	To evaluate validity and reliability	Rhinology	Malaysia	Clinical	Perception, Attitude, and Practice of primary care practitioner toward allergic rhinitis practice guidelines (PAP-PCP questionnaire)
nr	537	Dichotomous	Performance	Item focus	To evaluate the application of applying IRT methods to skin prick test data	Rhinology	USA	Clinical	Atopy score
nr	13,273	Polytomous	Performance	Item focus	To consider the assumption that ordinal responses have interval-scale properties	Otology/Neurology/Audiology	Australia, UK, Germany, and Sweden	Clinical	International Outcome Inventory for Hearing Aids (IOI-HA)
nr	181	Polytomous	QoL	Item focus	To evaluate psychometric properties	Rhinology	USA	Clinical	The Sinonasal Outcome Test-22 (SNOT-22)
nr	223	Polytomous	QoL	Item focus	To conduct an item-level analysis of the SNOT-22	Rhinology	Canada	Clinical	The Sinonasal Outcome Test-22 (SNOT-22)
nr	714	Polytomous	Appearance, QoL, performance	Item focus	To provide insights into the performance of the PROMs	Facial plastics	Canada, UK, and USA	Clinical	Cleft-Q, Child Oral Health Impact Profile-Oral-Symptom Scale (COHIP-OSS), Nasal Obstructive Symptom Evaluation questionnaire (NOSE), Intelligibility
nr	160	Polytomous	QoL	Item focus	To take advantage of the Rasch model characteristics	Laryngology	Iran	Clinical	Iran Voice Quality of Life Profile
Yes	165	Polytomous	Appearance and performance	Item focus	To evaluate psychometric properties	Rhinology	Canada	Clinical	The Standardized Cosmesis and Health Nasal Outcome Survey (SCHNOS)
Yes	80	Polytomous	QoL	Item focus	To test the fit between observed item mean scores and expected item mean scores	Laryngology	Denmark	Clinical	The Voice-Related Quality of Life Instrument (V-RQOL)
nr	503	Polytomous	Appearance and adverse effects	Item focus	To make decisions about the overall quality of the scale	Facial plastics	USA and Canada	Clinical	FACE-Q Skin, Lips and Facial Rhytides Appearance Scales
nr	209	Polytomous	Appearance and QoL	Item focus	To analyze and validate the PRO using a modern psychometric approach	Facial plastics	USA	Clinical	FACE-Q Skin Cancer Module

(Continues)

TABLE 1 (Continued)

CFA used on data	Sample size (n) ^a	Response format	Type of data	Person versus item focus	Rationale for using IRT programs	ENT content	Sample location	Author type	Scale used
nr	332	Polytomous	Appearance, adverse effects, and QoL	Item focus	To evaluate the overall quality of the scale	Facial plastics	USA and Canada	Clinical	FACE-Q Appearance Appraisal Scales and Adverse Effects Checklist for the Lower Face and Neck
nr	158	Polytomous	Appearance, adverse events, and QoL	Item focus	To determine whether the data collected from a sample fit the mathematical model	Facial plastics	USA, UK, and Canada	Clinical	FACE-Q Scale for Rhinoplasty
nr	702	Polytomous	Appearance, QoL, and adverse effect	Item focus	To identify a set of items that represented the best indicators of outcome for each of the five scales	Facial plastics	USA and Canada	Clinical	FACE-Q Scales
nr	369	Polytomous	Appearance, QoL	Item focus	To analyze psychometric properties	Facial plastics	USA, Canada, and UK	Clinical	Expectation and Psychological Distress
nr	960	Polytomous	Appearance and adverse effects	Item focus	To examine psychometric properties	Facial plastic	China, Canada, UK, USA, Spain, Brazil, Ireland, and Australia	Clinical	Ear-Q
nr	235	Polytomous	Appearance, QoL, and performance	Item focus	To evaluate psychometric properties	Facial plastics	Canada, UK, USA, Netherlands, and France	Clinical	FACE-Q Scale for facial nerve paralysis
nr	235	Polytomous	Appearance, function, QoL, adverse effects	Item focus	To evaluate the psychometric performance of the FACE-Q scales	Facial plastics	Canada	Clinical	Face-Q Craniofacial Module for Facial Nerve Paralysis
nr	239	Polytomous	Appearance, QoL	Item focus	To evaluate the construct validity of the anglicized FACE-Q Skin Cancer module	H&N	UK	Clinical	FACE-Q skin cancer module
nr	613	Polytomous	Performance	Item focus	To examine individual misophonic triggers	Otology/Neurotology/Audiology	UK	Clinical	Individual misophonia auditory triggers
Yes	1173	Polytomous	QoL	Item focus	To select the minimum number of the best-discriminating items	H&N	The Netherlands	Clinical	Basal and Squamous Cell Carcinoma Quality of Life (BaSQoL) Questionnaire

TABLE 1 (Continued)

CFA used on data	Sample size (n) ^a	Response format	Type of data	Person versus item focus	Rationale for using IRT programs	ENT content	Sample location	Author type	Scale used
nr	530	Polytomous	QoL	Item focus	To provide estimates of VHI item "difficulties" that can be used to transform ordinal VHI scores of patients into interval level severity measures and to re-examine the VHI-questionnaire to obtain more definite results regarding its dimension structure	Otology/ Neurology/ Audiology	The Netherlands	Clinical	Voice Handicap Index
nr	66	Polytomous	Performance	Item focus	To evaluate content validity	Pediatric otolaryngology	UK	Clinical	Speech, Spatial, and Qualities of Hearing Scale for Parents (SSQ-P)
Yes	205	Polytomous	QoL	Item focus	To assess psychometric properties through robust statistical techniques	Rhinology	UK and Ireland	Clinical	Food Allergy Anxiety Scale
Yes	800	Polytomous	QoL	Item focus	To analyze psychometric properties	Rhinology	USA	Clinical	The Sinonasal Outcome Test-22 (SNOT-22)
Yes	800	Polytomous	QoL	Item focus	To evaluate refinement possibilities	Rhinology	USA	Clinical	The Sinonasal Outcome Test-22 (SNOT-22)
nr	86	Polytomous	QoL	Item focus	To explore the relationship between TWVQ items and individuals' responses	Laryngology	USA	Clinical	Trans Woman Voice Questionnaire (TWVQ)
nr	117	Polytomous	Performance	Person focus	To determine whether patient features such as dizziness or fall history influence the measurement characteristics of the DGI	Otology/ Neurology/ Audiology	USA	Clinical	Dynamic Gait Index
nr	636	Polytomous	Performance	Item focus	To evaluate the reliability, validity, and identification accuracy	Laryngology	Spain, Turkey, Sweden, and Italy	Clinical	Eating Assessment Tool (EAT-10)
nr	114	Polytomous	Performance	Item focus	To test the construct validity	Otology/ Neurology/ Audiology	Australia	Clinical	Vestibular Screening Tool
Yes	316	Polytomous	QoL	Item focus	To evaluate psychometric properties	Rhinology	Denmark	Clinical	Taste and Smell Tool
nr	341	Polytomous	Performance	Item focus	To evaluate internal structure	Otology/ Neurology/ Audiology	China	Clinical	Cantonese Tone Identification Test
nr	108	Polytomous	QoL	Item focus	To examine the psychometric properties of the Dutch translation Swallowing QoL Questionnaire and its adjusted version	Laryngology	Belgium	Clinical	Dutch translation Swallowing QoL Questionnaire and its adjusted version

(Continues)

TABLE 1 (Continued)

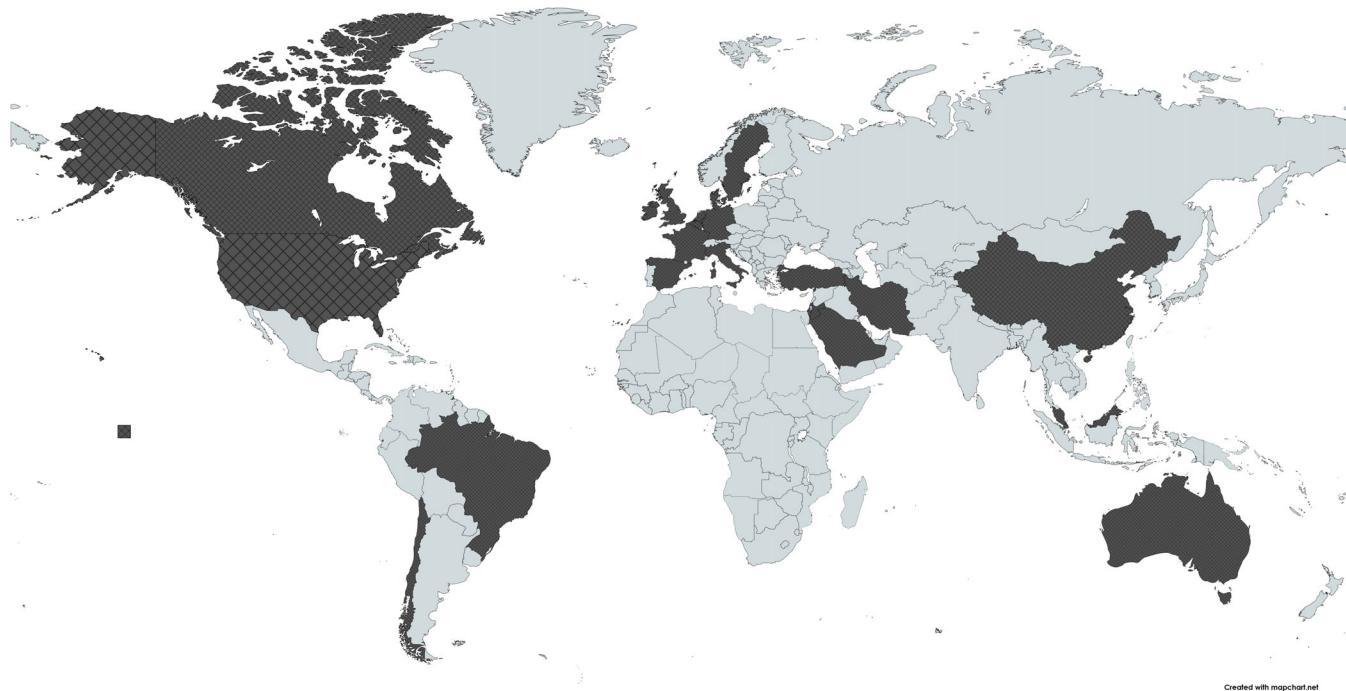
CFA used on data	Sample size (n) ^a	Response format	Type of data	Person versus item focus	Rationale for using IRT programs	ENT content	Sample location	Author type	Scale used
Yes	212	Polytomous	Attitude	Item focus	To assess scales within the HAROQ to determine their usefulness in a hearing screen setting	Otology/ Neurology/ Audiology	The Netherlands	Clinical	Hearing Aid Rehabilitation Questionnaire
Yes	1730	Polytomous	Performance	Item focus	To modify and verify the psychometric properties	Otology/ Neurology/ Audiology	China	Clinical	Infant-Toddler Meaningful Auditory Integration Scale
nr	353	Polytomous	Ability and QoL	Item focus	To determine whether the underlying latent ability could also be accurately represented by a subset of the items	Otology/ Neurology/ Audiology	USA	Clinical	Inner Effectiveness of Auditory Rehabilitation (Inner Ear) Scale
nr	1035	Dichotomous	Performance	Person focused	To investigate if there is an ordering in the speech sound development	Pediatric otalaryngology	The Netherlands	Clinical	Logo-Articulation Assessment
nr	453	Polytomous	Performance	Item focus	To evaluate objectivity and cross-cultural validity	Otology/ Neurology/ Audiology	USA, Germany, Jordan, and Saudi Arabia	Clinical	Vestibular Activities and Participation questionnaire (VAP)
Yes	115	Polytomous	Performance	Item focus	To improve the rating categories and delete items misfitting or showing local dependency	Otology/ Neurology/ Audiology	Italy	Clinical	Balance Evaluation System's Test: the mini-BESTest
nr	330	Polytomous	Effort	Item focus	To improve measurement precision and to assess psychometric measurement properties	Otology/ Neurology/ Audiology	UK	Clinical	Listening Effort Questionnaire-Cochlear Implant (LEQ-CI)
nr	380	Polytomous	QoL	Item focus	To assess the measurement properties of the HHIE	Otology/ Neurology/ Audiology	UK	Clinical	Hearing Handicap Inventory for the Elderly (HHIE)
nr	155	Polytomous	QoL	Item focus	To test the scalability of items and to test whether or not patients use the items response scale correctly	Rhinology	Italy	Clinical	RhinAsthma patient perspective
Yes	512	Polytomous	Appearance, QoL, adverse events, process of care	Item focus	To complement CTT and to further evaluate the scaling properties and construct validity	H&N	Canada	Clinical	McGill Body Image Concerns Scale for Use in Head and Neck Oncology (MBIS-HNC)
nr	100	Polytomous	Performance	Item focus	To evaluate cross-cultural equivalence	Rhinology	Belgium	Clinical	Dutch Face-Q Rhinoplasty Questionnaire
Yes	129	Polytomous	QoL	Item focus	To perform psychometric analysis	Otology/ Neurology/ Audiology	USA	Clinical	The Cochlear Implant Quality of Life-Expectations Instrument

TABLE 1 (Continued)

CFA used on data	Sample size (n) ^a	Response format	Type of data	Person versus item focus	Rationale for using IRT programs	ENT content	Sample location	Author type	Scale used
Yes	301	Polytomous	QoL and personality	Item focus	To confirm domain factor structure and to analyze items	H&N	USA	Clinical	Inventory to Measure and Assess image disturbance—Head & Neck (IMAGE-HN)
Yes	371	Polytomous	QoL	Item focus	To create an item bank and to assess CIQOL in the adult CI population	Otology/Neurology/Audiology	USA	Clinical	Cochlear Implant Quality of Life (CIQOL) Profile Instrument (CIQOL-35 Profile) and a Global Measure (CIQOL-10 Global)
Yes	371	Polytomous	QoL	Item focus	To determine psychometric properties of the CIQOL-35 Profile and the CIQOL-10 Global individual item difficulty, discrimination, and model fit to select the set of items for the profile instrument and global measure that would optimize their measurement characteristics	Otology/Neurology/Audiology	USA	Clinical	Cochlear Implant Quality of Life (CIQOL Profile Instrument) (CIQOL-35 Profile) and a Global Measure (CIQOL-10 Global)

Abbreviations: CFA, confirmatory factors analysis; CI, cochlear implant; CIQOL, Cochlear Implant Quality of Life; CTT, Classical Test Theory; GRM, graded response model; IRT, item response theory; PRO, patient-reported outcomes; QoL, quality of life; nr, not reported.

^aWhen more than two groups are included, the group with the highest number of participants is reported.



Created with mapchart.net

FIGURE 2 World map highlighting countries (dark gray area) from which study participants were recruited for IRT analyses ([Mapchart.net](#)). IRT, item response theory.

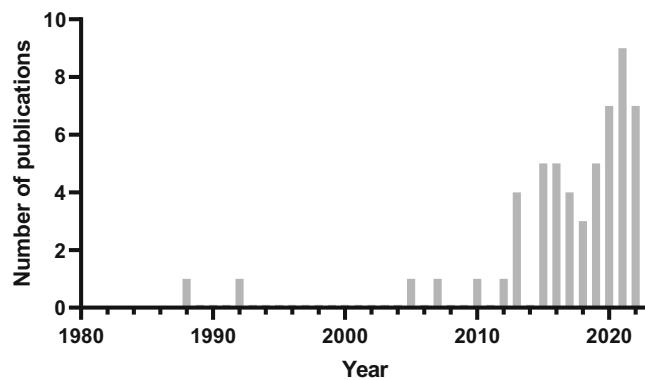


FIGURE 3 Number of studies using IRT in otolaryngology literature sorted by year. IRT, item response theory.

17 studies (52.7%), followed by the Winsteps®²¹ software, used in 12 studies (21.8%). R®^{22,23} was used in nine studies (16.4%), followed by four studies that used STATA®²⁴ (7.3%).

One study used both SAS®²⁵ and M-PLUS®,²⁶ whereas another one used a combination of SAS® and R®.^{22,23} Another study used both the BICAL®²⁷ and DIGRAM® software.²⁸ ACER ConQuest®,²⁹ IRTPRO®,³⁰ Jmetric®,³¹ OPLM®,³² and Python®³³ were each used once (1.8% each).

3.7 | Response format and data types analyzed

Most studies analyzed polytomous data ($n = 49$, 89.1%), whereas only six studies (10.9%) evaluated dichotomous data. Regarding data type,

39 studies analyzed one data type (70.9%), whereas 16 studies (29.1%) analyzed more than one data type. The most common data type in those studies that only used one data type was: quality of life ($n = 19$, 34.6%), followed by performance data (i.e., speech perception skills, balance assessment, functional gait abilities) analyzed in 17 studies (30.9%), attitude data in 2 studies (5.1%), and data related to effort in one study (2.6%). Regarding study design, 22 studies (40%) used a mixed-methods approach, whereas 33 studies (60%) were conducted based on a quantitative study design.

3.8 | Basic IRT assumptions

Studies using traditional IRT methods follow basic assumptions, such as (i) unidimensionality of the measured trait/outcome (i.e., that the items only measure one dominant latent trait) and (ii) local independence of analyzed items (i.e., responses from participants are not statistically related to each other, while controlling for the latent trait). Previous studies recommended evaluating those basic assumptions before using IRT-based methods.^{11–15,18}

In our study, we found that 30 (54.5%) reported the evaluation of unidimensionality, whereas 29 (52.7%) reported assessing checks on local independence. Eighteen studies (32.7%) reported assessing both unidimensionality and local independence. Regarding other IRT-related variables, 47 studies (85.5%) reported goodness-of-fit statistics, 15 studies (27.3%) also incorporated CFA during their analysis, and 6 studies (10.9%) reported assessing multiple IRT models for the best-fitting model.

TABLE 2 IRT models used in otolaryngological literature.

	Number of studies	Sample size (range)	Mean/median	SD	IQR
Partial credit model	11	86–960	379.5/330	268.8	357
Rating scale model	10	100–636	246.4/162.5	173	254.7
Graded response model	9	66–1730	507.8/223.0	530.4	607
Two-parameter logistic model	4	302–1173	656.3/575.0	369.1	672.2
Bayesian IRT model	1	13,273	–	–	–
Mokken analysis	1	1035	–	–	–
Dichotomous logistic model	1	530	–	–	–

Abbreviations: IQR, interquartile range; SD, standard deviation.

4 | DISCUSSION

Patients' perspectives on their own health and health care have become increasingly important in recent years, leading to the widespread use of PROMs in clinical practice.¹ However, the quality of these PROs can vary greatly, with many lacking strong psychometric properties.⁴ To address this issue, new test methods such as IRT and RA have emerged, offering significant advantages over traditional CTT methods for evaluating psychometric properties.¹⁸ In otolaryngology, the use of IRT and RA has become particularly important, providing valuable insights into research results. To date, however, no study has comprehensively evaluated the use of IRT methodology in the otolaryngology scientific literature. In this study, we have synthesized original contributions in otolaryngology literature that used modern test theory for statistical analyses. We discovered that IRT methodology had been used globally and across all subspecialties of ENT, with a significant increase in the last two decades. Moreover, we found that the most common data type analyzed using IRT methodology was QoL-related and that the most common item response was polytomous. Furthermore, we found that commercial software such as RUMM2030[®] or Winsteps[®], as well as open-source software such as R[®], has made implementation of modern test theory more accessible to researchers. Although these results will hopefully serve to motivate and direct future use of IRT in otolaryngology, we note that there is still room for more consistent reporting of IRT-model assumptions.

In the 1970s, IRT became prominent for its use in education, creating standardized tests.³⁴ Since then, it has become indispensable for developing, evaluating, and scoring psychology tests. Not surprisingly, IRT has become a cornerstone in the development of clinical assessment tools as healthcare providers prioritize patient-centered care, especially as the need for accurate and reliable data from PROMs has become increasingly important.¹⁸ Considering the widespread use of IRT methodology in the biomedical sciences it is essential to better understand their utilization in the otolaryngology literature. Gathering information on how these new testing methods are being used might provide valuable insights into interpreting research results. This, in turn, may inform future studies and motivate use of IRT methods in otolaryngology research.

There are various software programs available for conducting IRT analyses, including commercial and dedicated software packages such

as Winsteps[®], RUMM 2030[®], M-PLUS[®], BICAL[®], and DIGRAM[®]. In addition, IRT can also be conducted using the nonspecific open-source packages of the programming language R[®]. In the otolaryngology literature, we found that RUMM2030[®] and Winsteps[®] were the most commonly used software for IRT analyses. One possible explanation for this could be the user-friendly and dedicated interface of both RUMM2030[®] and Winsteps[®], which make them easier to use compared to the nonspecific interfaces of R[®] or STATA[®].

Researchers, and test- and PROM-developers widely use IRT and Rasch models to explore the connection between an individual's ability level and their responses to specific items on a given scale.^{11–16} These models offer valuable insights into human behavior and performance by analyzing the likelihood of certain responses based on latent traits, providing significant benefits for unlocking new insights and achieving better outcomes. Depending on several factors, such as the response format (i.e., dichotomous or polytomous answers), various IRT models can be used. Additionally, it is important to evaluate whether the discrimination parameter can be consistently applied across all items and whether guessing is a potential variable. Furthermore, it is necessary to determine if different categorical response parameters must be estimated for each item on the scale. A thorough examination of these aspects is imperative before making an informed decision. For dichotomous item responses, the simplest model is the 1-parameter logistic (1PL), followed by the 2-parameter logistic (2PL) and 3-parameter logistic (3PL) models. Rasch model and the 1PL are often used interchangeably, as the results from both models are nearly identical. However, the algorithm is slightly different (for an overview on this topic, see reference 35). Other IRT models exist for polytomous responses, such as the GRM,³⁶ partial credit, or rating scale models. In this study, we found that the partial credit, the rating scale, and the GRM were the most popular among otolaryngology studies using IRT methodology. This was not surprising, as QoL measures are usually assessed based on polytomous answer options. Our findings underscore the importance of understanding how to interpret outcome data from different models.

After selecting the appropriate IRT model for the data analysis, testing for goodness of fit is essential. Although there is currently no general agreement on assessing model-data fit, various methods have been used, such as evaluating a model's relative fit based on information criteria or comparing different models.¹⁸ In this review, we found

that most studies provided model-data fit information. It is crucial to note that, as with other statistical tests, IRT methodology also has certain assumptions that need to be taken into consideration and reported during the analysis: (i) local independence (i.e., items should not be related to each other except for measuring the same trait), (ii) unidimensionality (i.e., all items are measuring the same trait), (iii) monotonicity (i.e., as the trait level increases, the likelihood of providing a correct response also increases), and (iv) invariance (i.e., item parameters are independent of the population measured). In our analysis, we found that more than half of the included studies reported testing for unidimensionality and local independence. However, only 19/55 articles reported testing both assumptions. Our findings are encouraging in that within otolaryngology literature using IRT-based methodology, most studies checked and reported model-data fit. Interestingly, whereas many studies checked for at least one underlying assumption of IRT models, only a few conducted tests for more than one assumption. We believe that our results reinforce the importance of testing for the main assumptions and reporting the results to ensure an IRT analysis's accuracy and validity.

One of the main drawbacks of CTT is that it assumes that reliability is constant across the measurement scale and assesses reliability as the sum of items. However, this is not always the case, as reliability can vary significantly between individual items. On the other hand, the IRT methodology is based on items rather than summary scores, providing item-level information, which is especially useful during the development and refinement of PROMs.

It is interesting to note that IRT methodology was first introduced in 1950. Still, it was not until 1988 that it was reported in otolaryngology literature.³⁷ In terms of its use across different subspecialties within ENT, we found that IRT methodology was most frequently used in the literature related to otology and neurotology, followed by rhinology and facial plastics, laryngology, pediatric otolaryngology, and oncology. It was interesting that IRT was used least commonly in head and neck surgery, although several explanations could be posited. For example, head and neck surgical oncology outcomes may rely most prominently on objective measurements of cancer size, spread, recurrence and mortality. However, with continued improvements in survival from head and neck cancer and increasing focus now on QoL, there may be greater opportunities for the utilization of IRT in the subspecialty field of head and neck surgical oncology.

It is worth noting that in the review conducted, nearly half of the included studies analyzed QoL-related data, followed by performance and appearance data. This finding is particularly important as it highlights the increasing scientific interest in measuring health outcomes in otolaryngology from the patient's perspective. PROs allow us to track patients' symptoms and monitor how they evolve, providing valuable insights into the effectiveness of the healthcare system and the extent to which it delivers meaningful improvements in patients' lives. This trend represents a positive development that has the potential to benefit both patients and healthcare providers significantly.

5 | CONCLUSION

In this review, we have discussed and explained the different IRT methodologies. We have shown that IRT has been used across all subspecialties of otolaryngology to develop and improve existing PROMs. We believe that the results of our study may serve to motivate and direct usage of IRT methods in future studies. In turn, future IRT-based outcome studies, using standardized reporting guidelines, can improve otolaryngology-outcome research sustainably by improving response rates and reducing patient's response burden.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

David T. Liu  <https://orcid.org/0000-0001-6948-737X>

Christian A. Mueller  <https://orcid.org/0000-0001-6172-5417>

Ahmad R. Sedaghat  <https://orcid.org/0000-0001-6331-2325>

REFERENCES

- Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas*. 2018;9:353. doi:[10.2147/PROM.S156279](https://doi.org/10.2147/PROM.S156279)
- Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evid Based Med*. 2022;27(3):153. doi:[10.1136/BMJEBO-2020-111573](https://doi.org/10.1136/BMJEBO-2020-111573)
- Hansen CF, Jensen J, Brodersen J, Siersma V, Comins JD, Krosgaard MR. Are adequate PROMs used as outcomes in randomized controlled trials? An analysis of 54 trials. *Scand J Med Sci Sports*. 2021;31(5):972-981. doi:[10.1111/sms.13896](https://doi.org/10.1111/sms.13896)
- Gagnier JJ, Johnston BC. Poor quality patient reported outcome measures bias effect estimates in orthopedic randomized studies. *J Clin Epidemiol*. 2019;116:36-38. doi:[10.1016/J.JCLINEPI.2019.07.012](https://doi.org/10.1016/J.JCLINEPI.2019.07.012)
- Chang EM, Gillespie EF, Shaverdian N. Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality. *Patient Relat Outcome Meas*. 2019;10:171. doi:[10.2147/PROM.S178344](https://doi.org/10.2147/PROM.S178344)
- Rothrock NE, Kaiser KA, Celli D. Developing a valid patient-reported outcome measure. *Clin Pharmacol Ther*. 2011;90(5):737-742. doi:[10.1038/CLPT.2011.195](https://doi.org/10.1038/CLPT.2011.195)
- Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. *Clin Ther*. 2014;36(5):648. doi:[10.1016/J.CLINTHERA.2014.04.006](https://doi.org/10.1016/J.CLINTHERA.2014.04.006)
- Miller MD. Classical test theory reliability. *International Encyclopedia of Education*. 3rd ed. Elsevier; 2009:27-30. doi:[10.1016/B978-0-08-044894-7.00235-9](https://doi.org/10.1016/B978-0-08-044894-7.00235-9)
- De Ayala RJ. (Rafael J) *The Theory and Practice of Item Response Theory*. Guilford Press; 1957.
- Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health*. 2015;18(1):25-34. doi:[10.1016/J.JVAL.2014.10.005](https://doi.org/10.1016/J.JVAL.2014.10.005)
- Embreton SE. The new rules of measurement. *Psychol Assess*. 1996;8(4):341-349. doi:[10.1037/1040-3590.8.4.341](https://doi.org/10.1037/1040-3590.8.4.341)
- Reeve B. *An Introduction to Modern Measurement Theory*; National Cancer Institute 2005.

13. 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. doi:[10.1007/s11136-007-9198-0](https://doi.org/10.1007/s11136-007-9198-0)
14. Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educ Meas Issues Pract.* 1993;12(3):38-47.
15. Hambleton RK, Swaminathan H, Rogers WH. *Fundamentals of Item Response Theory*. Sage Publications; 1991.
16. Baker FB. *The Basics of Item Response Theory*. 2nd ed.; 2001.
17. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Paedagogiske Institut; 1960 Accessed May 31, 2023. https://books.google.at/books/about/Probabilistic_Models_for_Some_Intelligence.html?id=aB9qlGcEACAAJ&redir_esc=y
18. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient.* 2014;7(1):23-35. doi:[10.1007/s40271-013-0041-0](https://doi.org/10.1007/s40271-013-0041-0)
19. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34. doi:[10.1016/j.jclinepi.2009.06.006](https://doi.org/10.1016/j.jclinepi.2009.06.006)
20. RUMM Laboratory Pty Ltd. RUMM2030. RUMM Laboratory Pty Ltd; 2012.
21. Linacre J. *Winsteps version 3.80.0*. Winsteps.com; 2013.
22. Rizopoulos D. Irtm: An R package for latent variable modeling and item response analysis. *J Stat Softw.* 2007;17(5):1-25. doi:[10.18637/JSS.V017.I05](https://doi.org/10.18637/JSS.V017.I05)
23. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2022 <https://www.r-project.org/>
24. StataCorp. *Stata Statistical Software Release 14*. StataCorp LP; 2015 Accessed May 31, 2023. [https://www.scirp.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/referencespapers.aspx?referenceid=1947409](https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/referencespapers.aspx?referenceid=1947409)
25. SAS Institute. *The SAS System for Windows* SAS Institute Inc. 1999.
26. Muthén LK, Muthén BO. *Statistical Analysis With Latent Variables: User's Guide*. Muthén & Muthén; 1998 Accessed May 31, 2023. www.StatModel.com
27. Wright BD, Mead RJ, Bell SR. *BICAL: Calibrating Items with the Rasch Model*. Statistical Laboratory, Department of Education, The University of Chicago; 1980.
28. Kreiner S, Nielsen T. *Item Analysis in DIGRAM 3.04. Part I: Guided Tours. Research Report 13/6*. 2013 Accessed June 1, 2023. https://www.researchgate.net/publication/316215618_Item_analysis_in_DIGRAM_304_Part_I_Guided_tours
29. Adams RJ, Wu ML, Cloney D, Bereznner A, Wilson MR. ACER ConQuest: Generalized Item Response Modeling Software (Version 5.29). 2020 <https://www.acer.org/au/conquest>
30. Paek I, Han KT. IRTPRO 2.1 for Windows (Item Response Theory for Patient-Reported Outcomes). *Appl Psychol Meas.* 2012;37(3):242-252. doi:[10.1177/0146621612468223](https://doi.org/10.1177/0146621612468223)
31. Aksu G, Güzeller CO, Eser MT. JMTRIK: Classical test theory and item response theory data analysis software. *J Meas Eval Educ Psychol.* 2019;10(2):165-178. doi:[10.21031/EPOD.483396](https://doi.org/10.21031/EPOD.483396)
32. Verhelst N, Glas C, Verstraten H. *OPLM: Computer Manual and Program Cito*. 1995.
33. van Rossum G. *Python Reference Manual*. CWI (Centre for Mathematics and Computer Science); 1995.
34. Bock RD. A brief history of item theory response. *Educ Meas Issues Pract.* 1997;16(4):21-33. doi:[10.1111/j.1745-3992.1997.tb00605.x](https://doi.org/10.1111/j.1745-3992.1997.tb00605.x)
35. Stemler SE, Naples A. Rasch measurement v. item response theory: knowing when to cross the line. *Pract Assess Res Eval.* 2021;26:2021. doi:[10.7275/v2gd-4441](https://doi.org/10.7275/v2gd-4441)
36. Samejima F. Graded response model. In: van der Linden WJ, Hambleton RK, eds. *Handbook of Modern Item Response Theory*. Springer; 1997:85-100. doi:[10.1007/978-1-4757-2691-6_5](https://doi.org/10.1007/978-1-4757-2691-6_5)
37. Bement L, Wallber J, DeFilippo C, Bochner J, Garrison W. A new protocol for assessing viseme perception in sentence context: the lip-reading discrimination test. *Ear Hear.* 1988;9(1):33-40. doi:[10.1097/00003446-198802000-00014](https://doi.org/10.1097/00003446-198802000-00014)
38. Bochner J, Garrison W, Palmer L. Simple discrimination isn't really simple. A confirmatory analysis of the Speech Sound Pattern Discrimination Test. *Scand Audiol.* 1992;21(1):37-49. doi:[10.3109/0150399209045980](https://doi.org/10.3109/0150399209045980)
39. Klassen AF, Rae C, Riff W, et al. FACE-Q craniofacial module: Part 2 Psychometric properties of newly developed scales for children and young adults with facial conditions. *J Plast Reconstr Aesthet Surg.* 2021;74(9):2330-2340. doi:[10.1016/j.bjps.2021.03.009](https://doi.org/10.1016/j.bjps.2021.03.009)
40. La Padula S, Coiante E, Pizza C, et al. The Face- and Neck-Lift Objective Photo-Numerical Assessment Scale: a complete scale for face-lift evaluation. *Plast Reconstr Surg.* 2023;151(1):64-71. doi:[10.1097/PRS.00000000000009797](https://doi.org/10.1097/PRS.00000000000009797)
41. Lauritsen MBG, Kreiner S, Söderström M, Dørup J, Lous J. A speech reception in noise test for preschool children (the Galter-test): validity, reliability and acceptance. *Int J Pediatr Otorhinolaryngol.* 2015; 79(10):1694-1701. doi:[10.1016/j.ijporl.2015.07.028](https://doi.org/10.1016/j.ijporl.2015.07.028)
42. Pusic AL, Klassen AF, Scott AM, Cano SJ. Development and psychometric evaluation of the FACE-Q satisfaction with appearance scale. A new patient-reported outcome instrument for facial aesthetics patients. *Clin Plast Surg.* 2013;40(2):249-260. doi:[10.1016/j.cps.2012.12.001](https://doi.org/10.1016/j.cps.2012.12.001)
43. Leong KP, Yeak SCL, Saurajen ASM, et al. Why generic and disease-specific quality-of-life instruments should be used together for the evaluation of patients with persistent allergic rhinitis. *Clin Exp Allergy.* 2005;35(3):288-298. doi:[10.1111/j.1365-2222.2005.02201.X](https://doi.org/10.1111/j.1365-2222.2005.02201.X)
44. Abdullah B, Kandiah R, Hassan NFHN, Ismail AF, Mohammad ZW, Wang DY. Assessment of perception, attitude, and practice of primary care practitioners towards allergic rhinitis practice guidelines: development and validation of a new questionnaire. *World Allergy Organ J.* 2020;13(12):100482. doi:[10.1016/j.waojou.2020.100482](https://doi.org/10.1016/j.waojou.2020.100482)
45. Sucharew H, Khouri JC, Rao M, et al. Predicting allergic disease at age four using an atopy predisposition score at age two: the application of item response theory. *Pediatr Allergy Immunol.* 2012;23(2): 195-201. doi:[10.1111/j.1399-3038.2011.01251.x](https://doi.org/10.1111/j.1399-3038.2011.01251.x)
46. Leijon A, Dillon H, Hickson L, Kinkel M, Kramer SE, Nordqvist P. Analysis of data from the International Outcome Inventory for Hearing Aids (IOI-HA) using Bayesian Item Response Theory. *Int J Audiol.* 2021;60(2):81-88. doi:[10.1080/14992027.2020.1813338](https://doi.org/10.1080/14992027.2020.1813338)
47. Liu DT, Phillips KM, Houssein FA, et al. Dedicated olfaction and taste items do not improve psychometric performance of the SNOT-22. *Laryngoscope.* 2022;132(8):1644-1651. doi:[10.1002/lary.30120](https://doi.org/10.1002/lary.30120)
48. Crump RT, Liu G, Janjua A, Sutherland JM. Analyzing the 22-item Sino-Nasal Outcome Test using item response theory. *Int Forum Allergy Rhinol.* 2016;6(9):914-920. doi:[10.1002/alr.21770](https://doi.org/10.1002/alr.21770)
49. Apon I, van Leeuwen N, Allori AC, et al. Rasch analysis of patient- and parent-reported outcome measures in the International Consortium for Health Outcomes Measurement Standard Set for Cleft Lip and Palate. *Value Health.* 2021;24(3):404-412. doi:[10.1016/j.jval.2020.10.019](https://doi.org/10.1016/j.jval.2020.10.019)
50. Dehqan A, Yadegari F, Asgari A, Scherer RC, Dabirmoghadam P. Development and validation of an Iranian Voice Quality of Life Profile (IVQLP) based on a classic and Rasch Rating Scale Model (RSM). *J Voice.* 2017; 31(1):113.e19-113.e29. doi:[10.1016/j.jvoice.2016.03.018](https://doi.org/10.1016/j.jvoice.2016.03.018)
51. Atallah MR, Milad D, Benamer YH, Saltychev M, Most SP, Moubayed SP. Translation, cultural adaptation and validation of the SCHNOS in French. *J Otolaryngol Head Neck Surg.* 2019;48(1):17. doi:[10.1186/s40463-019-0339-6](https://doi.org/10.1186/s40463-019-0339-6)
52. Wulff NB, Møller PR, Christensen KB, et al. The voice-related quality of life (V-RQOL) instrument: cross-cultural translation and test of validity and reliability of the Danish version. *J Voice.* 2021;35(5):806-806.e14. doi:[10.1016/j.jvoice.2020.01.010](https://doi.org/10.1016/j.jvoice.2020.01.010)

53. Klassen AF, Cano SJ, Schwitzer JA, et al. Development and psychometric validation of the FACE-Q skin, lips, and facial rhytids appearance scales and adverse effects checklists for cosmetic procedures. *JAMA Dermatol.* 2016;152(4):443-451. doi:[10.1001/jamadermatol.2016.0018](https://doi.org/10.1001/jamadermatol.2016.0018)
54. Lee EH, Klassen AF, Cano SJ, Nehal KS, Pusic AL. FACE-Q Skin Cancer Module for measuring patient-reported outcomes following facial skin cancer surgery. *Br J Dermatol.* 2018;179(1):88-94. doi:[10.1111/bjd.16671](https://doi.org/10.1111/bjd.16671)
55. Klassen AF, Cano SJ, Scott AM, Pusic AL. Measuring outcomes that matter to face-lift patients: development and validation of FACE-Q appearance appraisal scales and adverse effects checklist for the lower face and neck. *Plast Reconstr Surg.* 2014;133(1):21-30. doi:[10.1097/PRS.0000000000000895](https://doi.org/10.1097/PRS.0000000000000895)
56. Klassen AF, Cano SJ, East CA, et al. Development and psychometric evaluation of the FACE-Q scales for patients undergoing rhinoplasty. *JAMA Facial Plast Surg.* 2016;18(1):27-35. doi:[10.1001/jamafacial.2015.1445](https://doi.org/10.1001/jamafacial.2015.1445)
57. Klassen AF, Cano SJ, Schwitzer JA, Scott AM, Pusic AL. FACE-Q scales for health-related quality of life, early life impact, satisfaction with outcomes, and decision to have treatment: development and validation. *Plast Reconstr Surg.* 2015;135(2):375-386. doi:[10.1097/PRS.00000000000000895](https://doi.org/10.1097/PRS.00000000000000895)
58. Klassen AF, Cano SJ, Alderman A, et al. Self-report scales to measure expectations and appearance-related psychosocial distress in patients seeking cosmetic treatments. *Aesthet Surg J.* 2016;36(9):1068-1078. doi:[10.1093/asj/sjw078](https://doi.org/10.1093/asj/sjw078)
59. Klassen AF, Rae C, Bulstrode NW, et al. An international study to develop the EAR-Q patient-reported outcome measure for children and young adults with ear conditions. *J Plast Reconstr Aesthet Surg.* 2021;74(9):2341-2348. doi:[10.1016/j.bjps.2021.01.014](https://doi.org/10.1016/j.bjps.2021.01.014)
60. Klassen AF, Rae C, Gallo L, et al. Psychometric validation of the FACE-Q Craniofacial module for facial nerve paralysis. *Facial Plast Surg Aesthet Med.* 2022;24(1):1-7. doi:[10.1089/fpsam.2020.0575](https://doi.org/10.1089/fpsam.2020.0575)
61. Dobbs TD, Ottenhof M, Gibson JAG, Whitaker IS, Hutchings HA. The Patient-Reported Outcome Measures In Skin Cancer Reconstruction (PROMISCR) study: anglicisation and initial validation of the FACE-Q Skin Cancer Module in a UK cohort. *J Plast Reconstr Aesthet Surg.* 2021;74(3):615-624. doi:[10.1016/j.bjps.2020.09.005](https://doi.org/10.1016/j.bjps.2020.09.005)
62. Vitoratou S, Uglik-Marucha N, Hayes C, Erfanian M, Pearson O, Gregory J. Item response theory investigation of misophonia auditory triggers. *Audiol Res.* 2021;11(4):567-581. doi:[10.3390/audiolres11040051](https://doi.org/10.3390/audiolres11040051)
63. Waalboer-Spuij R, Hollestein LM, Timman R, Van De Poll-Franse LV, Nijsten TEC. Development and validation of the basal and squamous cell carcinoma quality of life (BaSQoL) questionnaire. *Acta Derm Venereol.* 2018;98(2):234-239. doi:[10.2340/00015555-2806](https://doi.org/10.2340/00015555-2806)
64. Bogaardt HCA, Hakkesteeg MM, Grolman W, Lindeboom R. Validation of the voice handicap index using Rasch analysis. *J Voice.* 2007; 21(3):337-344. doi:[10.1016/j.jvoice.2005.09.007](https://doi.org/10.1016/j.jvoice.2005.09.007)
65. Killan CF, Baxter PD, Killan EC. Face and content validity analysis of the Speech, Spatial and Qualities of Hearing Scale for Parents (SSQ-P) when used in a clinical service without interviews or week-long observation periods. *Int J Pediatr Otorhinolaryngol.* 2020;133:109964. doi:[10.1016/j.ijporl.2020.109964](https://doi.org/10.1016/j.ijporl.2020.109964)
66. Coelho GLDH, Byrne A, Hourihane J, DunnGalvin A. Development of the Food Allergy Anxiety Scale in an adult population: psychometric parameters and convergent validity. *J Allergy Clin Immunol Pract.* 2021;9(9):3452-3458.e1. doi:[10.1016/j.jacip.2021.04.019](https://doi.org/10.1016/j.jacip.2021.04.019)
67. Liu DT, Phillips KM, Speth MM, Besser G, Mueller CA, Sedaghat AR. Item response theory for psychometric properties of the SNOT-22 (22-item Sinonasal Outcome Test). *Otolaryngol Head Neck Surg.* 2022; 166(3):580-588. doi:[10.1177/01945998211018383](https://doi.org/10.1177/01945998211018383)
68. Liu DT, Phillips KM, Speth MM, Besser G, Mueller CA, Sedaghat AR. Exploring possibilities for shortening the 22-item Sino-Nasal Outcome Test (SNOT-22) using item response theory. *Int Forum Allergy Rhinol.* 2022;12(2):191-199. doi:[10.1002/alr.22878](https://doi.org/10.1002/alr.22878)
69. Zhao NW, Mason JM, Blum AM, et al. Using item-response theory to improve interpretation of the Trans Woman Voice Questionnaire. *Laryngoscope.* 2023;133(5):1197-1204. doi:[10.1002/lary.30360](https://doi.org/10.1002/lary.30360)
70. Dye DC, Eakman AM, Bolton KM. Assessing the validity of the dynamic gait index in a balance disorders clinic: an application of Rasch analysis. *Phys Ther.* 2013;93(6):809-818. doi:[10.2522/ptj.20120163](https://doi.org/10.2522/ptj.20120163)
71. Cordier R, Joosten A, Clavé P, et al. Evaluating the psychometric properties of the Eating Assessment Tool (EAT-10) using Rasch analysis. *Dysphagia.* 2017;32(2):250-260. doi:[10.1007/s00455-016-9754-2](https://doi.org/10.1007/s00455-016-9754-2)
72. Stewart V, Mendis MD, Rowland J, Choy NL. Construction and validation of the vestibular screening tool for use in the emergency department and acute hospital setting. *Arch Phys Med Rehabil.* 2015; 96(12):2153-2160. doi:[10.1016/j.apmr.2015.08.413](https://doi.org/10.1016/j.apmr.2015.08.413)
73. Niklasson AS, Christensen KB, Fjaeldstad AW, Ovesen T. Development and psychometric validation of the Taste And Smell Tool for Evaluation (TASTE) questionnaire. *JAMA Otolaryngol Head Neck Surg.* 2022;148(12):1164-1172. doi:[10.1001/jamaoto.2022.3392](https://doi.org/10.1001/jamaoto.2022.3392)
74. Lee KYS, Lam JHS, Chan KTY, van Hasselt CA, Tong MCF. Applying Rasch model analysis in the development of the Cantonese Tone Identification Test (CANTIT). *Int J Audiol.* 2017;56:S60-S73. doi:[10.1080/14992027.2017.1294766](https://doi.org/10.1080/14992027.2017.1294766)
75. Simpelaere IS, Van Nuffelen G, De Bodt M, Vanderwegen J, Hansen T. Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health Qual Life Outcomes.* 2017;15(1):66. doi:[10.1186/s12955-017-0639-3](https://doi.org/10.1186/s12955-017-0639-3)
76. Chenault M, Berger M, Kremer B, Anteunis L. Item response theory applied to factors affecting the patient journey towards hearing rehabilitation. *Audiol Res.* 2016;6(2):159. doi:[10.4081/audiores.2016.159](https://doi.org/10.4081/audiores.2016.159)
77. Yang F, Zhao F, Zheng Y, Li G. Modification and verification of the Infant-Toddler Meaningful Auditory Integration Scale: a psychometric analysis combining item response theory with classical test theory. *Health Qual Life Outcomes.* 2020;18(1):367. doi:[10.1186/s12955-020-01620-9](https://doi.org/10.1186/s12955-020-01620-9)
78. Jessen A, Ho AD, Corrales CE, Yueh B, Shin JJ. Improving measurement efficiency of the Inner EAR scale with item response theory. *Otolaryngol Head Neck Surg.* 2018;158(6):1093-1100. doi:[10.1177/0194599818760528](https://doi.org/10.1177/0194599818760528)
79. Priester GH, Post WJ, Goorhuis-Brouwer SM. Measuring speech sound development: an item response model approach. *Int J Pediatr Otorhinolaryngol.* 2013;77(9):1469-1473. doi:[10.1016/j.ijporl.2013.06.011](https://doi.org/10.1016/j.ijporl.2013.06.011)
80. Mueller M, Whitney SL, Alghwiri A, et al. Subscales of the Vestibular Activities and Participation questionnaire could be applied across cultures. *J Clin Epidemiol.* 2015;68(2):211-219. doi:[10.1016/j.jclinepi.2014.10.004](https://doi.org/10.1016/j.jclinepi.2014.10.004)
81. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTTest. *J Rehabil Med.* 2010;42(4):323-331. <https://www.embase.com/search/results?subaction=viewrecord&id=L358846383&from=export>
82. Hughes SE, Watkins A, Rapport F, Boisvert I, McMahon CM, Hutchings HA. Rasch analysis of the Listening Effort Questionnaire-Cochlear Implant. *Ear Hear.* 2021;42(6):1699-1711. doi:[10.1097/AUD.0000000000001059](https://doi.org/10.1097/AUD.0000000000001059)
83. Heffernan E, Weinstein BE, Ferguson MA. Application of Rasch analysis to the evaluation of the measurement properties of the Hearing Handicap Inventory for the Elderly. *Ear Hear.* 2020;41(5):1125-1134. doi:[10.1097/AUD.0000000000000832](https://doi.org/10.1097/AUD.0000000000000832)

84. Molinengo G, Baiardini I, Braido F, Loera B. RhinAsthma patient perspective: a Rasch validation study. *J Asthma*. 2018;55(2):119-123. doi:[10.1080/02770903.2017.1316391](https://doi.org/10.1080/02770903.2017.1316391)
85. Rodriguez AM, Frenkel S, Desroches J, et al. Development and validation of the McGill body image concerns scale for use in head and neck oncology (MBIS-HNC): a mixed-methods approach. *Psychooncology*. 2019;28(1):116-121. doi:[10.1002/pon.4918](https://doi.org/10.1002/pon.4918)
86. Declau F, Pingnet L, Verkest V, Hansen T. Cross-cultural evaluation of the Dutch FACE-Q rhinoplasty questionnaires using Rasch analysis. *Aesthet Surg J*. 2021;41(12):NP1916-NP1930. doi:[10.1093/asj/sjab217](https://doi.org/10.1093/asj/sjab217)
87. McRackan TR, Hand BN, Chidarala S, Dubno JR. Understanding patient expectations before implantation using the Cochlear Implant Quality of Life-Expectations instrument. *JAMA Otolaryngol Head Neck Surg*. 2022;148(9):870-878. doi:[10.1001/jamaoto.2022.2292](https://doi.org/10.1001/jamaoto.2022.2292)
88. Graboyes EM, Hand BN, Ellis MA, et al. Validation of a novel, multidomain head and neck cancer appearance- and function-distress patient-reported outcome measure. *Otolaryngol Head Neck Surg*. 2020;163(5):979-985. doi:[10.1177/0194599820927364](https://doi.org/10.1177/0194599820927364)
89. McRackan TR, Hand BN, Velozo CA, Dubno JR. Development of the cochlear implant quality of life item bank. *Ear Hear*. 2019;40(4):1016-1024. doi:[10.1097/AUD.0000000000000684](https://doi.org/10.1097/AUD.0000000000000684)
90. McRackan TR, Hand BN, Velozo CA, Dubno JR. Cochlear Implant Quality of Life (CIQOL): development of a profile instrument (CIQOL-35 Profile) and a global measure (CIQOL-10 Global). *J Speech Lang Hear Res*. 2019;62(9):3554-3563. doi:[10.1044/2019_JSLHR-H-19-0142](https://doi.org/10.1044/2019_JSLHR-H-19-0142)

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu DT, Mueller CA, Sedaghat AR. A scoping review of Rasch analysis and item response theory in otolaryngology: Implications and future possibilities. *Laryngoscope Investigative Otolaryngology*. 2024;9(1):e1208. doi:[10.1002/lia2.1208](https://doi.org/10.1002/lia2.1208)