Establishing Convergent Validity of the FACE-Q Aesthetics Module Scales

Lucas Gallo, MD, MSc, PhD(c)◉; Patrick Kim, MD◉;
Isabella Churchill, MD, MSc◉; Charlene Rae, PhD◉;
Sophocles H. Voineskos, MD, MSc; Achilles Thoma, MD, MSc◉;
Andrea L. Pusic, MD, MHS; Stefan J. Cano, PhD;
and Anne F. Klassen, BA (Hons), DPhil◉

OXFORD
UNIVERSITY PRESS

Abstract

Background: FACE-Q Aesthetics scales can be used to assess patient-important outcomes following both surgical and nonsurgical facial cosmetic interventions. Convergent validity is the degree to which the scores of one measurement relate to another measuring a similar construct.

Objectives: The aim of this study was to establish the convergent validity of 11 FACE-Q Aesthetics appearance scales vs the MERZ Aesthetics (Raleigh, NC) scales.

Methods: Data were collected from an online international sample of participants aged ≥20 years, who had presented to a dermatologist or plastic surgeon within the previous 12 months for a minimally invasive facial aesthetic treatment. Participants provided demographic and clinical data and completed 11 FACE-Q Aesthetics scales and 12 MERZ Aesthetics scales. Hypotheses regarding the strength of correlations between these scales were generated a priori. Adequate convergent validity was based on the percentage of correct hypotheses (>75%) and/or correlation ≥0.50 with an instrument measuring a similar construct.

Results: In total, 1259 participants were included in this survey. The mean [standard deviation] age of the participants was 42.6 [11.9] years old, and most were female (72.5%), Caucasian (76.9%), and living in the United States (49.9%) or the United Kingdom (42.9%). FACE-Q Lines Overall, Lower Face and Jawline, Appraisal of Lines—Forehead/Between Eyebrows/Crow's Feet/Lips/Nasolabial Folds/Marionette, and Lips scales demonstrated adequate convergent validity with patient-reported MERZ Aesthetics scales. The FACE-Q Face Overall and Cheeks scales did not show adequate convergent validity.

Conclusions: This study provides evidence of convergent validity for FACE-Q Aesthetics appearance scales. Establishing the validity of these scales remains an iterative process and further studies comparing the FACE-Q to other related measurement tools are required to strengthen this evidence.

Level of Evidence: 4 (Diagnostic)

FACE-Q Aesthetics is a patient-reported outcome measure (PROM) designed to evaluate outcomes following both surgical and minimally invasive facial aesthetics procedures.[1,2] This PROM has 4 key domains (ie, facial appearance, health-related quality of life, naturalness, and adverse effects of treatment) and consists of 37 independently functioning scales and 6 checklists.[3-5]

Although the FACE-Q Aesthetics module has been found to be reliable and responsive to change, establishing the validity of this PROM remains an iterative process.[1,6,7] To date, the convergent validity of these scales has yet to be established, representing a gap within the current literature. "Convergent validity" refers to how closely a scale or test relates to another tool designed to measure the same or similar

Drs Gallo, Kim, and Churchill are plastic surgery residents, and Dr Thoma is a plastic surgeon, Division of Plastic Surgery, McMaster University, Hamilton, ON, Canada. Dr Rae is a research associate and Dr Klassen is a professor, Department of Pediatrics, McMaster University, Hamilton, ON, Canada. Dr Voineskos is a plastic surgeon, Division of Plastic and Reconstructive Surgery, University of Toronto, Toronto, ON, Canada. Dr Pusic is a plastic surgeon, Division of Plastic Surgery, Brigham and Women's Hospital, Boston, MA, USA. Dr Cano is chief scientific officer, Modus Outcomes, Statfold, United Kingdom.

Corresponding Author:
Dr Lucas Gallo, 3N27, 1280 Main Street W, Hamilton, ON L8N 3Z5, Canada.
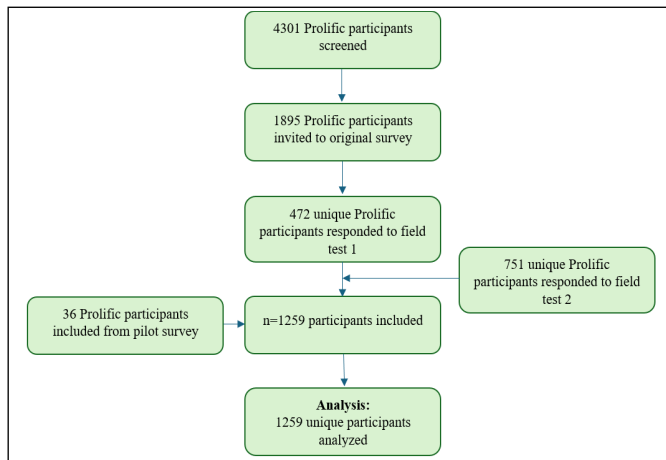E-mail: lucas.gallo@medportal.ca

Figure 1. Flow diagram of participant selection.

Table 2. Hypotheses to Test Construct Validity

Hypothesis	
A	Correlations with instruments measuring similar constructs should be ≥0.50 (strong correlation)
B	Correlations with instruments measuring related, but dissimilar constructs should be lower, ie, 0.30-0.50 (moderate correlation)
C	Correlations with instruments measuring unrelated constructs should be <0.30 (weak to negligible correlation)

Table 1. Selected FACE-Q Aesthetic Modules and MERZ Aesthetic Scales

FACE-Q Aesthetic scales	MERZ Aesthetic scales
Lower Face and Jawline	Jawline—At Rest
Appraisal of Lines—Forehead	Forehead Lines—Dynamic
Appraisal of Lines—Between Eyebrows	Glabella Lines—Dynamic
Appraisal of Lines—Crow's Feet	Crow's Feet—Dynamic
Appraisal of Lines—Nasolabial Folds	Nasolabial Folds—At Rest
Appraisal of Lines—Marionette	Marionette Lines—At Rest
Appraisal of Lines—Lips	Lip Wrinkles—At Rest
Lips	Upper Lip Fullness—At Rest Lower Lip Fullness—At Rest Oral Commissure—At Rest
Cheeks	Upper Cheek Fullness—At Rest Lower Cheek Fullness—At Rest
Face Overall	
Lines Overall	

Participant Recruitment

This study was completed as a part of a larger planned analysis to validate new FACE-Q Aesthetics "natural" scales and item libraries for satisfaction with face and psychological well-being.[7] A detailed description of participant selection and recruitment are reported elsewhere.[7]

Following a pilot survey of 144 individuals, a larger sample of participants from the United States and Canada were invited to complete the survey in December 2022, and the United Kingdom in August 2023, through the online crowd-sourcing platform Prolific (London, UK). Participants were compensated the equivalent of 10.80 GBP per hour for their time.

Participants were eligible to participate in this survey if they self-selected that: (1) they were ≥20 years old; (2) lived in Canada, the United Kingdom, or the United States; and (3) had been to a dermatology or a plastic surgery clinic in the past 12 months to receive one of 14 facial aesthetic treatments (Supplemental Table 1). Figure 1 demonstrates a flow diagram of participant selection.

Scales

Participants were required to complete 11 FACE-Q Aesthetics appearance scales and 12 MERZ Aesthetics scales. We selected these scales a priori by because they were deemed to measure similar constructs. The selected scales are shown in Table 1.

FACE-Q Aesthetics scales are scored by converting raw ordinal scores into a continuous score between 0 and 100, where higher scores correspond to higher satisfaction with appearance or better health-related quality of life. The description of individual scales can be found in the FACE-Q Aesthetics User's Guide (https://qportfolio.org/face-q/aesthetics/).[4]

The MERZ Aesthetics scales are scored ordinally between 0 to 3 or 0 to 4 and correspond to a spectrum of photographs for a selected aspect of the face. For example, the MERZ Jawline—At Rest can be scored between 0 (no sagging) to 4 (very severe sagging). Higher scores correspond to less favorable appearance for all MERZ

construct (ie, the concept being studied). Specifically, if 2 scales designed to measure similar constructs (eg, facial appearance) have a moderate or high correlation, this can be viewed as evidence of convergent validity.[8] Ultimately, this aspect of validity gives clinicians and researchers confidence that the scale being used accurately measures what it was intended to measure.

Typically, convergent validity is established through hypothesis testing, where investigators make a priori predictions about the strength of the correlation between 2 related measures. The evidence supporting an instrument's convergent validity becomes stronger as more predefined hypotheses are tested and determined to be correct.[8]

To establish the convergent validity of the FACE-Q Aesthetics module, it can be compared with another validated measure of facial aesthetics commonly used within the literature such as the MERZ Aesthetics (Raleigh, NC) scales. This measure consists of photonumeric scales designed to evaluate the appearance of the upper, middle, and lower face.[9] To date, the MERZ Aesthetics scales have been validated for use by clinician/expert raters and have been shown to have good reliability.[10-12]

The primary aim of this study was to establish the convergent validity of FACE-Q Aesthetics scales vs the MERZ Aesthetics scales in an online, international sample of participants.

METHODS

This online cross-sectional survey was coordinated at McMaster University (Hamilton, ON, Canada) and received ethics approval through the Hamilton Integrated Research Ethics Board (#13603; Hamilton, ON, Canada).

Table 3. Participant Demographics (N = 1259)

Participant characteristics	n	%
Age (years)		
20-29	202	16.0
30-39	277	22.0
40-49	427	33.9
50-59	234	18.6
60-69	99	7.9
70-79	20	1.6
BMI (kg/m^2)		
Underweight (<18.5)	88	7.0
Normal (18.5-24.9)	559	44.4
Overweight (25.0-29.9)	338	26.8
Obese Class I (30.0-34.9)	152	12.1
Obese Class II (35.0-39.)	46	3.7
Obese Class III (>40.0)	42	3.3
Missing/prefer not to answer	34	2.7
Self-identified race		
Caucasian	968	76.9
Black	90	7.1
Hispanic/Latino	29	2.3
South Asian	48	3.8
East Asian	42	3.3
Middle Eastern	8	0.6
Other/prefer not to answer	74	5.9
Gender		
Male	333	26.4
Female	913	72.5
Other	13	1.1
Fitzpatrick Scale		
I	92	7.3
II	343	27.2
III	458	36.4
IV	246	19.5
V	102	8.1
VI	18	1.4

Table 3. Continued

Participant characteristics	n	%
Highest level of education		
Completed some or all of high school	113	9.0
Completed some or all of college/trade school/university	815	64.7
Completed some or all of masters/doctoral degree	330	26.2
Prefer not to answer	1	0.1
Marital status		
Married/common-law	701	55.7
Widowed	13	1.0
Separated/divorced	141	11.2
Single, never married	389	30.9
Other/prefer not to answer	15	1.2
Country of residence		
Canada	89	7.1
United States of America	628	49.9
United Kingdom	540	42.9
Missing/prefer not to answer	2	0.2

Aesthetics scales except for Upper Lip Fullness—At Rest and Lower Lip Fullness—At Rest, for which higher scores correspond to a more favorable appearance.[9]

Sample Size

As per COSMIN guidelines for convergent validity, we aimed to have a minimum of 100 participants provide responses for each scale being examined.[13]

Data Analysis

To assess convergent validity, hypotheses regarding the correlation between the FACE-Q Aesthetics and MERZ Aesthetics scales were generated a priori (Table 2). Hypotheses were adapted from the "COSMIN Methodology for Systematic Reviews of Patient-Reported Outcome Measures User Manual."[14]

Given that the selected FACE-Q Aesthetics scales include items that evaluate participant perceptions of their appearance (eg, how bothered they are by the area being evaluated) and the MERZ scales use standardized photographs for participants to match their appearance (eg, degree of jawline sagging), it was hypothesized that these scales likely measure related, but dissimilar constructs for the purpose of hypothesis testing.

Spearman's rank correlation coefficient (ρ) was used to calculate correlation. Convergent validity is generally considered adequate if >75% of hypotheses are correct or if a correlation with an instrument

Table 4. Summary of Hypothesis Testing

FACE-Q Aesthetic scale	Hypothesis A correct	Hypothesis B correct	Hypothesis C correct	Total correct hypotheses	Adequate convergent validity?
Face Overall	NA	0/12 (0.0%)	NA	0/12 (0.0%)	No
Lines Overall	NA	8/8 (100.0%)	4/4 (100.0%)	12/12 (100.0%)	Yes
Lower Face and Jawline	NA	1/1 (100.0%)	11/11 (100.0%)	12/12 (100.0%)	Yes
Appraisal of Lines: Forehead	NA	1/1 (100.0%)	11/11 (100.0%)	12/12 (100.0%)	Yes
Appraisal of Lines: Between Eyebrows	NA	1/1 (100.0%)	9/11 (81.8%)	10/12 (83.3%)	Yes
Appraisal of Lines: Crow's Feet	NA	1/1 (100.0%)	10/11 (90.9%)	11/12 (91.7%)	Yes
Appraisal of Lines: Nasolabial Folds	NA	1/1 (100.0%)	9/11 (81.8%)	10/12 (83.3%)	Yes
Appraisal of Lines: Marionette	NA	2/2 (100.0%)	8/10 (80.0%)	10/12 (83.3%)	Yes
Appraisal of Lines: Lips	NA	1/1 (100.0%)	11/11 (100.0%)	12/12 (100.0%)	Yes
Lips	NA	3/3 (100.0%)	9/9 (100.0%)	12/12 (100.0%)	Yes
Cheeks	NA	0/2 (0.0%)	10/10 (100.0%)	10/12 (83.3%)	No

n, number of correct hypotheses (numerator); N, total number of hypotheses (denominator). Values are given as n/N (%). NA, not applicable; Hypothesis A: correlations with instruments measuring similar constructs should be ≥0.50; Hypothesis B: correlations with instruments measuring related, but dissimilar constructs should be lower, ie, 0.30-0.50; Hypothesis C: correlations with instruments measuring unrelated constructs should be <0.30;

measuring the same construct is ≥0.50 (Hypothesis A).[8,15] An additional precondition was added by the study authors that all Hypotheses A and/or B should be true for a scale to demonstrate adequate convergent validity.

An analysis of variance (ANOVA) was performed to determine if statistically significant differences exist between related FACE-Q Aesthetics scale scores (ie, continuous data) and MERZ Aesthetic scale scores (ie, categorical data). Statistical significance was considered to be a P-value ≤.05. Missing data were handled using an available-case analysis. All analyses were performed with SPSS version 26.0 for Windows (IBM Corporation, Armonk, NY).

RESULTS

Following completion of this cross-sectional survey, n = 1259 participants were included in this analysis. The mean [standard deviation] age of the sample was 42.6 [11.9] years (range, 20-75 years), with 913 (72.5%) female, 332 (26.4%) male, and 14 (1.1%) gender diverse/other. Most survey respondents self-identified as Caucasian (76.9%) and lived in the United States (49.9%) or the United Kingdom (42.9%). Additional sample demographics are reported in Table 3.

Convergent Validity

The results of hypothesis testing are reported in Supplemental Table 2 and summarized in Table 4. Specifically, the FACE-Q: Lines Overall scale was found to have adequate convergent validity. This scale had moderate correlations with all chosen MERZ Aesthetics photonumeric scales except for Upper/Lower Lip fullness and Upper/Lower Cheek fullness, which were hypothesized to have weaker correlations because these scales do not measure the presence of lines. Overall, 100% of hypotheses were correct.

Additionally, the region-specific FACE-Q Aesthetics appearance and lines scales (ie, Lower Face and Jaw, Appraisal of Lines—Forehead/ Between Eyebrows/Crow's Feet/Nasolabial Folds/Marionette/Lips,

and Lips) were found to have adequate convergent validity. These scales demonstrated moderate correlations with MERZ scales that measured related constructs and weaker correlations with MERZ scales that measured unrelated constructs. Overall correct hypotheses ranged from 83.3% to 100%.

The FACE-Q scales that did not demonstrate convergent validity through this analysis were the Face Overall and Cheeks scales. Notably, the Face overall scale was hypothesized to have moderate correlation with all MERZ scales but was only weakly correlated with all scales. Thus both Hypotheses B and C were incorrect. Furthermore, the Cheeks scale was hypothesized to have moderate correlation with the MERZ: Upper/Lower Cheek fullness scales but was found to be weakly correlated with all MERZ scales. As such, Hypothesis B for both these 2 scales was incorrect, resulting in inadequate convergent validity, despite 83.3% of the hypotheses being correct overall.

ANOVA

Table 5 demonstrates the results of the ANOVA and the mean FACE-Q Aesthetics scale scores for moderately correlated MERZ scales. All FACE-Q scale scores were statistically significantly different between MERZ photonumeric scale scores. For example, patients scoring 0 (no sagging) on the MERZ: Jawline at rest scale had a mean FACE-Q: Lines overall score of 73.5 [21.9], while patients scoring 4 (very severe sagging) on the scale MERZ scale had a mean FACE-Q: Lines overall score of 25.5 [17.7].

DISCUSSION

FACE-Q Aesthetics was designed to measure outcomes that are important to patients undergoing surgical or nonsurgical cosmetic facial interventions. These scales have been broadly validated to demonstrate various properties of validity, reliability, and responsiveness, with a subset of the most frequently used scales qualified for use by the FDA as Medical Device Development Tools.[1-3,5,16-18]

Table 5. FACE-Q Scores for Moderately Correlated Merz Scales

FACE-Q scale	MERZ scale	FACE-Q score (mean [SD])					
		MERZ 0	MERZ 1	MERZ 2	MERZ 3	MERZ 4	*P*-value
Lines Overall	Jawline—At Rest	73.5 [21.9]	61.2 [22.7]	52.2 [21.3]	42.4 [19.4]	25.5 [17.7]	<.001
Lines Overall	Forehead Lines—Dynamic	75.9 [22.5]	68.9 [22.0]	58.9 [21.3]	50.4 [23.3]	43.1 [27.0]	<.001
Lines Overall	Glabella Lines—Dynamic	75.1 [20.9]	66.2 [23.0]	59.7 [21.3]	51.4 [24.1]	49.2 [27.6]	<.001
Lines Overall	Crow's Feet—Dynamic	77.7 [22.2]	67.1 [20.9]	55.2 [21.7]	45.2 [19.8]	38.3 [22.2]	<.001
Lines Overall	Nasolabial Folds—At Rest	76.0 [21.8]	65.5 [22.1]	57.8 [20.9]	41.3 [19.7]	30.9 [21.9]	<.001
Lines Overall	Marionette Lines—At Rest	74.1 [21.4]	63.3 [21.2]	54.1 [21.8]	45.4 [23.5]	35.9 [23.8]	<.001
Lines Overall	Lip Wrinkles—At Rest	71.6 [22.2]	59.5 [21.1]	50.4 [21.8]	40.5 [18.0]	27.0 [18.4]	<.001
Lines Overall	Oral Commissure—At Rest	72.6 [22.4]	64.6 [21.4]	56.1 [22.7]	48.3 [22.8]	36.4 [29.6]	<.001
Lower Face and Jawline	Jawline—At Rest	62.3 [26.6]	50.9 [25.4]	40.9 [25.2]	25.2 [27.3]	25.0 [12.7]	<.001
Appraisal of Lines: Forehead	Forehead Lines—Dynamic	NA	76.3 [21.6]	63.8 [23.4]	54.3 [25.9]	45.4 [28.2]	<.001
Appraisal of Lines: Between Eyebrows	Glabella Lines—Dynamic	NA	73.4 [23.1]	60.1 [25.2]	43.9 [29.2]	44.5 [34.1]	<.001
Appraisal of Lines: Crow's Feet	Crow's Feet—Dynamic	NA	74.6 [22.5]	58.4 [25.0]	45.9 [28.0]	35.6 [30.6]	<.001
Appraisal of Lines: Nasolabial Folds	Nasolabial Folds—At Rest	NA	74.8 [24.8]	62.8 [26.6]	41.9 [26.4]	26.7 [28.5]	<.001
Appraisal of Lines: Marionette	Marionette Lines—At Rest	NA	73.5 [23.7]	57.8 [27.2]	34.4 [28.3]	27.2 [28.9]	<.001
Appraisal of Lines: marionette	Oral Commissure—At Rest	81.9 [21.9]	67.8 [25.4]	56.9 [27.4]	39.9 [29.7]	26.3 [29.2]	<.001
Appraisal of Lines: Lips	Lip Wrinkles—At Rest	NA	75.9 [22.8]	55.7 [30.0]	38.9 [29.0]	11.8 [18.2]	<.001
Lips	Upper Lip Fullness—At Rest	39.6 [30.4]	52.5 [23.9]	66.5 [20.5]	76.5 [19.5]	87.4 [19.1]	<.001
Lips	Lower Lip Fullness—At Rest	44.2 [34.3]	52.4 [25.9]	64.0 [21.3]	72.1 [21.9]	87.0 [19.0]	<.001
Lips	Oral Commissure—At Rest	75.2 [22.2]	63.5 [23.2]	58.9 [22.6]	48.2 [26.4]	46.9 [33.1]	<.001

NA, not applicable; SD, standard deviation. *P*-values were determined from analysis of variance.

The concept of validity as it applies to measurement instruments is defined by the COSMIN group as "the degree to which [a PROM] measures the construct(s) it purports to measure."[19] Within the domain of validity, there exist 3 specific properties: content validity (ie, does the content of the PROM reflect the construct to be measured?), criterion validity (ie, does the score of the PROM adequately reflect the scores of the "gold standard"?), and construct validity (ie, does the PROM adequately measure the intended construct?).[19,20]

In the present study, we assessed the convergent validity of FACE-Q Aesthetics with the MERZ Aesthetics scales and provide evidence of convergent validity for 9 out of 11 scales tested, namely, Lines Overall, Lower Face and Jaw, Appraisal of Lines—Forehead/Between Eyebrows/Crow's Feet/Nasolabial Folds/Marionette/Lips, and Lips scales. Although >75% of a priori defined hypotheses were correct for all scales, we concluded that convergent validity could not be sufficiently supported for the Face Overall and Cheeks scales because hypotheses suggesting a moderate correlation (ie, Hypothesis B) for these scales were incorrect—a precondition at the onset of the study.

There are several hypotheses for why the Face Overall scale may not have been convergent with the MERZ scales. First, the Face Overall scale evaluates the entire face as a whole; therefore satisfaction/dissatisfaction about specific areas of their face may not be reflected in their Face Overall scores.[4] Taking the example of the MERZ scale with the lowest correlation, Lower Lip Fullness—At Rest, raters who believe

they have thin lips do not necessarily have the same degree of dissatisfaction about their overall facial appearance. Notably, the Lines Overall scale was found to have convergent validity despite also evaluating the entire face. However, in the Face Overall scale, the questions ask the rater to evaluate holistic aspects of their appearance, such as symmetry, balance, freshness, etc.[4] In contrast, the Lines Overall scale asks specifically about static and dynamic lines on the face.[4] These questions therefore more closely reflect the various lines (forehead, glabellar, nasolabial folds, etc) depicted in the MERZ scales.[9]

The FACE-Q Cheeks scale was also not convergent with its respective MERZ Upper/Lower Cheek Fullness—At Rest scales. This may be because the FACE-Q scale asks the rater to assess multiple domains of their cheek appearance (symmetry, smoothness, attractiveness, contour, fullness).[4] In the MERZ scales, symmetry and smoothness cannot be graded because they remain constant for all photographs.[9] As such, it could be argued that the Cheeks scale and its respective MERZ scales measure dissimilar constructs, explaining its poor convergent validity.

Overall, this study provides evidence of convergent validity for the majority of FACE-Q Aesthetics scales used in this analysis and may be relevant in an eventual systematic review of the measurement properties of the FACE-Q Aesthetics module, which in turn can be used to make evidence-based recommendations for clinical and research use.[14] The study findings are in line with other studies

investigating the measurement properties of the FACE-Q Aesthetics scales, which show favourable results.[1,6,7] Additionally, this study demonstrates mean FACE-Q scores for its corresponding MERZ Aesthetics photonumeric scale. Ultimately, these results may be used to aid in clinicians' and researchers' interpretation of how these scales relate to one another. In a clinical setting, clinicians may use the study findings to provide a visual estimation of projected improvement on the MERZ Aesthetics photonumeric scale, based on the expected improvement in FACE-Q scores. However, although the 2 instruments show evidence of convergent validity, the MERZ Aesthetics photonumeric scales should not be used to replace the FACE-Q scale and vice versa in clinical research. These tools measure similar, but not identical constructs and each adds important elements to the overall assessment of an aesthetic outcome.

This study has several limitations. First, given that the selected FACE-Q Aesthetics scales include items which evaluate participant perceptions of their appearance (eg, how bothered they are by the area being evaluated) and the MERZ scales use standardized photographs for participants to match their appearance (eg, degree of jawline sagging), it was hypothesized that these scales likely measure related, but dissimilar, constructs for the purpose of hypothesis testing. As such, the authors did not generate any hypotheses about strong correlations that would better support the convergent validity of the FACE-Q scales. Second, the MERZ scales had previously been validated for use by clinicians and expert raters only.[10-12] In our study, MERZ scales were completed directly by the participants based on our assessment of the face validity of the MERZ scales for direct participant use. Especially in an online survey format, differences in picture quality between participants may also have affected the ratings. Third, the sample only included English-speaking participants who were mostly from two countries (United States, United Kingdom), and mostly Caucasian, limiting the generalizability of study findings. Fourth, participants self-selected to enroll in the study and were provided monetary compensation, potentially leading to volunteer bias, and therefore the sample may not be fully representative of the study population. Finally, the data provided by participants were self-reported and therefore could not be independently verified—which may limit the accuracy of the data.

CONCLUSIONS

This study demonstrates evidence of adequate convergent validity of 9 FACE-Q Aesthetics scales (Lines Overall, Lower Face and Jaw, Appraisal of Lines—Forehead/Between Eyebrows/Crow's Feet/Nasolabial Folds/Marionette/Lips, and Lips). Further research should be conducted comparing the FACE-Q Aesthetics scales to other PROMs measuring similar constructs to strengthen the available evidence supporting the construct validity of these scales. Additionally, this study provides estimates of FACE-Q scale scores based on their corresponding MERZ scales that can be used to aid in the interpretation of scores for both instruments.

Supplemental Material

This article contains supplemental material located online at www.aestheticsurgeryjournal.com.

Disclosures

Funding

REFERENCES

1. Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL. Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg*. 2010;26:303-309. doi: 10.1055/s-0030-1262313

2. 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;40:249-260. doi: 10.1016/j.cps.2012.12.001

3. 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;135:375-386. doi: 10.1097/PRS.0000000000000895

4. FACE-Q® AESTHETICS A User's Guide for Researchers and Clinicians. January 2023. Accessed October 16, 2023. https://qportfolio.org/wp-content/uploads/2023/01/FACE-Q-AESTHETICS-USERS-GUIDE.pdf

5. MDDT summary of evidence and basis of qualification decision for FACE-Q Aesthetics. Accessed October 16, 2023. https://www.fda.gov/media/157956/download

6. Gama JT, Rossetto LA, Brito NB, Veiga DF, Ferreira LM. Cross-cultural validation of the FACE-Q satisfaction with facial appearance overall scale (FACE-Q SFAOS) in Brazilian rhytidoplasty patients. *Clinics (Sao Paulo)*. 2020;75:e1568. doi: 10.6061/clinics/2020/e1568

7. Gallo L, Rae C, Kim PJ, et al. Establishing test-retest reliability and the smallest detectable change of FACE-Q aesthetic module scales. *J Plast Reconstr Aesthet Surg*. 2024;95:231-238. doi: 10.1016/j.bjps.2024.06.002

8. Abma IL, Rovers M, van der Wees PJ. Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. *BMC Res Notes*. 2016;9:226. doi: 10.1186/s13104-016-2034-2

9. Stella E, Di Petrillo A. *Standard evaluation of the patient: the Merz scale*. Springer eBooks; 2013: 33-50. doi: 10.1007/978-88-470-5361-8_3

10. Flynn TC, Carruthers A, Carruthers J, et al. Validated assessment scales for the upper face. *Dermatol Surg*. 2012;38:309-319. doi: 10.1111/j.1524-4725.2011.02248.x

11. Carruthers J, Flynn TC, Geister TL, et al. Validated assessment scales for the mid face. *Dermatol Surg*. 2012;38:320-332. doi: 10.1111/j.1524-4725.2011.02251.x

12. Narins RS, Carruthers J, Flynn TC, et al. Validated assessment scales for the lower face. *Dermatol Surg*. 2012;38:333-342. doi: 10.1111/j.1524-4725.2011.02247.x

13. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN study design checklist for patient-reported outcome measurement instruments. 2019. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf

14. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN manual for systematic reviews of PROMs. 2018. https://www.cosmin.nl/wpcontent/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf

15. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42. doi: 10.1016/j.jclinepi.2006.03.012

16. Panchapakesan V, Klassen AF, Cano SJ, Scott AM, Pusic AL. Development and psychometric evaluation of the FACE-Q aging appraisal scale and patient-perceived age visual analog scale. *Aesthet Surg J*. 2013;33:1099-1109. doi: 10.1177/1090820X13510170

17. 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;133:21-30. doi: 10.1097/01.prs.0000436814.11462.94

18. 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;152:443-451. doi: 10.1001/jamadermatol.2016.0018

19. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745. doi: 10.1016/j.jclinepi.2010.02.006

20. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27:1159-1170. doi: 10.1007/s11136-018-1829-0

Supplementary Table 1: Screening Questions Used in Prolific

AESTHETICS TREATMENTS - FACE	
In the PAST 12 MONTHS, have you been to a DERMATOLOGY or a PLASTIC SURGERY CLINIC to have a FACIAL AESTHETIC treatment?	0, No 1, Yes
In the PAST 12 MONTHS, have you had any of these FACIAL AESTHETIC Treatments: Choose all that apply.	0, NONE 1, BOTULINUM TOXIN A - ie, Botox, Dysport, Xeomin or Jeuveau, Xeomin 2, FILLER - eg, Restylane, Juvederm, Radiesse, Sculptra 3, FAT REDUCTION - eg, Kybella to treat a double chin 4, SKIN BOOSTER (eg, Prophilo) (*asked in field-test screen) 5, PLATELET RICH PLASMA (PRP) injections 6, SKIN TIGHTENING with ultrasound - eg, Ultherapy 7, SKIN TIGHTENING with Radio-frequency - eg, Thermage, Morpheus8, Exilis, Profound RF 8, CHEMICAL PEEL 9, MICRODERMABRASION 10, LASER - eg, CO2, Vbeam, Fraxel 11, INTENSE PULSED LIGHT Light (IPL) - eg, Lynton Lumina IPL 12, MICRONEEDLING 13, HYDRAFACIAL 14, THREADLIFT 15, Other
You said you had BOTOX injected. What was the MAIN REASON for having BOTOX?	1, Cosmetic reasons - to look better, younger, refreshed 2, Medical reasons - to treat migraines, to stop grinding teeth 3, Other reason 88, None of the above
You said you a SOFT TISSUE FILLER injected. The last time you had filler, where was the filler injected? Choose all that apply	1, Cheeks - to add volume and restore fullness 2, Lips - to plump or to smooth out lip lines 3, Other

Supplementary Table 2: Hypothesis Testing of FACE-Q And Merz Aesthetic Scales With Spearman's Rank Correlation Coefficient

Legend:
- = correct (B) hypothesis
- = incorrect (B) hypothesis
- = correct (C) hypothesis
- = incorrect (C) hypothesis

Hypotheses:
(A): correlations with instruments measuring similar constructs should be ≥0.50;
(B): correlations with instruments measuring related, but dissimilar constructs should be lower i.e. 0.30-0.50;
(C): correlations with instruments measuring unrelated constructs should be <0.30.

FACE-Q Aesthetic Scales	MERZ: Jawline – At Rest	MERZ: Forehead Lines - Dynamic	MERZ: Glabella Lines - Dynamic	MERZ: Crow's Feet - Dynamic	MERZ: Nasolabial Folds - At Rest	MERZ: Marionette Lines - At Rest	MERZ: Lip Wrinkles - At Rest	MERZ: Upper Lip Fullness - At Rest	MERZ: Lower Lip Fullness - At Rest	MERZ: Upper Cheek Fullness - At Rest	MERZ: Lower Cheek Fullness - At Rest	MERZ: Oral Commissure at Rest
Face overall	(B) -0.296	(B) -0.242	(B) -0.248	(B) -0.247	(B) -0.266	(B) -0.267	(B) -0.205	(B) 0.142	(B) 0.119	(B) -0.199	(B) -0.161	(B) -0.274
Lines overall	(B) **-0.372***	(B) **-0.344***	(B) **-0.306***	(B) **-0.434***	(B) **-0.392***	(B) **-0.410***	(B) **-0.410***	(C) 0.127	(C) 0.134	(C) -0.293	(C) -0.299	(B) **-0.339***
Lower face and jaw	(B) **-0.303***	(C) -0.130	(C) -0.144	(C) -0.133	(C) -0.225	(C) -0.225	(C) -0.133	(C) 0.102	(C) 0.056	(C) -0.150	(C) -0.107	(C) -0.235
Appraisal of lines: forehead	(C) -0.197	(B) **-0.347***	(C) -0.253	(C) -0.287	(C) -0.257	(C) -0.225	(C) -0.263	(C) 0.067	(C) 0.071	(C) -0.212	(C) -0.188	(C) -0.168
Appraisal of lines: between eyebrows	(C) -0.276	(C) -0.196	(B) **-0.366**	(C) -0.281	(C) -0.246	(C) **-0.326***	(C) **-0.322***	(C) 0.076	(C) 0.089	(C) -0.199	(C) -0.220	(C) -0.287
Appraisal of lines: crow's feet	(C) -0.194	(C) -0.197	(C) -0.133	(B) **-0.396***	(C) -0.205	(C) -0.199	(C) **-0.300***	(C) 0.058	(C) 0.086	(C) -0.224	(C) -0.219	(C) -0.210
Appraisal of lines: nasolabial folds	(C) -0.296	(C) -0.109	(C) -0.126	(C) -0.177	(B) **-0.356***	(C) **-0.336***	(C) -0.268	(C) 0.069	(C) 0.081	(C) -0.201	(C) -0.253	(C) **-0.345***
Appraisal of lines: marionette	(C) **-0.340***	(C) -0.108	(C) -0.156	(C) -0.189	(C) -0.285	(B) **-0.430***	(C) **-0.329***	(C) 0.091	(C) 0.046	(C) -0.185	(C) -0.237	(B) **-0.418***
Appraisal of lines: lips	(C) -0.216	(C) -0.135	(C) -0.175	(C) -0.200	(C) -0.188	(C) -0.256	(B) **-0.426***	(C) 0.130	(C) 0.104	(C) -0.155	(C) -0.231	(C) -0.271
Lips	(C) -0.257	(C) -0.167	(C) -0.237	(C) -0.236	(C) -0.241	(C) -0.265	(C) -0.254	(B) **0.451***	(B) **0.370***	(C) -0.196	(C) -0.217	(B) **-0.320***
Cheeks	(C) -0.248	(C) -0.169	(C) -0.149	(C) -0.178	(C) -0.228	(C) -0.206	(C) -0.194	(C) 0.077	(C) 0.096	(B) -0.277	(B) -0.183	(C) -0.214