


Development and Validation of the Jawline Subject Satisfaction Scale



Élaboration et validation de l'échelle de satisfaction des patients des patients sur l'aspect de leur mâchoire (Jawline Subject Satisfaction Scale)

Kaitlyn M. Enright, PhD, ACRP-CPI^{1,2} ,
John S. Sampalis, PhD, FACE, FAHA³, Anneke Andriessen, PhD^{4,5},
and Andreas Nikolis, MD, PhD, FRCSC^{1,6}

Abstract

Introduction: Research in aesthetic medicine commonly includes evaluations of subject satisfaction with treatment results. However, conventional analytic methods typically generate statistically imprecise ordinal scores. To overcome this limitation, researchers have begun employing the Rasch model, an analytical framework grounded in item response theory. The Rasch model permits scale modifications capable of enhancing measurement accuracy. This study focuses on using the Rasch model to evaluate a scale measuring subject satisfaction following aesthetic treatments to the jawline. **Objective:** To develop and validate a multiitem, self-administered questionnaire measuring patient satisfaction with aesthetic treatment of the jawline. **Methods:** A 10-item questionnaire [The Jawline Subject Satisfaction Scale (JS³)] was devised to measure subject satisfaction following aesthetic treatments of the jawline. Each question was responded to using a 5-point Likert scale, with response selections ranging from “very much satisfied” to “very much dissatisfied” or “strongly agree” to “strongly disagree.” The scale’s psychometric properties (reliability and separation for items and persons, item and person fit statistics, and unidimensionality and local independence) were validated using a Rasch model based on a dataset collected from a sample of forty subjects. **Results:** The results of the Rasch analysis revealed high internal consistency of the JS³, with a person reliability estimate of 0.86 and an item reliability estimate of 0.96. The separation estimates for persons and items were 2.50 and 4.72, respectively, demonstrating the scale’s ability to differentiate between high and low responders and validating the instrument’s construct. All infit and outfit values fell within the established range (0.5-1.5), and the data fit the model of unidimensionality and local independence. Raw score transformations into logits were conducted, which were then converted to Rasch measurements. These measurements are available for use in practice for conducting standard statistical analyses evaluating treatment and/or group effects. **Conclusions:** The application of the Rasch model produced a valid and reliable scale (ie, JS³) for measuring satisfaction with the appearance of the jawline following aesthetic treatments.

¹ Erevna Innovations Inc, Clinical Research Unit, Westmount, QC, Canada

² Klyncal Consulting & Services, Montreal, QC, Canada

³ JSS Medical Research, Saint-Laurent (Montreal), QC, Canada

⁴ Radboud University Medical Center, Nijmegen, the Netherlands

⁵ Andriessen Consultants, Malden, the Netherlands

⁶ Department of Plastic Surgery, McGill University, Montreal, QC, Canada

Received August 21, 2023. Revised December 8, 2023. Accepted March 12, 2024.

Corresponding Author:

Kaitlyn M. Enright, Erevna Innovations Inc., 376 Victoria Ave., Suite #400, Westmount, QC, Canada, H3Z 1C3.

Email: research@vicpark.com

Résumé

Introduction : La recherche en médecine esthétique inclut habituellement des évaluations sur la satisfaction des patients concernant les résultats du traitement. Toutefois, les méthodes d'analyse conventionnelles génèrent typiquement des cotes ordinales statistiquement imprécises. Pour contourner cette limitation, les chercheurs ont commencé à utiliser le modèle de Rasch, un cadre analytique ancré dans la théorie de la réponse d'item. Le modèle de Rasch permet des modifications d'échelle capables d'améliorer la précision des mesures. Cette étude porte sur l'utilisation du modèle de Rasch pour évaluer une échelle de mesure de la satisfaction des patients après des traitements esthétiques du contour de la mâchoire (ou « Jawline »). **Objectif :** Élaborer et valider un questionnaire autoadministré à plusieurs items de satisfaction des patients des patients sur le traitement esthétique du contour de leur mâchoire. **Méthodes :** Un questionnaire à dix items (le JS3 ou Jawline Subject Satisfaction Scale) a été mis au point pour mesurer la satisfaction des patients après des traitements esthétiques du contour de la mâchoire. La réponse à chaque question impliquait une échelle de Likert en cinq points avec des choix de réponse allant de « Très satisfait(e) » à « Très insatisfait(e) » ou de « Entièrement d'accord » à « Pas du tout d'accord ». Les propriétés psychométriques de l'échelle (fiabilité et statistiques sur l'adéquation de l'item et de la personne, caractère unidimensionnel et indépendance locale) ont été validées au moyen d'un modèle de Rasch s'appuyant sur un ensemble de données collectées auprès d'un échantillon de quarante sujets. **Résultats :** Les résultats de l'analyse de Rasch ont révélé la forte cohérence interne du JS3 avec une estimation de la fiabilité de la personne de 0,86 et une estimation de la fiabilité de l'item de 0,96. Les estimations de séparation pour les personnes et items étaient respectivement de 2,50 et 4,72 démontrant l'aptitude de l'échelle à faire la différence entre les répondeurs forts et faibles et validant le montage de l'instrument. Les valeurs Intit et Outfit sont tombées dans la plage établie (0,5 à 1,5) et les données correspondent au modèle d'unidimensionnalité et d'indépendance locale. Des transformations des scores bruts en logits ont été effectuées puis converties en modèle de Rasch, ces mesures sont utilisables en pratique pour réaliser des analyses statistiques standard évaluant les effets du traitement et/ou de groupe. **Conclusions :** L'application du Modèle de Rasch a produit une échelle valide et fiable (c'est-à-dire le JS3) pour la mesure de la satisfaction de l'aspect du contour de la mâchoire après traitements esthétiques.

Keywords

aesthetics, item response theory, Rasch model, patient-reported outcome measures, validation study

Mots clés

esthétique, modèle de Rasch, étude de validation, mesure des résultats rapportés par les patients, théorie de la réponse d'item

Introduction

Clinical researchers often utilize patient-reported outcome measures (PROMs) in the form of self-administered questionnaires comprised of qualitative or semiquantitative items that converge to quantitative scores or results that are intended to measure constructs such as quality of life, emotional well-being, and functional status. In the field of aesthetic medicine, such scales typically measure treatment satisfaction.

Scales used to measure patient satisfaction with treatment typically consist of a collection of questions or items with multiple choice responses that are ordinal or Likert scales. Levels of agreement (ie, strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree) are used to measure specific attributes of what the patient perceives to be satisfaction with treatment. Mathematical manipulation, including summation or averaging, is often used to summarize the “qualitative” responses to the questions into a quantitative score. Essentially, treating the ordinal categorical scale as a continuous one. While using these methods to measure change from baseline may have some validity, comparing the summative scores across patients may have serious flaws that can negatively affect the validity of the measurements.

The potential flaws of the approach described above, and specifically treating ordinal categorical responses as continuous scales to summarize the responses of questions on multiitem questionnaires, are related to the following nonfounded assumptions. First, all items of the questionnaire have the same relevance and importance for the construct being measured, in this case, patient satisfaction. As an example, satisfaction with the treatment process may have an entirely different importance and relevance than satisfaction with the final clinical outcome. Second, the spaces between different levels of the ordinal scale represent the same difference in the patient's actual satisfaction. As an example, in a 5-point Likert Scale ascertaining agreement with a specific statement, the difference between “strongly agree” and “agree” may not be the same numerically as the difference between “agree” and “neither agree nor disagree.” Nonetheless, the assumption of equal distances between these pivotal points is fundamental to the mathematical manipulation of the summation or averaging of the responses. Third, there is significant between- and within-patient as well as between- and within-item variation with respect to the actual value of the response that corresponds to the same score on the Likert Scale. As an example, refer to the illustration in Figure 1, wherein satisfaction response on

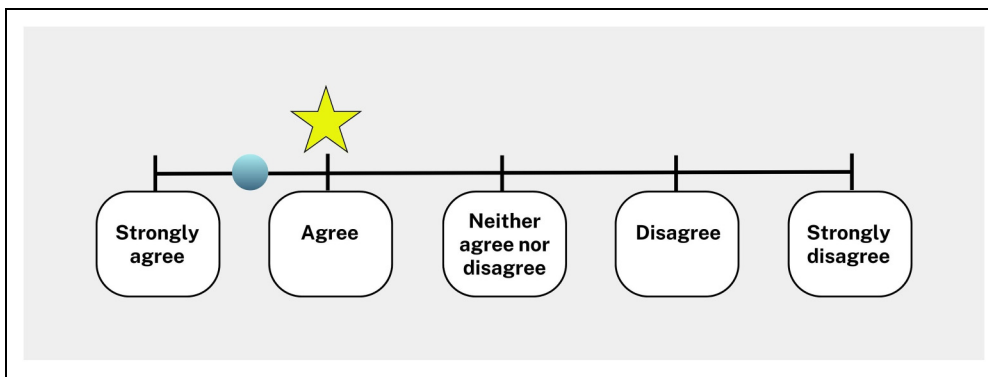


Figure 1. Assessing construct variability: discrepancy between actual values and scale responses in an illustrative context. In the illustration, the blue circle represents the actual value, and the yellow star represents the response on the scale.

the Likert scale may represent significantly different values as perceived by the respondent. In the illustration, the blue circle represents the actual value, and the yellow star represents the response on the scale. This creates significant variation with respect to the actual value of the construct within the same score on the scale.

There are some solutions to the problems with ordinal scales mentioned above such as using different weights for questionnaire items and employing visual analogue instead of ordinal scale for responses. However, these approaches have not been integrated into a comprehensive statistical model that resolves all the issues discussed above. Statistical techniques based on item response theory or, more specifically, the Rasch model, provide reproducible mathematical/statistical solutions to these issues that can be employed to develop multiitem questionnaires measuring a specific construct.¹⁻⁴ The Rasch model has become increasingly popular in scale development studies,⁵ with its application representing a paradigm shift in the analysis and interpretation of PROMs.² The Rasch model addresses weaknesses in traditional approaches as it is the only model based on fundamental measurement principles; a mathematical framework based on logical (not normative) axioms.⁶

The aim of this study was to develop a multiitem questionnaire measuring patient satisfaction with aesthetic treatment of the jawline using the Rasch approach.⁷

Objective

To develop and validate a multiitem self-administered questionnaire measuring patient satisfaction with aesthetic treatment of the jawline.

Methods

Scale Development

The Subject Satisfaction Scale (SSS) consists of 10 items (i), which are rated on a 5-point scale (i1: very much satisfied,

Table 1. Item Content of the Jawline Subject Satisfaction Scale (JS³).

Item	Content
1	Please rate your overall level of satisfaction with treatment results
2	The treatment improved my appearance
3	I am satisfied with the appearance of my jawline after treatment
4	I am satisfied with the appearance of my chin after treatment
5	I feel more attractive after treatment
6	I feel more comfortable being photographed or taking a “selfie” after treatment
7	The treatment results are natural looking
8	I look less tired after treatment
9	The treatment makes me feel better about myself
10	I am less self-conscious about the appearance of my lower face, after treatment

Note: The JS³ measures satisfaction with *treatment* and is therefore not performed at baseline. The JS³ is protected by copyright and can be obtained from the corresponding author under license.

satisfied, neutral, dissatisfied, and very much dissatisfied; i2-i10: strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree). Responses are assigned a corresponding numerical value (1-5), with higher scores reflecting a better outcome. Thus, the raw sum scores can range between 10 and 50. The raw sum scores are then converted to the equivalent Rasch-transformed measure [0 (worst) to 100 (best)]. The Jawline SSS (JS³) is a version of the generic SSS, which is specific to measuring satisfaction following aesthetic treatment to the jawline. Questions contained within the JS³ are depicted in Table 1.

Rasch Analysis

Ten items relating to satisfaction with treatment following aesthetic rejuvenation of the jawline were answered by 40 female subjects with ages ranging between 26 and 78 years (M: 49.27; SD: 11.02) as part of a clinical trial.⁸ The Ministep Rasch

(Version of Winsteps 5.5.1.0) was used to apply the Rasch model to the raw data.⁹ Transformation of the linear Rasch model log-odds units (logits) was performed for enhanced interpretability, as they differ from scores that users may be familiar with. Therefore, it was desirable to transform logits into more user-friendly ranges (ie, 0-100) while preserving their linear properties.

Reliability and Separation of Items and Persons. In the Rasch model, reliability has its usual (ie, Cronbach's alpha) definition of "true" person/item measure variance divided by observed person/item measure variance. Thus, the reliability coefficients can range from zero to one. Separation is the number of statistically different performance strata that the test can identify in the sample, and its coefficients can range from zero to infinity.

Person (Test) Reliability: Person reliability measures the degree of homogeneity among the scale items. In Winsteps, person reliability is equivalent to traditional test reliability (ie, Cronbach's alpha), with the main difference being that Cronbach's alpha approximates reliability using an analysis of variance, while the Rasch model approximates it using a measure of standard errors. Person, or test reliability, depends mainly on (i) the sample agreement variance, with wider ranges equating higher person reliability; (ii) the length of the rating scale, with longer lengths equating higher person reliability; (iii) the number of categories per item, with more categories equating higher person reliability; and (iv) the sample-item targeting, with better targeting equating higher person reliability. Person reliability is independent of sample size and is largely uninfluenced by model fit.

Item Reliability: The item reliability coefficient provided by Winstep has no traditional equivalent. Item reliability mainly depends on i) the item difficulty variance, with a wide difficulty, range equating to high item reliability, and ii) the person sample size, with large samples equating a high item reliability. It is independent of test length and is not influenced by model fit. Usually, low item reliability values are the result of a small person sample size that precedes the ability to establish a reproducible item difficulty hierarchy.

In the Rasch model, reliability coefficients are given by the following equation (Equation 1):

$$Reliability = (OV - EV) / OV$$

where OV = observed variance of agreeability measures, and EV = mean of squared standard errors of person agreeability measures.

Separation of Persons: Person separation is used to distinguish people, with low values (< 2) implying that the instrument may not be sensitive enough to distinguish between high and low responders.

Separation of Items: Item separation is used to verify the instrument's construct validity, with low item separation (< 3) implying that the person sample is not large enough to confirm the item difficulty hierarchy.

In the Rasch model, separation coefficients are given by the following equation (Equation 2):

$$Separation(G) = \sqrt{Reliability / (1 - Reliability)} \\ = \text{"True" standard deviation} / \text{Error standard deviation}$$

Item and Person Fit Statistics. Infit estimates measure the aberrations among inliers, and outfit estimates measure the aberrations among the outliers. Items were ordered based on their fit statistics and evaluated based on the critical values proposed by Linacre (ie, 0.50-1.59),¹⁰ as these values are valid for sample sizes ≤ 150 and items ≤ 10 .¹¹

The equations (derived from Bond and Fox)¹² for the fit statistics are presented below (Equation 3):

$$outfit(ui) = \sum_{n=1}^N \frac{(Z_{ni})^2}{N} \\ Infit(vi) = \frac{\sum_{n=1}^N (Z_{ni})^2}{\sum_{n=1}^N W_{ni}}$$

where Z_{ni} is the standardized residual variance for each item/person interaction, and W_{ni} is the individual residual variance.

Unidimensionality and Local Independence. Unidimensionality and local independence of items are 2 related assumptions underlying the Rasch model. Psychometric unidimensionality indicates that only one latent trait (ie, satisfaction) is necessary to account for respondents' agreeability.¹³ Relatedly, the assumption of local independence means that a latent trait is the only construct that determines the probability of agreeing to items. When the assumption of local independence of items is met, there should not be any correlation between 2 items after the effect of the underlying trait is conditioned out (ie, the correlation of residuals should be \sim zero). In effect, this ensures that agreement or disagreement on one item should not lead to an agreement or disagreement on another item.¹⁴ If there are significant correlations among the items after the contribution of the latent trait is removed, then there are dimensions assessed by the scale which are not accounted for by the model.¹⁵ While there are no universally agreed-upon critical values for indicating unidimensionality and local dependence, unidimensionality can be assumed when the unexplained variance is relatively low in comparison to the variance explained by measures, and local independence can be assumed when correlations between questions < 0.50 .¹⁶

Mathematically, the local independence assumption is formulated as (Equation 4):

$$P(X = x|\theta) = \prod_{i=1}^I P_i(X_i = x|\theta)$$

Table 2. Reliability and Separation Estimates for Items and Persons.

Person	Input		Measured		Infit		Outfit	
	Total	Count	Measure	Real SE	IMNSQ	ZSTD	OMNSQ	ZSTD
Mean	28.5	9.2	2.3	0.84	1.03	-0.1	0.96	-0.2
P.SD	6.4	1.1	2.48	0.38	0.73	1.3	0.66	1.1
Real RMSE	0.92	True SD	2.31	Separation	2.5	Person reliability		0.86
Mean	214	69.2	0	0.26	1.01	-0.1	0.97	-0.2
P.SD	27.3	3.6	1.28	0.3	0.3	1.8	0.31	1.7
Real RMSE	0.27	True SD	1.25	Separation	4.72	Item reliability		0.96

Table 3. The Misfit Order for Item Statistics.

Entry number	Total score	Total count	JMLE measure	Model SE	Infit		Outfit		PTmeasur-AL		Exact Obs.%	Match Exp.%	Item
					MNSQ	ZSTD	MNSQ	ZSTD	Corr.	Exp.			
2	199	67	0.37	0.24	1.45	2.32	1.55	2.66	A 0.64	0.77	60.7	65.3	3
6	245	69	-2.09	0.28	1.43	2.10	1.13	0.46	B 0.60	0.66	66.1	73.0	7
7	188	69	1.71	0.24	1.33	1.80	1.34	1.84	C 0.73	0.79	51.6	63.2	8
5	187	69	1.61	0.24	1.15	0.90	1.16	0.94	D 0.78	0.80	58.1	64.1	6
3	215	69	-0.15	0.24	0.88	-0.65	0.99	0.04	E .073	0.75	76.2	67.5	4
10	269	75	-2.27	0.27	0.99	0.10	0.77	-0.53	e 0.67	0.63	75.0	73.1	1
9	188	62	0.37	0.25	0.91	-0.42	0.88	-0.59	d 0.81	0.76	67.9	66.3	10
1	244	75	-0.67	0.24	0.75	-1.56	0.75	-1.28	c 0.77	0.72	75.0	67.7	2
4	205	70	0.74	0.24	0.60	-2.65	0.58	-2.75	b 0.87	0.78	77.8	64.6	5
8	200	67	0.37	0.24	0.60	-2.62	0.59	-2.68	a 0.83	0.75	80.6	65.3	9
Mean	214	69.2	0	0.25	1.01	-0.08	0.97	-0.19			68.9	67.0	
P.SD	27.3	3.6	1.28	0.02	0.3	1.75	0.31	1.68			9.2	3.3	

where θ represents a respondent's latent trait estimates, X represents a vector of possible responses from items and their responses (x), and i represents items.

Person Agreeability and Item Difficulty. A Wright map was used as a tool for evaluating the distributions of a person's agreeableness and the item difficulty levels. The Wright map provided a visual of the scale parameters by placing the difficulty of the items on the same measurement scale as the agreeableness of the respondents.¹⁷

Results

Rasch Analysis

Table 2 displays the reliability and separation estimates for items and persons. Table 3 displays the misfit order for item statistics. Tables 4 and 5 display measures of unidimensionality and local independence, respectively.

Reliability and Separation of Items and Persons.

Person (Test) Reliability: The person reliability estimate of the sample was 0.86, corresponding to a high level of

internal consistency (ie, items of the scale are measuring the same construct).

Item Reliability: The item reliability estimate was 0.96, indicating that the sample was large enough to precisely locate the items on the latent variable.

Separation of Persons: The person separation estimate was 2.50, indicating that the scale was able to distinguish between high and low responders.

Separation of Items: The item separation estimate was 4.72, validating the construct of the instrument (ie, the scale accurately assesses what it is supposed to).

Item and Person Fit Statistics. All infit/outfit values fell within established critical values (0.50-1.59).

Unidimensionality and Local Independence. The unexplained variance in the first contrast was low (~2.0 eigenvalue units) in comparison to the variance explained by measures (14.5239 eigenvalue units). All correlations between questions were <0.50.

Person Agreeability/Satisfaction and Item Difficulty. The Wright map displaying the difficulty of items and agreeability/

Table 4. Measures of Unidimensionality and Local Independence.

	Eigenvalue	Observed	Expected
Total raw variance in observations	24.5239	100.00%	100.00%
Raw variance explained by measures	14.5239	59.20%	59.20%
Raw variance explained by persons	9.6503	39.40%	39.40%
Raw variance explained by items	4.8736	19.90%	19.90%
Raw unexplained variance (total)	10.0000	40.80%	100.00% 40.80%
Unexplained variance in first contrast	2.0657	8.40%	20.70%
Unexplained variance in second contrast	1.7826	7.30%	17.80%
Unexplained variance in third contrast	1.2432	5.10%	12.40%
Unexplained variance in fourth contrast	1.0459	4.30%	10.50%
Unexplained variance in fifth contrast	0.9984	4.10%	10.00%

Observed: variance components for the observed data.

Expected: variance components expected for these data if they exactly fit the Rasch model, i.e., the variance that would be explained if the data accorded with the Rasch definition of unidimensionality.

Total raw variance in observations: total raw-score variance in the observations.

Raw unexplained variance (total): raw-score variance in the observations not explained by the Rasch measures.

Raw variance explained by measures: raw-score variance in the observations explained by the Rasch item difficulties, person abilities and polytomous scale structures.

Raw variance explained by persons: raw-score variance in the observations explained by the Rasch person abilities (and apportioned polytomous scale structures).

Raw variance explained by items: raw-score variance in the observations explained by the Rasch item difficulties (and apportioned polytomous scale structures).

Raw unexplained variance (total): raw-score variance in the observations not explained by the Rasch measures.

Unexplained variance in 1st, 2nd, ... contrast: variance that is not explained by the Rasch measures is decomposed into Principal Component Analysis (PCA), components = Contrasts. The size of the first, second, ... contrast (component) in the PCA decomposition of standardized residuals (or as set by $PRCOMP-*/R4=$), i.e., variance that is not explained by the Rasch measures, but that is explained by the contrast. At most, 5 contrasts are reported. If less than 5 are reported, then the other contrasts have negative eigenvalues, usually due to the data overfitting the Rasch model.

satisfaction of persons is displayed in Figure 2. The relative position of the questionnaire items on the Wright map was based on the likelihood that patients would indicate agreeing or being satisfied with the parameter being measured.

Score Transformations to Logits and Conversion to Rasch Measurements. Given that the data fit the Rasch model, person and item locations (scores) were logarithmically transformed into a common unit of measurement (ie, logits). To facilitate interpretability, logits were converted to Rasch

Table 5. Largest Standardized Residual Correlations Used to Identify Dependent Item.

Correlation	Entry		Entry	
	Number	Item	Number	Item
0.27	5	6	9	10
0.25	1	2	10	1
-0.40	2	3	5	6
-0.35	7	8	10	1
-0.34	1	2	6	7
-0.33	3	4	9	10
-0.32	2	3	8	9
-0.27	5	6	6	7
-0.26	1	2	7	8
-0.26	3	4	5	6
-0.25	1	2	9	10
-0.24	2	3	4	5
-0.23	9	10	10	1
-0.22	3	4	4	5
-0.22	7	8	9	10
-0.21	3	4	8	9
-0.20	2	3	7	8
-0.20	6	7	10	1
-0.19	5	6	7	8
-0.18	4	5	7	8

measurements using Winsteps' internal software (Table 6). In practice, these measurements would be available for use in standard statistical analyses investigating treatment and/or group effects.

Discussion

This report describes the development of JS³, a multiitem questionnaire assessing subject satisfaction with treatment results following aesthetic rejuvenation of the jawline using the Rasch approach (Figures 3 and 4). The results of the Rasch analysis support the psychometric strengths, specifically reliability and validity, of the JS³, making it a suitable tool for measuring posttreatment subject satisfaction. Among the strengths of the JS³, it is notable that all 10 items of the scale fit the Rasch model. In addition, it returns a unidimensional measure of satisfaction with treatment effects and possesses high construct validity and reliability.

The hierarchical orders of the items on the Wright map propose a hypothesis concerning theoretical aspects related to subject satisfaction. It is worth noting that the easiest items for subjects to positively endorse were Items 7 and 1, which concerned the naturalness of results and overall satisfaction, respectively. This is an important point, as many clinicians only assess overall satisfaction in practice and these insights may be beneficial for optimizing clinical treatment regimens and marketing purposes. Moreover, given that the lower third of the face is a highly complex area containing multiple aging anatomic structures (eg, increased bone resorption, changes in

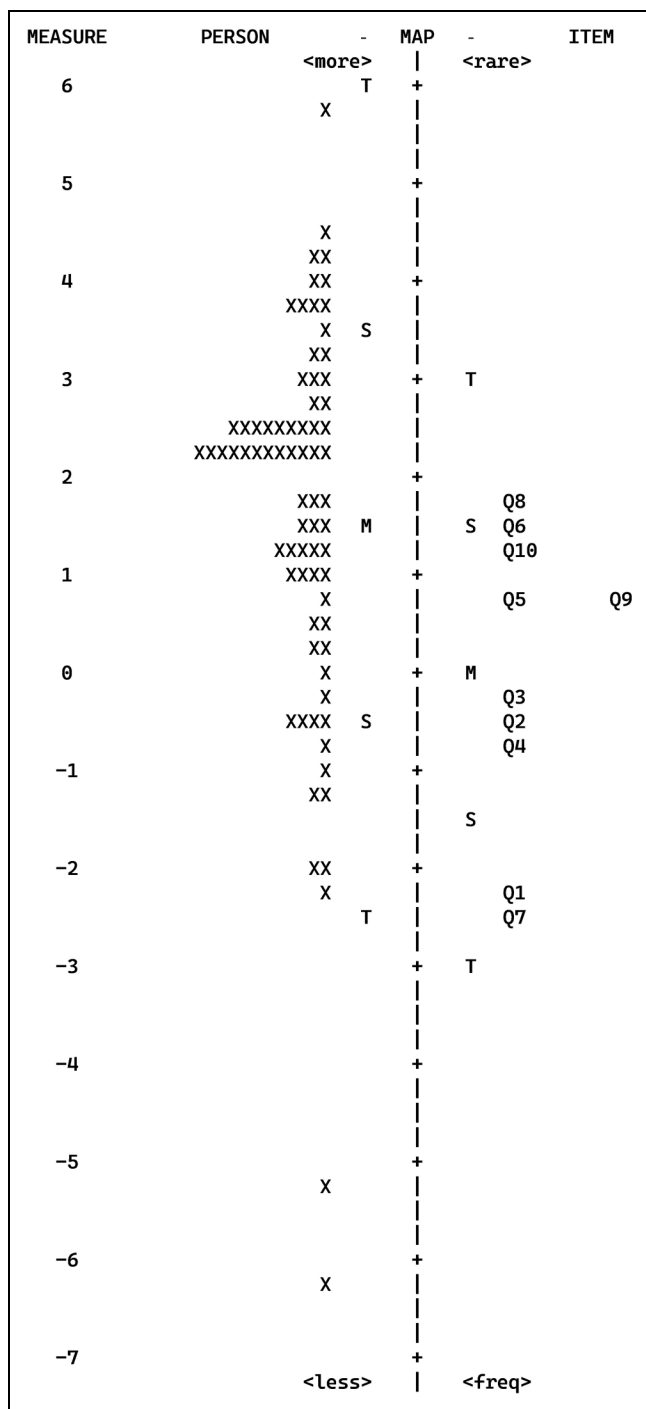


Figure 2. Wright’s map is organized as 2 vertical histograms, with the difficulty of each item on the right and the agreeability of each person on the left. The difficulty of items and agreeability of persons are distributed from most (top) to least (bottom).

the position of deep and superficial fat pads, decreased collagen, reduced elasticity),¹⁸ the JS³ represents a relevant tool for precisely evaluating patient satisfaction with multiple treatment effects. This becomes even more significant when considering that the lower third of the face is often the hardest area to treat in older patients. Therefore, the development of a

measurement tool capable of isolating the lower third of the face becomes even more pertinent.

Limitations and Future Directions

The “ceiling effect” is a limitation of many satisfaction questionnaires,^{19,20} and is a result of the difficulty in subjects being completely satisfied with all items of the scale (ie, there is no room for improvement).²¹ In this study sample, only 4/40 (10%) subjects totaled the maximum score, and this may have hidden some of the association between subjects’ clinical improvement and treatment satisfaction.²²

Currently, the JS³ is only available in English. Certified translations into other languages should be conducted and future studies could replicate the current methodology on different language versions, while taking into consideration standardized guidelines for cross-cultural adaptation of PROMs.²³ In addition, the authors plan on validating other versions of the generic SSS, to be applicable to aesthetic treatments in different regions of the face (eg, lips, midface, and tear troughs).

Previously, the Rasch model has been used to develop and validate other scales used in the aesthetics field. Most notably is the FACE-Q aesthetics module,²⁴ which consists of 34 scales and 6 checklists that measure a variety of important concepts of interest to surgical and nonsurgical aesthetic patients, including facial appearance, adverse events, patient expectations, appearance-related psychosocial distress, and quality of life.²⁵ Other condition-specific PROMs that have been developed using Rasch measurement theory include those for acne and related scarring, body contouring, and craniofacial deformities.²⁶ In addition to their favorable psychometric properties, the clinical utility of these PROMs has been evidenced in many clinical trials evaluating the results of blepharoplasties, face-lifts, nonsurgical facial rejuvenation, orthognathic surgery, rhinoplasty, lip augmentation, facial feminization/masculinization procedures, chin augmentation, among others.²⁴ To the authors’ knowledge, the only PROM previously developed specifically for the jawline includes the FACE-Q – Satisfaction with Lower Face and Jawline scale (FQ-SLFJ), which consists of 5 questions rated on a 4-point scale (1 = very dissatisfied, 2 = somewhat dissatisfied, 3 = somewhat satisfied, 4 = very satisfied). Each of the 5 questions measures a different parameter associated with patient satisfaction with the appearance of the lower face and jawline (ie, prominence and definition of the jawline, side view, how nice, and smooth the lower face appears). Conversely, the JS³ consists of 9 questions which are answered using a 5-point scale, with a neutral option anchoring the middle of the response categories. Although Rensis Likert (1932) advocated for the use of the 5-point scale,²⁷ the inclusion of a neutral option in surveys has since been debated but is generally substantiated by adhering to best practices when a direct comparison is being made (eg, before vs after treatment).²⁸ Including a neutral option as the midway point also avoids the exaggeration of negative and/or positive responses by forcing responders to make a choice in the case of no preference/difference. Moreover, while the

FQ-SLFJ is a static (cross-sectional) measurement tool, the JS³ measures satisfaction with *treatment results* and thus has an

Table 6. Raw Sum Score to Rasch-Transformed Measures Conversion Table.

Score	Measure	SE	Score	Measure	SE	Score	Measure	SE
10	.00E	10.58	24	35.44	2.87	38	61.99	3.41
11	7.47	6.12	25	36.91	2.89	39	64.05	3.42
12	12.37	4.65	26	38.40	2.93	40	66.14	3.46
13	15.68	4.06	27	39.95	2.99	41	68.30	3.52
14	18.34	3.72	28	41.56	3.06	42	70.54	3.60
15	20.64	3.50	29	43.27	3.16	43	72.88	3.68
16	22.69	3.33	30	45.10	3.27	44	75.34	3.77
17	24.57	3.20	31	47.05	3.38	45	77.92	3.88
18	26.31	3.09	32	49.13	3.47	46	80.68	4.03
19	27.96	3.01	33	51.29	3.52	47	83.71	4.28
20	29.53	2.95	34	53.49	3.53	48	87.30	4.80
21	31.04	2.91	35	55.58	3.50	49	92.42	6.20
22	21.53	2.88	36	57.82	3.46	50	100.00E	10.62
23	33.99	2.87	37	59.92	3.43			



Figure 3. Three-dimensional images of a 58-year-old woman at baseline (a) and month 1 (b) following hyaluronic acid (Restylane Lyft) injections for contouring the jawline [total volume administered over 2 visits; baseline = 4.0 mL; week 2 = 4.3 mL (total = 8.3 mL)].

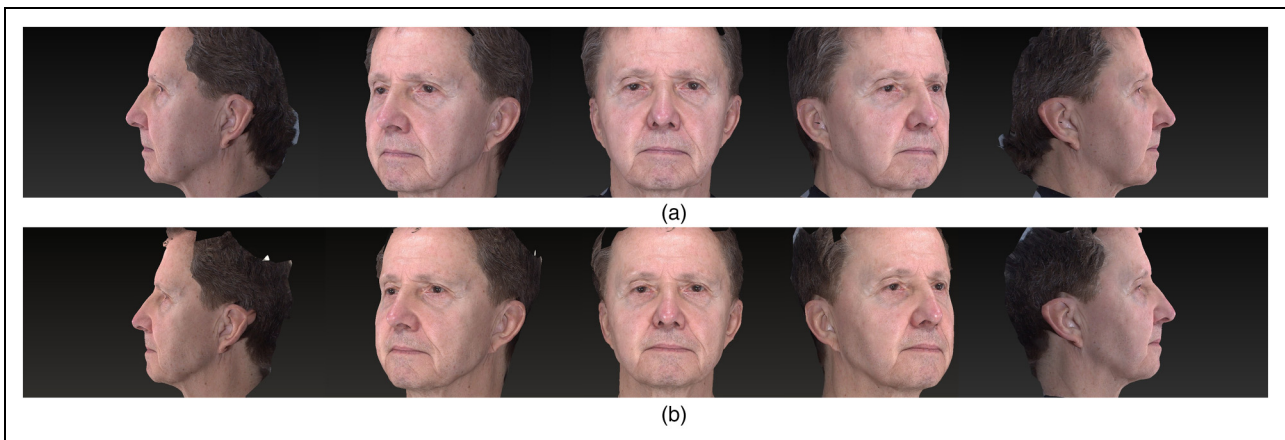


Figure 4. Three-dimensional images of a 66-year-old man at baseline (a) and month 1 (b) following hyaluronic acid (Restylane Lyft) injections for contouring the jawline [total volume administered over 2 visits; baseline = 4.6 mL; week 2 = 1.4 mL (total = 6.0 mL)].

implicit temporal feature associated with changes over time. This makes statistical models based on the JS³ less challenging to handle the dynamics of causality as direct inferences are captured, whereas the FQ-SLFJ ignores treatment assignments and can only demonstrate a correlation. Nonetheless, given the differences (no overlap) in the parameters assessed by the 2 scales, both could be used to investigate the heterogeneity of treatment effects.

Conclusion

Application of the Rasch model produced a valid and reliable scale (ie, JS³) for measuring satisfaction with the appearance of the jawline following aesthetic treatments.

Author's Contributions

Kaitlyn M. Enright: conceptualization, methodology, formal analysis, and writing—original draft. John S. Sampalis: validation, writing—review and editing. Anneke Andriessen: validation, writing—review and editing. Andreas Nikolis: supervision, writing—reviewing and editing.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Ethical Approval

Written informed consent was obtained from the patients for the publication of their photographs. This research involved the secondary analysis of an anonymous dataset, and as such, was exempt from review by a Research Ethics Board (TCPS2 article 2.4). This article does not contain any primary studies with human or animal subjects.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Kaitlyn M. Enright  <https://orcid.org/0000-0001-6555-4838>

References

- Pallant JF, Miller RL, Tennant A. Evaluation of the Edinburgh post natal depression scale using Rasch analysis. *BMC Psychiatry*. 2006;6(1):28. doi:10.1186/1471-244X-6-28
- Muller S, Roddy E. A Rasch analysis of the Manchester foot pain and disability index. *J Foot Ankle Res*. 2009;2(1):29. doi:10.1186/1757-1146-2-29
- Tor E, Steketee C. Rasch analysis on OSCE data : an illustrative example. *Australas Med J*. 2011;4(6):339-345. doi:10.4066/AMJ.2011.755
- Andrich D. *Final Report on the Psychometric Analysis of the Early Development Instrument (EDI) Using the Rasch Model: A Technical Paper Commissioned for the Development of the Australian Early Development Instrument (AEDI)*. Royal Children's Hospital; 2004.
- DeVellis RF, Thorpe CT. *Scale Development: Theory and Applications*. Sage Publications; 2021.
- Hendriks J, Fyfe S, Styles I, Skinner SR, Merriman G. Scale construction utilising the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. *Australas Med J*. 2012;5(5):251-261. doi:10.4066/AMJ.2012.952
- Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Paedagogiske Institut; 1960.
- Nikolis A. Comparative trial evaluating two hyaluronic acid fillers manufactured with different technologies for contouring the jawline.
- Linacre JM. *A User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs*; 2006. Accessed December 7, 2023. <http://archive.org/details/B-001-003-730>.
- What do Infit and Outfit, Mean-square and Standardized mean? Accessed December 7, 2023. <https://www.rasch.org/rmt/rmt162f.htm>.
- Müller M. Item fit statistics for Rasch analysis: Can we trust them? *J Stat Distrib Appl*. 2020;7(1):5. doi:10.1186/s40488-020-00108-7
- Applying the Rasch Model | Fundamental Measurement in the Human Science. Accessed December 7, 2023. <https://www.taylorfrancis.com/books/mono/10.4324/9781315814698/applying-rasch-model-trevor-bond>
- Hambleton RK, Swaminathan H. *Item Response Theory: Principles and Applications*. Springer Science & Business Media; 2013.
- Lord FM, Novick MR, Birnbaum A. *Statistical Theories of Mental Test Scores*. Addison-Wesley; 1968.
- Lee YW. Examining passage-related local item dependence (LID) and measurement construct using Q 3 statistics in an EFL reading comprehension test. *Lang Test - LANG TEST*. 2004;21:74-100. doi:10.1191/0265532204lt260oa
- Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Meas*. 2017; 41(3):178-194. doi:10.1177/0146621616677520
- Reliability and Separation. Accessed December 7, 2023. <https://www.rasch.org/rmt/rmt94n.htm>.
- Safran T, Swift A, Cotofana S, Nikolis A. Evaluating safety in hyaluronic acid lip injections. *Expert Opin Drug Saf*. 2021; 20(12):1473-1486. doi:10.1080/14740338.2021.1962283
- Moret L, Nguyen JM, Pillet N, Falissard B, Lombrail P, Gasquet I. Improvement of psychometric properties of a scale measuring inpatient satisfaction with care: A better response rate and a reduction of the ceiling effect. *BMC Health Serv Res*. 2007;7(1):197. doi:10.1186/1472-6963-7-197
- Caronni A, Ramella M, Arcuri P, et al. The Rasch analysis shows poor construct validity and low reliability of the Quebec user evaluation of satisfaction with assistive technology 2.0 (QUEST 2.0) questionnaire. *Int J Environ Res Public Health*. 2023;20(2): 1036. doi:10.3390/ijerph20021036
- Badejo MA, Ramtin S, Rossano A, Ring D, Koenig K, Crijns TJ. Does adjusting for social desirability reduce ceiling effects and

- increase variation of patient-reported experience measures? *J Patient Exp.* 2022;9:23743735221079144. doi:10.1177/23743735221079144
22. Carreon LY, Sanders JO, Diab M, Sturm PF, Sucato DJ, Spinal Deformity Study Group. Patient satisfaction after surgical correction of adolescent idiopathic scoliosis. *Spine.* 2011;36(12):965-968. doi:10.1097/BRS.0b013e3181e92b1d
 23. Beaton D, Bombardier C, Guillemin F, Ferraz M. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures Contributors to this Document. Inst Work Health. 2007;1. - Google Search. Accessed December 7, 2023. <https://www.google.com/search?client=safari&rls=en&q=Beaton+D%2C+Bombardier+C%2C+Guillemin+F%2C+Ferraz+M.+Recommendations+for+the+Cross-Cultural+Adaptation+of+the+DASH+%26+QuickDASH+Outcome+Measures+Contributors+to+this+Document.+Inst+Work+Health.+2007%3B1.&ie=UTF-8&oe=UTF-8>.
 24. Gallo L, Kim P, Yuan M, et al. Best practices for FACE-Q aesthetics research: a systematic review of study methodology. *Aesthet Surg J.* 2023;43(9):NP674-NP686. doi:10.1093/asj/sjad141
 25. 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
 26. Klassen AF, Rae C, Wong Riff KW, et al. FACE-Q craniofacial module: Part 1 validation of CLEFT-Q scales for use in children and young adults with facial conditions. *J Plast Reconstr Aesthet Surg.* 2021;74(9):2319-2329. doi:10.1016/j.bjps.2021.05.040
 27. Likert R. A technique for the measurement of attitudes. *Arch Psychol.* 1932;22(140):55-55.
 28. Johns R. One size doesn't fit all: selecting response scales for attitude items. *J Elections Public Opin Parties.* 2005;15(2):237-264. doi:10.1080/13689880500178849