Community Dentistry and Oral Epidemiology

WILEY

Community Dentistry and Oral Epidemiology

ORIGINAL ARTICLE OPEN ACCESS

Psychometric Validation of the FACE-Q Dental Module in Patients With Malocclusions

Ali Tassi[1] | Anne F. Klassen[2] | Jessica Li[3] | Karen W. Y. Wong Riff[4] | Charlene Rae[2]

[1]Graduate Orthodontics and Dentofacial Orthopaedics Program, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada | [2]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada | [3]Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada | [4]Division of Plastic and Reconstructive Surgery, Department of Surgery, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

Correspondence: Anne F. Klassen (aklass@mcmaster.ca)

Received: 25 April 2025 | **Revised:** 25 April 2025 | **Accepted:** 12 May 2025

Funding: The research described in this study was supported by a grant from the Canadian Institute of Health Research (FRN 148779).

Keywords: children | malocclusion | oral health | orthodontics | patient reported outcome measures | quality of life | validation study | young adults

ABSTRACT

Objective: The FACE-Q Craniofacial Module is a patient-reported outcome measure (PROM) developed for children and young adults with craniofacial conditions. We hypothesised that some of its scales may be applicable to other populations. The aim of this study was to assess the validity and reliability of FACE-Q scales for patients with dental malocclusions.

Methods: The FACE-Q Dental Module includes 5 scales from the Craniofacial Module that measure appearance (Face, Jaws, Smile and Teeth) and function (Eating/Drinking). Data were collected from patients aged 8–29 years who presented with a dental malocclusion (pre-treatment) or 1–2 years after orthodontic treatment (post-treatment) at a large university-based orthodontic specialty clinic in Canada between September 2018 and March 2020. Patients completed a paper questionnaire booklet, and data were entered into a Research Electronic Data Capture (REDCap) survey. The psychometric analysis was performed using Rasch Measurement Theory (RMT) analysis.

Results: The sample of 434 patients was aged 9 to 29 years, with 249 female and 185 male participants. The sample included 252 pre-treatment and 182 post-treatment patients. The 4 appearance scales evidenced strong psychometric performance; all 37 items had ordered thresholds with good item fit to the Rasch model. Reliability was high, with person separation index and Cronbach alpha values, with and without extremes ≥ 0.86. As hypothesised, those participants who had a major difference in appearance, and those who reported liking their appearance less, scored lower on the appearance scales ($p < 0.001$). In the RMT analysis, the Eating/Drinking scale evidenced low reliability and poor targeting with close to 40% of particpants scoring at the ceiling.

Conclusion: The FACE-Q Dental Module provides a means to collect evidence-based outcomes data from children and young adults who undergo orthodontic care for dental malocclusions.

1 | Introduction

Malocclusions are common and encompass a wide range of jaw and dental alignment discrepancies perceived as aesthetically and functionally poor [1]. Major concerns reported by patients with malocclusions include alignment of teeth, presence of increased overjet, and factors impacting their quality of life (e.g., feeling embarrassed or sad about one's teeth) [2]. With orthodontic treatment, many patients report improved dental and facial aesthetics, better self-image, and greater confidence to eat and smile in front of other people [2, 3]. A strong association between aesthetics and psycho-social outcomes has

also been noted in orthodontic patients [3]. This association is especially notable for adolescents due to social media use, where the appearance of their teeth and smile in images is of particular importance. Given the significance of aesthetics in the treatment of malocclusions, having a measure of dental appearance would be informative for both clinical practice and research. Appearance is a construct that is best measured from the patient perspective due to its subjective nature, and this could be accomplished using a patient-reported outcome measure (PROM) [4, 5].

To address the lack of PROMs for children and young adults (8 to 29 years of age) with cleft and non-cleft craniofacial conditions, our team developed the CLEFT-Q [6] and FACE-Q Craniofacial Module (CM) [7–9]. These PROMs measure 4 domains: appearance, function, health-related quality of life (HRQL) and adverse effects of treatment [7–9]. Scales were developed and tested in patients with any type of congenital or acquired craniofacial conditions. Psychometric analysis on a combined sample of 4743 participants provided evidence of validity and reliability of PROMs [8]. We hypothesized that 4 appearance scales (Face, Jaws, Smile, Teeth) and 1 function scale (Eating/Drinking) from these PROMs may also be applicable to patients with more common orthodontic needs, including those seen in orthodontic practices in the community. In order to assess whether these scales had content validity (i.e., were relevant, comprehendible, and comprehensive) in a general orthodontic population, qualitative research methods were used to cognitively assess this sub-set of scales based on best practice guidelines for PROM development [4]. This process involved cognitive debriefing interviews with 15 patients who had a wide range of malocclusion traits and orthodontic treatment plans, as well as input from 21 clinical experts, which confirmed the scales had content validity within this population [10]. Given that the scales had content validity, 153 patients who had orthodontic treatment were recruited as part of the FACE-Q CM field test [8, 9]. This field test study validated CLEFT-Q scales in a non-cleft craniofacial sample [8], and finalised new scales for the broader population of patients with facial differences [9].

Many PROMs have been developed to examine outcomes in oral disease [11–14], with only a few specifically for use in the orthodontic population [13, 14]. A comparison of commonly used PROMs in dentistry is provided in Table S1. The Malocclusion Impact Questionnaire (MIQ) is the only commonly used PROM in orthodontics that was developed with a modern psychometric approach [15]. An important limitation of traditional psychometric methods is that these scales do not produce interval level measurement and instead provide ordered counts [16]. The use of parametric statistics is based on the assumption of having normally distributed interval or ratio level data. The FACE-Q CM differs from PROMs described in Table S1 in a variety of ways. The FACE-Q CM was developed and validated in a broad sample of patients with craniofacial conditions as well as patients with malocclusions seen in hospital and community-based orthodontic clinics [8–10]. The FACE-Q CM sample included a wide age range (children, adolescents and young adults) of patients who are often treated together in the same clinics, and who may require multiple phases of treatment throughout their development [8, 9]. Items selected in the scales were cognitively tested and found to be relevant across this wide age range. Having a single tool that can be used in children and young adults can facilitate use in a diverse clinical practice. Also, the FACE-Q CM has scales that focus specifically on aspects of the appearance of the face, jaws, smile and teeth, all important aesthetic features that orthodontic patients are concerned about. Many commonly used PROMs measure quality of life outcomes and general oral health rather than appearance-related malocclusion traits.

Although the development of the FACE-Q CM included a general orthodontic sample, psychometric evidence was not reported for this group separately [8]. Psychometric evidence for a target population helps to inform users when selecting a PROM for clinical or research use. Evidence for the content validity of a sub-set of scales has already been established in patients with malocclusions [10], however, other psychometric properties (e.g., construct validity, reliability) have yet to be reported. This study expands on previous development work of the FACE-Q CM [8, 9] to further assess the psychometric properties of a subset of scales (Face, Jaws, Smile, Teeth, Eating/Drinking) in a large sample of orthodontic patients with malocclusions who attended a university orthodontic clinic.

2 | Methods

2.1 | Ethics Statement

Ethics Approval Was Obtained From the Health Sciences Research Ethics Board (HSREB) at the University of Western Ontario (#108129) and the Hamilton Integrated Research Ethics Board at McMaster University.

2.2 | Measure

2.2.1 | Face-Q CM [8, 9]

The scales assessed in this study were developed for patients 8 to 29 years of age with a facial difference [8, 9]. This age range was chosen so the scales could be applied in children and young adults, allowing for longitudinal analyses across this age span. Having a tool that is valid across a large age range can facilitate implementation in treatment centres that include children and young adults. All the FACE-Q CM scales are interval-level (i.e., can be utilised in parametric statistics) and are scored on a scale of 0 to 100, with a higher score indicating better outcomes. Scores are derived by summing the response values for the items within the scale. The scoring guide for the FACE-Q CM can then be used to look up Rasch-transformed scores for the corresponding summed values for each scale. Scale development was informed by best practice guidelines published by the Food and Drug administration (FDA) [4], International Society for Quality of Life Research [5], and the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) [17, 18]. Five scales relevant to the orthodontic population were tested, forming the FACE-Q Dental Module, including 4 appearance scales (Face, Jaw, Smile, Teeth) and a function scale (Eating/Drinking). The appearance scales ask about how much a patient likes how parts of their face or their smile looks now (Not at all, A little bit, Quite a bit, Very much). The Eating/Drinking scale asks about the frequency of functional problems

of the mouth within the past week (Always, Often, Sometimes, Never).

2.3 | Data Collection

Data were collected between September 2018 and March 2020 from patients seen in the Graduate Orthodontic Clinic at the Schulich School of Medicine and Dentistry, Western University (Canada). When selecting the sample, it is important to ensure it is representative of the group the tool is intended to be used in [18]. For the purposes of this study, the context of use for the PROM was patients seeking orthodontic care for malocclusions encountered in a general orthodontic setting. Inclusion criteria: patients aged between 8 and 29 years, fluent in English, with any type or severity of malocclusion, and before starting (pre-treatment) or recently completed (1–2 years post-treatment) orthodontic treatment. A diverse sample was obtained in order to determine if the FACE-Q Dental Module was psychometrically valid for use in the general orthodontic population. A member of the staff at the orthodontic clinic introduced eligible participants to the study. Patients (and their guardians) interested in participating were invited to complete an informed assent/consent letter and study questionnaire booklet. Patients were asked to complete the survey on their own.

The survey asked patients their age and gender and if they previously underwent orthodontic treatment (braces or aligners), followed by the FACE-Q Dental Module scales. A clinical form was used by the site orthodontist to collect patient information including age, gender, type of dentition and phase of treatment (pre-treatment, post-treatment). After reviewing the clinical data, a severity rating (i.e., No, Yes—minor, Yes—major) for appearance of the face, jaw, smile and teeth, as well as ability to eat and drink, was determined by a single on-site faculty orthodontist (AT). This was based on clinical expertise and commonly used assessments and indices for soft-tissue facial and smile analysis, orthodontic diagnosis, orthodontic treatment need, and case complexity [1]. Data were collected in this manner to be consistent with the methodology of the original field test, whereby the site recruiter would provide a severity rating for a patient based on their experience and expertise [8, 9]. More specific clinical data were also collected by the site orthodontist from the patients records, including dental crowding and spacing, diastemas, midline discrepancies, molar occlusion, overjet, overbite, openbite, crossbites, and dental anomalies (missing, impacted or supernumerary teeth). These data were collected to provide a comprehensive description of the sample involved in the validation of the scales. Data were entered into a REDCap database hosted at McMaster University (Canada) [19].

2.4 | Statistical Analysis

2.4.1 | Rasch Measurement Theory Analysis

Rasch Measurement Theory (RMT) was used to determine if data fit the Rasch model [20]. Data were analysed in RUMM2030 (RUMM version 2030, RUMM Laboratory Pty Ltd., Duncraig, Western Australia, 1998–2020) using the polytomous partial credit model [16]. Analysis details are provided in Table 1.

2.4.2 | Construct Validation

In SPSS v26.0 (IBM Corporation, Armonk NY, USA), using the scoring algorithm for each scale, raw scores were transformed into scores that ranged from 0 (worse) to 100 (best). Scale missing data were imputed using the mean, if at least 50% of items were completed. Data normality was assessed using Kurtosis (absolute > 2) and Skewness (absolute > 2) values [30]. For values over 2, scale scores were analysed using non-parametric statistics. Pre-defined hypotheses for construct validation are provided in Table 2.

3 | Results

3.1 | RMT Analysis

Table 3 provides detailed characteristics of the 434 participants. The sample ranged in age from 9 to 29 years (Mean = 15.9; SD = 3.3). A wide range of diagnoses, malocclusion traits, and orthodontic treatment plans were represented in the sample.

Table 4 shows the scale level RMT analysis results alongside the previously published field-test sample scale results [8, 9]. The analysis provided evidence of the reliability and validity of the 4 appearance scales. The Eating/Drinking scale, on the other hand, evidenced low reliability, with PSI values of 0.33 and 0.39 and Cronbach alpha values of 0.75 and 0.56, with and without extremes, respectively. The scale was not adequately targeted to the clinic-based orthodontic sample, with close to 40% of the sample not scoring on the scale. These findings differed from the FACE-Q field-test sample where reliability values were ≥ 0.77 and more than 70% of participants scored on scale [8, 9].

Table S2 shows the item level RMT results. For the appearance scales, thresholds were ordered for items; one item had significant *p*-value after Bonferroni adjustment, and item fit was outside of ±2.5 for 12 items. Reliability for the appearance scales was high, with person separation index (PSI) and Cronbach alpha values with and without extremes over ≥ 0.86. The residual correlations for a pair of items in the Face scale were 0.46. When a subtest was performed, the PSI values dropped by 0.01, indicating marginal impact on scale reliability. Altogether, 6 items evidenced DIF: 4 and 2 items in the Face scale for age and gender, respectively, and 1 item each in the Teeth and Smile scales for age. Pearson correlations between person locations for items before and after item split for DIF indicated marginal impact, with all correlations ≥ 0.998. Data fit the Rasch model for 2 scales, with marginal misfit for 2 scales.

3.2 | Construct Validation

Detailed results for hypothesis testing are in Tables S3–S11. Acceptance of the hypotheses ranged from 80% for the Jaws scale to 100% for the Face scale (Table 2). As expected, scores on relevant scales increased as participants reported liking an aspect of their face or smile more ($p < 0.001$). Scores for severity by facial part were lowest for major differences compared to the minor and no difference groups ($p < 0.001$). Participants who were post-treatment scored higher on appearance scales compared

TABLE 1 | Summary of psychometric analysis and findings.

Property	Description	Study findings & Importance
Target population	The population in which the PROM is to be used	This instrument is for individuals 8–29 years of age with any type of malocclusion
Content validity	The items, response options and instructions are relevant, comprehendible, and comprehensive to the target population	Previously demonstrated, in Tan et al. study [10], in sample of orthodontic patients 8 to 29 years of age that the scales were relevant, comprehensive, comprehendible. This shows that the items, response options and instructions for this set of constructs were meaningful and applicable to this age range
Psychometrics		
Rasch analysis	*Thresholds*: thresholds for item response options (e.g., Not at all, A little, Quite a bit, Very much) were examined to determine if they were ordered as intended [21] *Fit indicators*—three fit indicators log residuals (item-person interactions), chi-square values (item-trait interactions), and item characteristic curves [24] were examined to determine if the items of each scale worked as hypothesized to map out a clinical hierarchy for the concept being measured *Local dependency* was examined to determine if items in each scale were independent of each other, since correlations between residuals can artificially inflate reliability [25–27]. Residuals that correlated ≥0.30 were identified and a subtest was performed to determine the impact of any such correlations on scale reliability *Targeting*: to determine if person and item locations were evenly spread over a good range that overlapped. This study also computed the proportion of the sample to score on scale	Results showed that the data for the scales fit the Rasch model, meaning that the items can be summed to generate a score for the construct of interest [22]. It was also found that the response options had ordered thresholds. This result creates a probabilistic Guttman pattern that reflects a hierarchy that is clinically sensible, where endorsing an item lower on the scale positively should be harder (see Figure S1). For example in a scale measuring mobility, we would not expect people to be more likely to respond 'very difficult' to an item about the ability to stand up, and 'not at all difficult' to an item about running a mile. Local dependency results demonstrate that the items selected for a scale function independently of each other. When item residuals are correlated, the scale appears more reliable, artificially [16, 22]. None of the correlated items within scales were found to have an impact on scale reliability. If a scale is well targeted the outcome in the sample will overlap with the range of concepts that are measured by the set items. A well targeted scale will have a mean person location close to the center of the items [23]. The targeting of the scales is shown in Figure S1. For a scale to have acceptable targeting the pink bars, representing distribution of person scores in the sample, should cover the blue bars that represent the items in the scales. All the scales in this study, except Eating/Drinking, had acceptable targeting
Measurement invariance	*Differential Item Functioning* (*DIF*) was examined to identify if any items in a scale worked differently by age group (8–15 years vs. 16–29 years) and gender (male vs. female). For DIF analysis, since the size of the sample by gender differed, 3 random samples were chosen and the DIF analysis was repeated to see if the results were stable. Chi-square values that were significant after Bonferroni adjustment were used to identify items with potential DIF. Adjustments for DIF were made by splitting the sample based on the characteristic and then correlating the original and new person locations to determine the impact of DIF on scoring [24]	DIF was observed for both gender and age for some items (Table S3). However, upon further examination the observed DIF did not impact the scoring of the scales. This supports that the scales function the same in both the age and gender sub-groups examined [16, 22]. This finding does not mean that these groups experience the same outcome or have similar scores, just that the items work the same way within the sub-groups.

(Continues)

TABLE 1 | (Continued)

Property	Description	Study findings & Importance
Internal consistency	Reliability was examined in terms of *Person Separation Index* (*PSI*), which examines the spread of people on the scale, and *Cronbach alpha*, which measures how closely related a set of items are as a group. Higher scores indicate greater reliability, with values > 0.70 considered satisfactory [28, 29]	For the appearance scales the PSI values were all ≥ 0.86. This result provides evidence that the items within the scales discriminate between participants [16]. Each scale, except for Eating \ Drinking, had a Cronbach's alpha of at least 0.70 indicating sufficient evidence for internal consistency in this target population. This finding means that the items in the scales work together to measure the same construct.
Construct validity	Demonstrates how well the PROM represents the construct it intends to measure. It is tested by using predefined hypotheses of known group differences. Three sets of hypotheses were tested. The first set represented Likert scales of the overall self-reported construct, which provided obvious group differences that should increase incrementally. The second set were based on clinical variables and represent more difficult differences to detect. Finally, correlation hypotheses were tested to demonstrate relationships between the scales, where it was expected that correlations between the scales measuring similar constructs would be > 0.50, for related but dissimilar constructs would be 0.30–0.50, and for unrelated constructs would be < 0.30 [18]. As per COSMIN criteria we deemed construct validity was 'sufficient' for a scale if at least 75% of hypotheses were met [17]	All scales met the 75% threshold for acceptance of hypotheses specified by COSMIN for sufficient evidence of construct validity (Table 2) [18]. This finding provides evidence that the scales measure the constructs as intended. Validation is an ongoing process; further work can be done looking at known differences in clinical groups or comparing to other measures that capture similar or related constructs.

TABLE 2 | Pre-defined hypotheses by scale for construct validation assessment.

Predicted hypotheses	Face	Jaws	Smile	Teeth
As participants self-report liking their face more scores will...	Increase**	Increase**	Increase**	Increase**
As participants self-report liking their smile more scores will...	Increase**	NA	Increase**	Increase**
As participants self-report liking their jaw more scores will...	Increase**	Increase**	NA	NA
As participants self-report liking their teeth more scores will...	Increase**	NA	Increase**	Increase**
Participants who are post treatment will score...	Higher**	NA	Higher**	Higher**
Participants who have a major teeth difference will score...	Lower**	NA	Lower**	Lower**
Participants who have a major smile difference will score...	Lower**	NA	Lower**	Lower**
Participants who have a major jaw difference will score...	Lower**	Lower**	NA	NA
Scores on the Face scale will correlate with...	NA	Moderately(NM)**	Moderately(NM)**	Moderately(NM)**
Scores on the Teeth scale will correlate with...	NA	Moderately**	Highly**	NA
Scores on the Jaws scale will correlate with...	NA	NA	Moderately**	NA
Proportion of hypotheses accepted	8/8	4/5	8/9	6/7

Abbreviations: NA, not applicable; NM, hypothesis not met (strength of correlation was greater than expected.
**$p < 0.001$.

with pre-treatment participants (Figure 1; $p \leq 0.002$). The magnitude of the difference between pre- and post-treatment groups was greatest for the Teeth and Smile scales. In terms of convergent validity, most hypothesised relationships met the COSMIN criteria (Table 2 and Table S11).

4 | Discussion

This study provides evidence for the reliability and validity of the FACE-Q Dental Module in a large community-based sample of children and young adults with malocclusion. The FACE-Q Dental Module scales ask specific questions of the patients about their perception of particular malocclusion traits. The content of the scales is highly relevant in orthodontics, as diagnosis is typically presented as a problem list of these traits [1]. Furthermore, the FACE-Q Dental Module does not just focus on teeth, but also measures aesthetics of the smile, jaws, and face overall.

The psychometric findings for the 4 appearance scales examined in this study adhere to COSMIN guidelines. An interpretation of the findings is provided in Table 1. The scales in the Dental Module overcome limitations of currently available scales used in orthodontics, which were mainly developed using methods based on Classical Test Theory (CTT) (Table S1). CTT differs from modern psychometrics because it focuses solely on the total score and its reliability. The CTT approach creates an overall quantification (e.g., sum, average) that summarises the items. Whereas modern psychometrics uses a probabilistic approach that examines the response to an item and its relation to the amount of the construct being measured [31]. Limitations of CTT have been summarised by Cano and Hobart [16, 22]. The main advantage of the FACE-Q Dental Module is that the scale scores have interval-level properties, making them more suitable as outcome measures compared with ordinal-level rating scales. Also, scales developed using RMT methods can be applied to individual patient measurement, which is a limitation of CTT where the results are only applicable at the group-level [16, 22]. The FACE-Q Dental Module can be used in a wider age range than the MIQ, which is for children aged 10 to 16 year olds [15]. The MIQ differs from the FACE-Q Dental Module, as it measures constructs focusing more on the "impact of malocclusion on the daily lives of young people" whereas the FACE-Q Dental Module focuses mainly on appearance concerns.

This study found that the Eating/Drinking scale evidenced low reliability and was not well targeted to the sample. Many of the participants in the sample scored at the ceiling, suggesting eating or drinking issues were not an important concern for them. This finding aligns with the fact that non-ideal occlusions (underbites, overbites, etc.), which are common reasons for seeking treatment for malocclusion, do not always affect the ability to eat or drink in the same way a craniofacial condition might [32]. Functional issues measured by the Eating/Drinking scale, such as food falling out or difficulty using a straw, are likely not applicable to the average orthodontic patient without a craniofacial

TABLE 3 | Demographic and clinical characteristics.

Characteristic	Category	$N=434$	%
Age	8–12	67	15.4
	13–17	245	56.5
	18–29	122	28.1
Gender	Female	249	47.4
	Male	185	42.6
Treatment stage	Pre-treatment	252	58.1
	Post-treatment	182	41.9
Malocclusion traits[a]	**Category**	$N=252$	%
Dentition	Mixed	61	24.2
	Permanent	191	75.8
Upper crowding or spacing	None	20	7.9
	Crowding	173	68.7
	Spacing	59	23.4
Lower crowding or spacing	None	31	12.3
	Crowding	179	71.0
	Spacing	42	16.7
Maxillary midline diastema	Not present	191	75.8
	Present	61	24.2
Maxillary midline to facial	On	112	44.4
	Off	140	55.6
Mandibular midline to facial	On	83	32.9
	Off	169	67.1
Right molar occlusion	Class I	118	46.8
	Class II	99	39.3
	Class III	35	13.9
Left molar occlusion	Class I	117	46.4
	Class II	98	38.9
	Class III	37	14.7
Overjet	Normal	124	49.2
	Excess	97	38.5
	Reverse	31	12.3
Overbite	Normal	91	36.1
	Deep bite	128	50.8
	Open bite	33	13.1
Lateral openbite	Yes	25	9.9
Anterior crossbite	Yes	75	29.8
Posterior crossbite	Yes	76	30.2
Buccal crossbite	Yes	13	5.2
Missing teeth	Yes	32	12.7
Impacted teeth	Yes	35	13.9
Supernumerary teeth	Yes	6	2.4

[a]Pre-treatment group only.

condition. And if they have occurred in the past, they may have successfully adjusted or may no longer occur frequently enough to report. For this scale, the findings differed from the FACE-Q CM field-test study where 74% of participants scored on the scale. The results suggest that the Eating/Drinking scale may not be applicable to an orthodontic population with less severe malocclusions [8].

PROMs are useful tools that can be incorporated into both research and clinical practice to help inform orthodontic care from the patient perspective. Within the clinic, orthodontists could use the FACE-Q Dental Module to aid in shared decision-making [33]. Along with traditional clinical measures, PROMs can add valuable information that can help better understand both the patient's values and perceptions [33]. For example, by asking patients about specific malocclusion traits, the clinician is able to understand where the patient perceives the problem to be. This information could be incorporated into care to help manage patient expectations and needs, as well as to create personalised treatment plans, and help facilitate the informed consent process [34]. Results from PROMs can also be used to monitor progress throughout the course of treatment by tracking changes in scores [35]. The FACE-Q Dental Module can also add standardised appearance-related outcomes to orthodontic research that are applicable across a large severity of malocclusions and age ranges, reducing the need for multiple measures. Decreasing heterogeneity of measures used in clinical research would help to facilitate systematic reviews that compile evidence across studies.

This study has some limitations. First, the sample was recruited from a single university-based orthodontic specialty clinic in Canada. Second, some aspects of the COSMIN criteria were not assessed, including test–retest reliability and responsiveness [17]. Third, minor vs. major ratings were made by the site orthodontist using clinical expertise guided by clinical metrics and diagnostic indices. Further work should assess additional psychometric properties of the scales in this population and include associations with the standardised clinical measurements. This study also did not assess the construct validity of the HRQL scales from the FACE-Q CM in children and young adults with malocclusions. Although these scales were not included in the phase one study of content validity for the FACE-Q Dental Module [10], they were previously tested extensively in cognitive interviews with a broad range of patients with craniofacial conditions, many of whom need or have had orthodontic treatment [8, 9]. Additional research could further assess the content validity and psychometric properties of these additional scales in patients with malocclusions.

TABLE 4 | Rasch Measurement Theory scale level statistics.

Scale		*N* to complete scale	*N* in RMT	% scored on scale	c^2	DF	*p*	PSI+	PSI-	α+	α-
Face	Dental	433	420	97.0	74.8	54	0.03	0.87	0.86	0.88	0.88
	Field test [8]	4159	3777	90.8	48.9	72	0.98	0.87	0.87	0.92	0.89
Jaws	Dental	408	327	80.1	30.7	28	0.33	0.90	0.88	0.95	0.91
	Field test [8]	1999	1480	74.0	24.5	56	0.99	0.91	0.89	0.96	0.92
Smile	Dental	431	396	91.9	66.5	54	0.12	0.91	0.90	0.94	0.92
	Field test [9]	497	442	88.9	70.6	45	0.01	0.91	0.89	0.94	0.91
Teeth	Dental	429	406	94.6	110.6	60	< 0.01	0.93	0.93	0.96	0.95
	Field test [8]	3022	2684	88.8	54.1	96	0.99	0.86	0.85	0.95	0.93
Eating \ Drinking	Dental	430	260	60.5	29.9	18	0.04	0.33	0.39	0.75	0.56
	Field test [8]	391	290	74.2	20.1	27	0.83	0.77	0.80	0.91	0.86

Abbreviations: α ± extr, Cronbach alpha with and without extremes; c^2, chi square; DF, degrees of freedom; PSI ± extr, Person Separation Index with and without extremes.

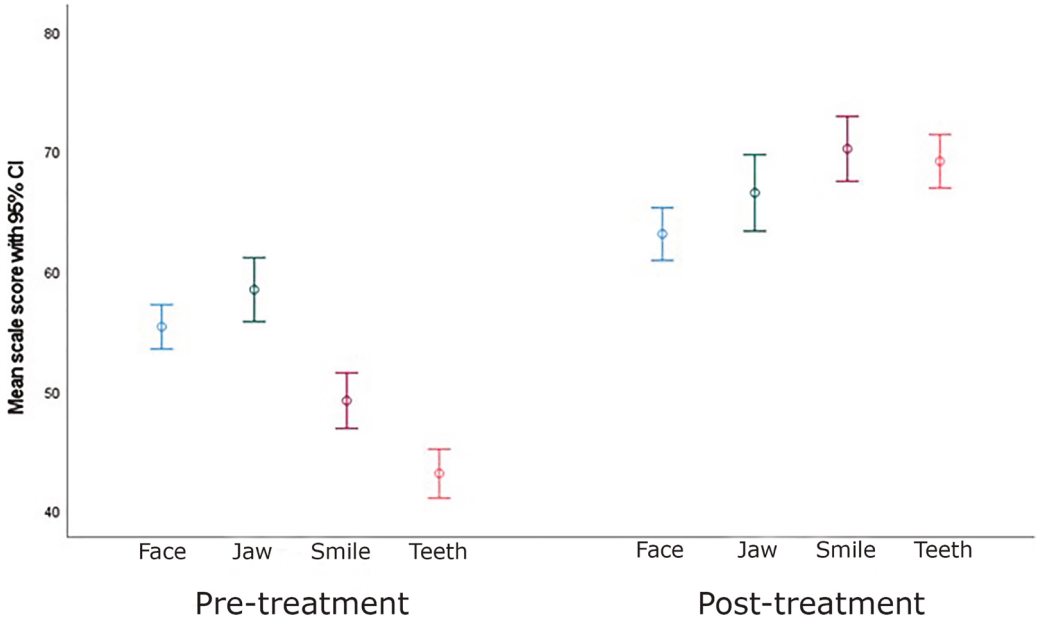


FIGURE 1 | Mean scores and 95% confidence intervals (CI) for 4 Appearance scales Pre and Post Treatment.

5 | Conclusions

The FACE-Q Dental Module scales provide practitioners and researchers with a means to measure the appearance of the teeth, smile, jaws, and face from the patient perspective in orthodontics. Since each scale is independently functioning, one or more scales can be used. The scales evidenced reliability and validity within the general orthodontic population, and further research will help optimise the FACE-Q Dental Module for use in clinical settings. These scales can be used in research studies and orthodontic clinics to measure and evaluate treatment need, severity, and outcomes from the perspective of patients aged 8 to 29 years. Further information about the FACE-Q Dental Module can be found at https://qportfolio.org/face-q/.

Acknowledgements

The research described in this study was supported by a grant from the Canadian Institute of Health Research (FRN 148779). We are grateful to CIHR for funding our research as well as to the many healthcare professionals and research staff in craniofacial sites around the world for their dedication to this project.

Ethics Statement

Ethics approval was obtained from the Health Sciences Research Ethics Board (HSREB) at the University of Western Ontario (approval number 108129) and the Hamilton Integrated Research Ethics Board at McMaster University.

Consent

Written consent/assent was obtained from participants. Parents or guardians also provided written consent where required. This study was conducted in full accordance with ethical principles, including the World Medical Association Declaration of Helsinki (version 2008).

Conflicts of Interest

Anne Klassen and Karen Wong are co-developers of the patient-reported outcome scales described in this publication and share in any licence revenues as royalties based on their institutions' inventor sharing policy for their use in for-profit study. The other authors have no conflicts of interest to declare in relation to this work.

Data Availability Statement

Participants included as part of this study did not provide written consent to allow their data to be shared publicly; therefore, for this ethical reason, supporting data is not available.

References

1. W. R. Proffit, H. W. Fields, and D. M. Sarver, *Contemporary Orthodontics*, 5th ed. (Elsevier Mosby, 2013).

2. E. Bradley, A. Shelton, T. Hodge, et al., "Patient-Reported Experience and Outcomes From Orthodontic Treatment," *Journal of Orthodontics* 47, no. 2 (2020): 107–115.

3. N. Al Quraini, R. Shah, and S. J. Cunningham, "Perceptions of Outcomes of Orthodontic Treatment in Adolescent Patients: A Qualitative Study," *European Journal of Orthodontics* 41, no. 3 (2019): 294–300.

4. U.S. Department of Health and Human Services Food and Drug Administration, "Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims," 2009, http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf.

5. D. L. Patrick, L. B. Burke, C. J. Gwaltney, et al., "Content Validity–Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1–Eliciting Concepts for a New PRO Instrument," *Value in Health* 14, no. 8 (2011): 967–977, https://doi.org/10.1016/j.jval.2011.06.014.

6. A. F. Klassen, K. W. Riff, N. M. Longmire, et al., "Psychometric Findings and Normative Values for the CLEFT-Q Based on 2434 Children and Young Adult Patients With Cleft Lip and/or Palate From 12 Countries," *Canadian Medical Association Journal* 190, no. 15 (2018): E455–E462.

7. N. M. Longmire, K. W. Y. W. Riff, J. L. O'Hara, et al., "Development of a New Module of the FACE-Q for Children and Young Adults With Diverse Conditions Associated With Visible and/or Functional Facial Differences," *Facial Plastic Surgery* 33, no. 5 (2017): 499–508.

8. A. F. Klassen, C. Rae, K. W. Riff, et al., "FACE-Q Craniofacial Module: Part 1 Validation of CLEFT-Q Scales for Use in Children and Young Adults With Facial Conditions," *Journal of Plastic, Reconstructive & Aesthetic Surgery* 74, no. 9 (2021): 2319–2329.

9. A. F. Klassen, C. Rae, W. Riff, et al., "FACE-Q Craniofacial Module: Part 2 Psychometric Properties of Newly Developed Scales for Children and Young Adults With Facial Conditions," *Journal of Plastic, Reconstructive & Aesthetic Surgery* 74, no. 9 (2021): 2330–2340.

10. A. Tassi, J. Tan, B. Piplani, N. Longmire, K. W. Y. Wong Riff, and A. F. Klassen, "Establishing Content Validity of an Orthodontic Subset of the FACE-Q Craniofacial Module in Children and Young Adults With Malocclusion," *Orthodontics & Craniofacial Research* 24, no. 4 (2021): 553–560.

11. C. Zaror, Y. Pardo, G. Espinoza-Espinoza, et al., "Assessing Oral Health-Related Quality of Life in Children and Adolescents: A Systematic Review and Standardized Comparison of Available Instruments," *Clinical Oral Investigations* 23 (2019): 65–79.

12. X. Huang, Z. Tao, P. Ngan, D. Qin, H. He, and F. Hua, "The Use of Dental Patient-Reported Outcomes Among Comparative Observational Studies in Orthodontics: A Methodological Study," *Journal of Evidence-Based Dental Practice* 24, no. 1 (2024): 101956.

13. Z. Tao, T. Zhao, P. Ngan, D. Qin, F. Hua, and H. He, "The Use of Dental Patient-Reported Outcomes Among Randomized Controlled Trials in Orthodontics: A Methodological Study," *Journal of Evidence-Based Dental Practice* 23, no. 1 (2023): 101795.

14. E. Peter, R. M. Baiju, J. M. Varughese, and N. O. Varghese, "Patient-Reported Outcome Measures in Orthodontics," *Dentistry and Medical Research* 7, no. 1 (2019): 3–11.

15. P. E. Benson, S. J. Cunningham, N. Shah, et al., "Development of the Malocclusion Impact Questionnaire (MIQ) to Measure the Oral Health-Related Quality of Life of Young People With Malocclusion: Part 2–Cross-Sectional Validation," *Journal of Orthodontics* 43, no. 1 (2016): 14–23.

16. J. Hobart and S. Cano, "Improving the Evaluation of Therapeutic Interventions in Multiple Sclerosis: The Role of New Psychometric Methods," *Health Technology Assessment* 13 (2009): 1–177.

17. L. B. Mokkink, C. B. Terwee, D. L. Patrick, et al., "The COSMIN Checklist for Assessing the Methodological Quality of Studies on Measurement Properties of Health Status Measurement Instruments: An International Delphi Study," *Quality of Life Research* 19 (2010): 539–549, https://doi.org/10.1007/s11136-010-9606-8.

18. L. B. Mokkink, C. A. Prinsen, D. L. Patrick, et al., "COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments," *Netherlands* 2019 (2019): 1–32.

19. P. A. Harris, R. Taylor, B. L. Minor, et al., "The REDCap Consortium: Building an International Community of Software Platform Partners," *Journal of Biomedical Informatics* 95 (2019): 103208.

20. G. Rasch, *Studies in Mathematical Psychology: 1. Probabilistic Models for Some Intelligence and Attainment Tests* (Danmarks pædagogiske Institute, 1960).

21. B. D. Wright and G. N. Masters, *Rating Scale Analysis* (MESA Press, 1982).

22. S. J. Cano and J. C. Hobart, "The Problem With Health Measurement," *Patient Preference and Adherence* 5 (2011): 279–290, https://doi.org/10.2147/PPA.S14399.

23. S. Cleanthous, S. Bongardt, P. Marquis, C. Stach, S. Cano, and T. Morel, "Psychometric Analysis From EMBODY1 and 2 Clinical Trials to Help Select Suitable Fatigue pro Scales for Future Systemic Lupus Erythematosus Studies," *Rheumatology and Therapy* 8, no. 3 (2021): 1287–1301, https://doi.org/10.1007/s40744-021-00338-4.

24. D. Andrich, *Rasch Models for Measurement* (SAGE Publications, 1988), 7–68.

25. K. B. Christensen, G. Makransky, and M. Horton, "Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations," *Applied Psychological Measurement* 41 (2017): 178–194.

26. I. Marais, "Local Dependence," in *Rasch Models in Health*, ed. K. B. Christensen, S. Kreiner, and M. Mesbah (Wiley-ISTE Ltd, 2013), 111–130.

27. D. Andrich, "An Elaboration of Guttman Scaling With Rasch Models for Measurement," *Sociological Methodology* 15 (1985): 33–80.

28. L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika* 16 (1951): 297–334.

29. J. C. Nunnally, *Psychometric Theory*, 3rd ed. (McGraw-Hill, 1994).

30. H. Y. Kim, "Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) Using Skewness and Kurtosis," *Restorative Dentistry and Endodontics* 38, no. 1 (2013): 52–54.

31. R. E. Tractenberg, "Classical and Modern Measurement Theories, Patient Reports, and Clinical Outcomes," *Contemporary Clinical Trials* 31, no. 1 (2010): 1–3, https://doi.org/10.1016/S1551-7144(09)00212-2.

32. J. S. Antoun, P. V. Fowler, H. C. Jack, et al., "Oral Health-Related Quality of Life Changes in Standard, Cleft, and Surgery Patients After Orthodontic Treatment," *American Journal of Orthodontics and Dentofacial Orthopedics* 148 (2011): 568–575.

33. O. C. Damman, A. Jani, B. A. de Jong, et al., "The Use of PROMs and Shared Decision-Making in Medical Encounters With Patients: An Opportunity to Deliver Value-Based Health Care to Patients," *Journal of Evaluation in Clinical Practice* 26, no. 2 (2020): 524–540.

34. D. Feu, C. C. Quintão, and J. A. Miguel, "Quality of Life Instruments and Their Role in Orthodontics," *Dental Press Journal of Orthodontics* 15 (2010): 61–70.

35. F. Hua, "Increasing the Value of Orthodontic Research Through the Use of Dental Patient-Reported Outcomes," *Journal of Evidence-Based Dental Practice* 19, no. 2 (2019): 99–105.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix

Table 1. Summary characteristics of FACE-Q Dental Module and PROMs commonly used in orthodontics

PROM	Number items	Constructs	Age range development	Original development used Rasch or IRT	Development paper
FACE-Q Dental Module	Face (9) Jaws (7) Teeth (12) Smile (9) Eating/Drinking (9)	• Facial Appearance • Teeth Appearance • Smile Appearance • Jaw Appearance • Eating and Drinking	8-29	Yes	Klassen AF, et al. 2021. 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. 74(9):2319-29. Klassen AF et al. 2021. 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. 74(9):2330-40.
Oral Aesthetic Subjective Impact Scale (OASIS)	5	• Self-perceived oral appearance (e.g. teased, avoid smiling)	14-15	No	Mandall N et al. Perceived aesthetic impact of malocclusion and oral self-perceptions in 14-15-year-old Asian and Caucasian children in greater Manchester. Eur J Orthod. 2000 Apr 1;22(2):175-83.
Orthognathic QoL Questionnaire (OQLQ)	22	• Facial esthetics • Oral function • Awareness of facial esthetics • Social aspects	Adults	No	Cunningham SJ, et al. Development of a condition-specific quality of life measure for patients with dentofacial deformity: I. Reliability of the instrument. Community Dent Oral Epidemiol 2000;28:195-201. Cunningham SJ, et al. Development of a condition-specific quality of life measure for patients with dentofacial deformity: II. Validity and responsiveness testing. Community Dent Oral Epidemiol 2002;30:81-90
Child Oral Health QoL Questionnaire (COHQoL)	**CPQ** 6-7yrs;8-10yrs 25 items 11-14 yrs 16 items	• Oral symptoms • Functional limitations • Emotional well-being • Social well-being	6-14	No	Jokovic A. Development of an oral health outcome measure for children aged 6 to 14 years. 2003. University of Toronto. Jokovic A, et al. Questionnaire for measuring oral health-related quality of life in eight-to ten-

					year-old children. Pediatr Dent. 2004 Nov 1;26(6):512-8. Jokovic A, et al. Short forms of the Child Perceptions Questionnaire for 11–14-year-old children (CPQ 11–14): development and initial evaluation. Health Qual Life Outcomes. 2006 Dec;4:1-9.
Psychosocial Impact of Dental Aesthetic Questionnaire (PIDAQ)	23	• Dental self confidence • Social impact • Psychological impact • Aesthetic concern	18-30 (Further validation in 11-17 yr olds)	No	Klages U, et al. Development of a questionnaire for assessment of the psychosocial impact of dental aesthetics in young adults. Eur J Orthod. 2006;28(2):103–111. Klages U, et al. Psychosocial impact of dental aesthetics in adolescence: validity and reliability of a questionnaire across age-groups. Qual Life Res. 2015;24(2):379–390.
Malocclusion Impact Questionnaire (MIQ)	17	• Feel about appearance of teeth • Effect of teeth on life • Oral health and function	10-16	Yes	Benson PE, et al. Development of the Malocclusion Impact Questionnaire (MIQ) to measure the oral health-related quality of life of young people with malocclusion: part 2–cross-sectional validation. Journal of orthodontics. 2016 Jan 2;43(1):14-23.
Demand for orthodontic treatment questionnaire (DOTQ)	70	• Psychological and social • Malocclusion related • Treatment demand	13	No	Taghavi Bayat J, et al. Predicting orthodontic treatment need: reliability and validity of the Demand for Orthodontic Treatment Questionnaire. Eur J Orthod. 2017 Jun 1;39(3):326-33. Taghavi Bayat J, et al. Determinants of orthodontic treatment need and demand: a cross-sectional path model study. Eur J Orthod. 2017 Feb 1;39(1):85-91.
Child Oral Health Impact Profile (COHIP)	34	• oral health, • functional well-being, • social/emotional well-being, • school environment • self-image	8-15	No	Broder HL, et al Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. Community Dent Oral Epidemiol. 2007 Aug;35:8-19.
Oral Impacts on Daily Performance (OIDP)	9	• Physical performance (eg, eating)	35-44 adult 11-12 Child	No	Adulyanon S, et al. Oral impacts affecting daily performance in a low dental disease Thai

& CHILD-OIDP		• Psychological performance (eg, smiling, emotional stability) • Social performance (contact with people)			population. Community Dent Oral Epidemiol 1996. 24, 385-389.

Table 2: RMT item level fit statistics and differential item function results

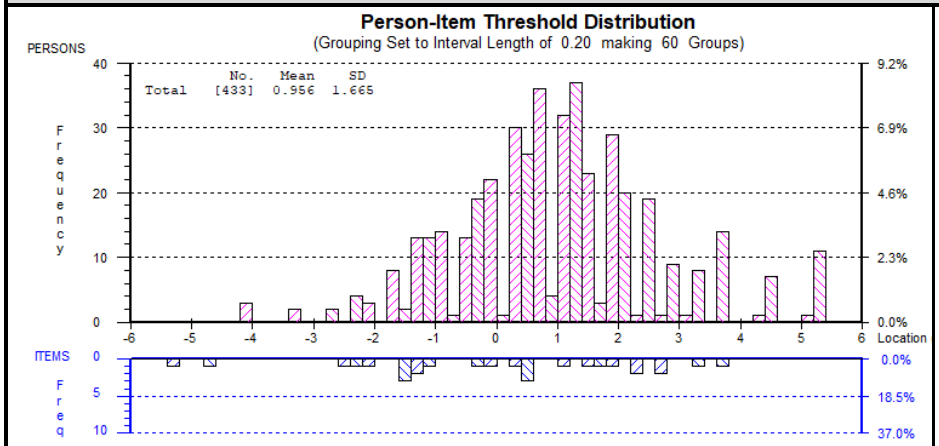
		ITEM FIT TO RASCH MODEL							DIF	
SCALE	**ITEM**	**Location**	**SE**	**Fit Residual**	**DF**	**χ^2**	**DF**	**Prob**	**Age**	**Gender**
FACE	Look best	-1.73	0.09	-1.04	363	5.83	6	0.44	no	no
	Go out	-1.56	0.09	-0.31	365	3.53	6	0.74	yes	yes
	Shape	-0.17	0.08	1.68	365	5.73	6	0.45	yes	no
	Smile	0.58	0.08	-2.28	364	15.99	6	0.01	no	no
	Laugh	-0.18	0.08	2.78	364	12.82	6	0.05	yes	no
	Photos	0.62	0.07	1.13	367	4.36	6	0.63	yes	yes
	Match	0.67	0.08	-1.78	369	6.57	6	0.36	no	no
	Profile	0.59	0.08	1.36	365	4.06	6	0.67	no	no
	Up close	1.19	0.08	-1.95	362	15.93	6	0.01	no	no
JAWS	Size	-0.10	0.11	-1.03	298	2.37	4	0.67	no	no
	Mouth closed	-0.04	0.11	-5.33	295	2.81	4	0.59	no	no
	Shape	-0.32	0.11	-5.15	296	5.38	4	0.25	no	no
	Mirror	0.06	0.11	-3.91	296	2.75	4	0.60	no	no
	Photos	-0.21	0.11	-4.58	297	4.82	4	0.31	no	no
	Smile	-0.03	0.11	0.67	298	9.65	4	0.05	no	no
	Profile	0.64	0.10	-1.80	298	2.91	4	0.57	no	no
SMILE	Expresses	-0.90	0.09	2.82	346	8.32	6	0.22	no	no
	Mirror	-0.16	0.09	-2.07	349	7.68	6	0.26	no	no
	Wide	-0.26	0.09	2.93	345	7.45	6	0.28	no	no
	Shape	-0.22	0.09	-0.69	347	4.93	6	0.55	no	no
	Even	-0.40	0.09	0.06	351	4.48	6	0.61	no	no
	Straight	-0.11	0.08	0.98	342	11.30	6	0.08	no	no
	Photo	0.52	0.09	-2.03	349	7.30	6	0.29	no	no
	Teeth	0.71	0.08	-1.36	348	5.75	6	0.45	yes	no
	Other people	0.82	0.08	-2.97	349	9.26	6	0.16	no	no

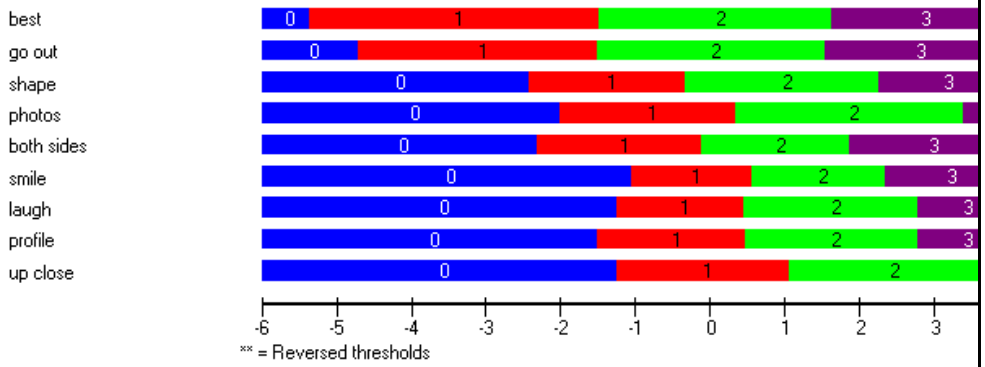
TEETH	Size	-0.82	0.09	3.99	372	20.97	5	0.00	no	no
	Close together	-0.51	0.08	-0.60	371	8.51	5	0.13	no	no
	Room	-0.22	0.08	1.72	364	3.85	5	0.57	no	no
	Shape	-0.24	0.08	-0.56	370	9.70	5	0.08	no	no
	Gum/teeth	-0.10	0.08	2.88	370	14.61	5	0.01	no	no
	Photos	0.07	0.08	-3.90	371	14.00	5	0.02	no	no
	Profile	0.32	0.08	1.12	369	5.29	5	0.38	no	no
	Straight	0.24	0.08	-2.24	368	5.23	5	0.39	yes	no
	Top/bottom	0.26	0.08	-2.15	370	6.40	5	0.27	no	no
	Smile	0.13	0.08	-2.40	371	6.01	5	0.30	no	no
	Line up	0.30	0.08	-3.51	370	7.91	5	0.16	no	no
	Compared	0.59	0.08	-2.36	368	8.08	5	0.15	no	no
EAT	Food falls	-1.78	0.16	0.97	232	2.97	2	0.23	no	no
	Liquid spills	-2.56	0.21	-0.71	231	1.13	2	0.57	no	no
	Straw	-3.68	0.39	-0.82	232	0.47	2	0.79	no	no
	Open mouth	0.34	0.16	0.16	230	2.61	2	0.27	no	no
	Avoid foods	1.82	0.13	-1.22	232	4.29	2	0.12	no	no
	Trouble biting	1.88	0.13	-0.68	232	4.88	2	0.09	no	no
	Chew	1.40	0.13	-0.94	230	5.18	2	0.07	no	no
	Small bits	0.67	0.14	-0.53	229	6.56	2	0.04	no	no
	Eat slow	1.93	0.13	0.87	230	1.81	2	0.41	no	no

SE = standard error; χ^2 = chi square; DF = degrees of freedom

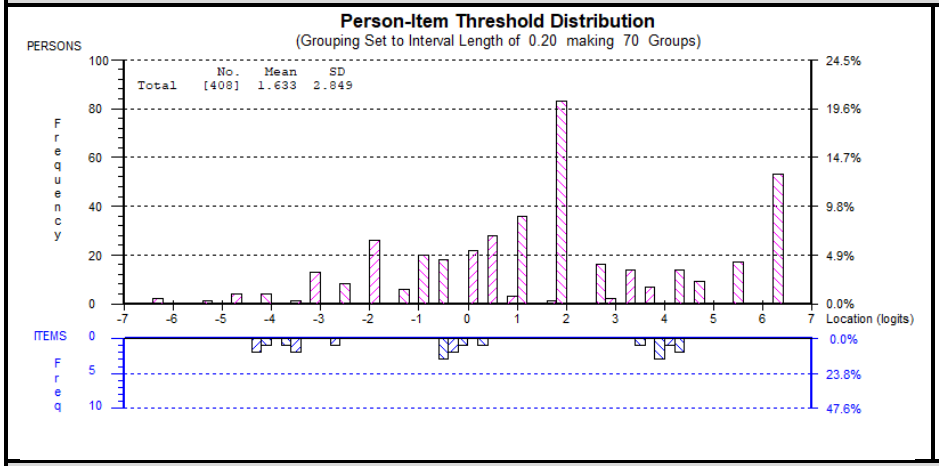
Figure 1

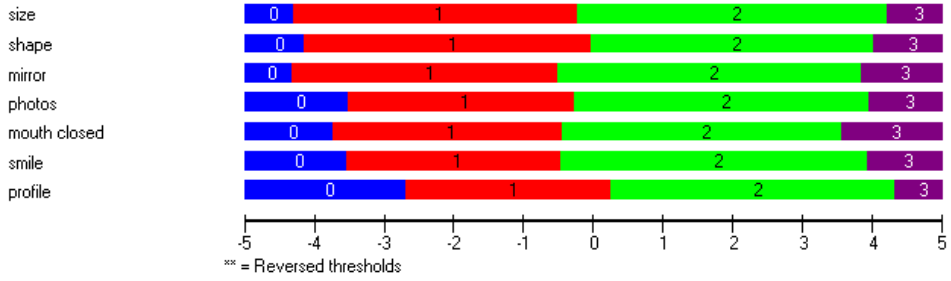
Face



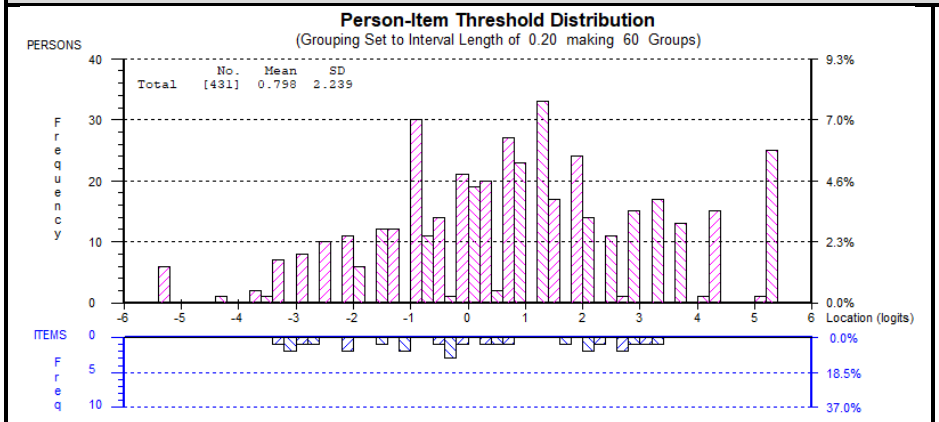


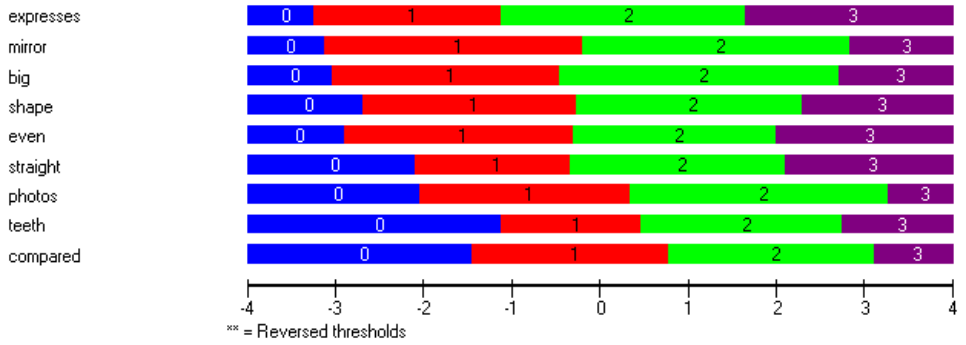
Jaws



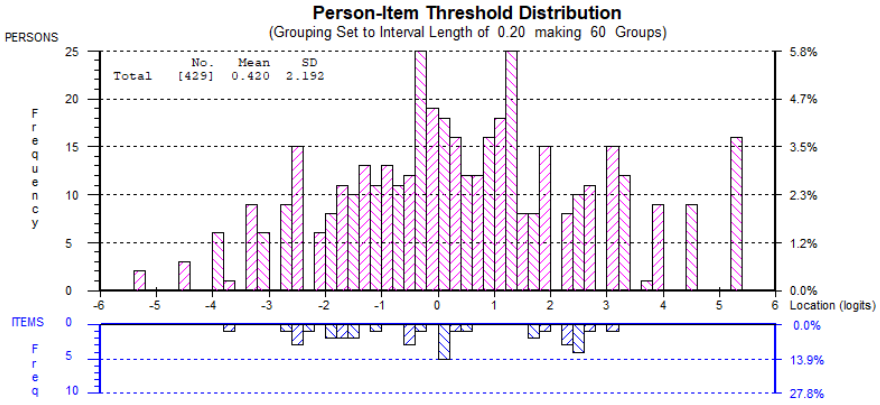


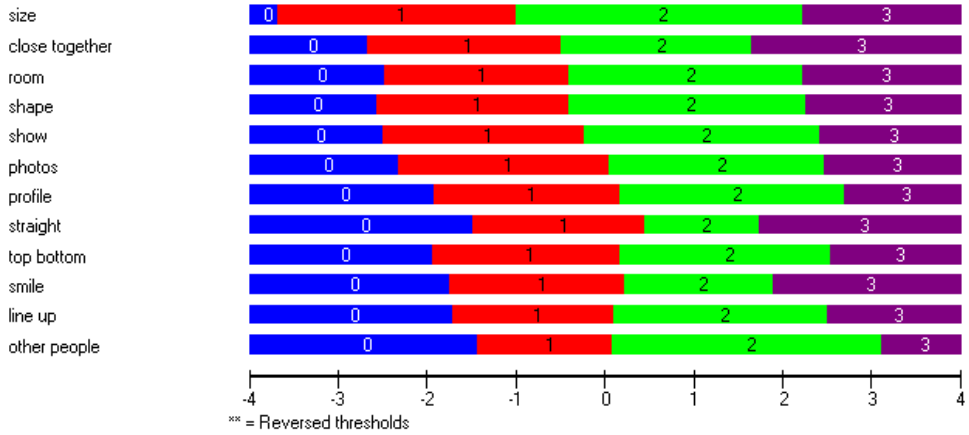
Smile


Person-Item Threshold Distribution
(Grouping Set to Interval Length of 0.20 making 60 Groups)
PERSONS
No. Mean SD
Total [431] 0.798 2.239
Frequency
Location (logits)
ITEMS
Freq

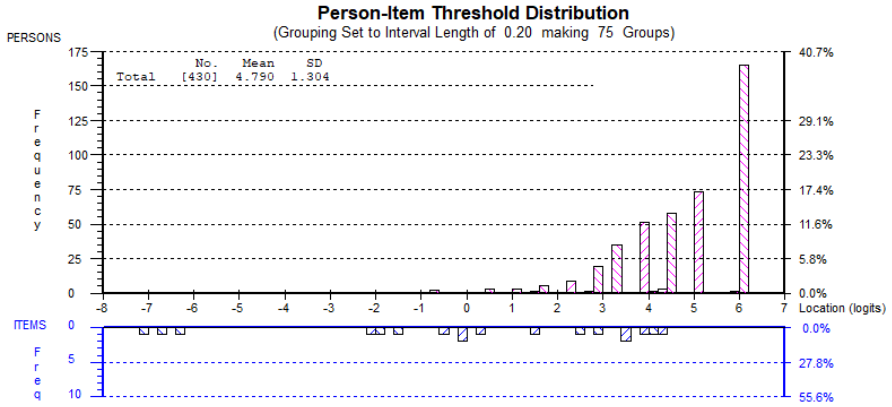

expresses
mirror
big
shape
even
straight
photos
teeth
compared
** = Reversed thresholds

Teeth


Person-Item Threshold Distribution
(Grouping Set to Interval Length of 0.20 making 60 Groups)
PERSONS
No. Mean SD
Total [429] 0.420 2.192
Frequency
Location (logits)
ITEMS
Freq


size
close together
room
shape
show
photos
profile
straight
top bottom
smile
line up
other people
** = Reversed thresholds

Eating & Drinking



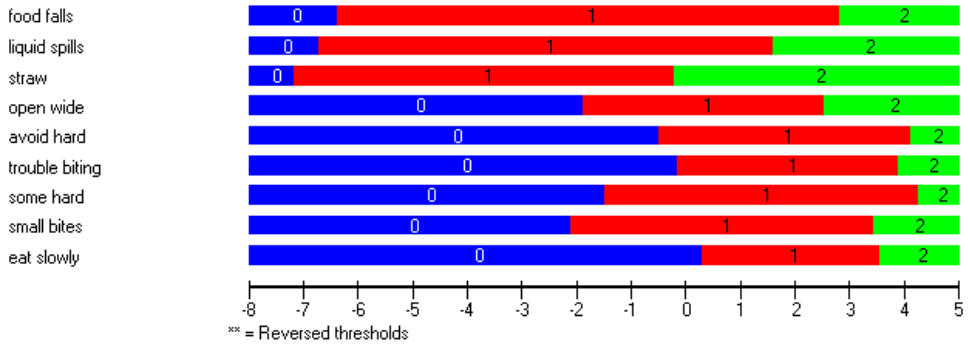


Table 3. Descriptive Statistics by sub-group for “How much do you like how your face looks overall?”

Scale	How much do you like how your face looks overall?	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	Not at all	6	26	10	4	16	36	14	40	p<0.001
	A little bit	65	41	10	1	39	43	14	59	
	Quite a bit	242	57	9	1	56	58	34	84	
	Very much	110	73	15	1	70	76	40	100	
Jaws	Not at all	5	28	8	4	18	39	17	36	p<0.001
	A little bit	64	43	18	2	39	48	0	100	
	Quite a bit	241	59	17	1	57	61	0	100	
	Very much	108	80	19	2	76	84	36	100	
Smile	Not at all	6	18	12	5	6	31	0	36	p<0.001
	A little bit	64	36	16	2	32	40	0	77	
	Quite a bit	241	58	17	1	56	60	19	100	
	Very much	110	73	18	2	70	77	36	100	
Teeth	Not at all	5	25	11	5	11	38	8	38	p<0.001
	A little bit	63	36	16	2	32	40	0	78	
	Quite a bit	241	54	17	1	52	56	8	100	
	Very much	110	67	20	2	63	70	18	100	
	A little bit	49	64	15	2	60	68	38	100	
	Quite a bit	169	74	18	1	71	77	27	100	
	Very much	71	77	17	2	73	81	27	100	

Table 4. Descriptive Statistics by sub-group for “How much do you like how your smile looks overall?”

Scale	How much do you like how your smile looks overall?	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	Not at all	29	41	16	3	35	47	14	76	<0.001
	A little bit	106	48	12	1	46	51	14	84	
	Quite a bit	159	59	10	1	57	60	34	100	
	Very much	125	70	14	1	68	73	40	100	
Smile	Not at all	29	22	15	3	16	27	0	57	<0.001
	A little bit	106	41	11	1	39	43	7	77	
	Quite a bit	160	59	10	1	57	60	30	100	
	Very much	126	80	14	1	77	82	41	100	
Teeth	Not at all	29	27	14	3	22	32	0	54	<0.001
	A little bit	103	38	13	1	35	41	0	81	
	Quite a bit	160	55	12	1	53	57	14	100	
	Very much	126	73	16	1	70	75	18	100	
	A little bit	85	68	16	2	65	72	27	100	
	Quite a bit	102	76	17	2	72	79	35	100	
	Very much	81	77	18	2	73	81	27	100	

Table 5. Descriptive Statistics by sub-group for “How much do you like how your jaw looks overall?”

Scale	How much do you like how your jaw looks overall?	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	Not at all	12	39	12	4	31	47	20	61	p<0.001
	A little bit	81	46	12	1	43	48	14	84	
	Quite a bit	218	57	10	1	56	59	20	91	
	Very much	111	72	15	1	69	75	31	100	
Jaws	Not at all	12	19	10	3	12	26	0	33	p<0.001
	A little bit	81	40	8	1	38	41	11	58	
	Quite a bit	217	59	11	1	58	61	22	100	
	Very much	112	88	15	1	85	90	46	100	

Table 6. Descriptive Statistics by sub-group for “How much do you like how your teeth look overall?”

Scale	How much do you like how your teeth look overall?	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	Not at all	59	44	14	2	40	47	14	84	p<0.001
	A little bit	110	53	11	1	51	55	20	84	
	Quite a bit	152	60	11	1	59	62	31	91	
	Very much	99	71	16	2	67	74	40	100	
Smile	Not at all	58	31	15	2	27	35	0	59	p<0.001
	A little bit	109	47	13	1	44	49	0	80	
	Quite a bit	153	61	12	1	59	63	23	100	
	Very much	99	81	16	2	78	84	39	100	
Teeth	Not at all	59	26	11	1	23	29	0	54	p<0.001
	A little bit	110	41	9	1	40	43	18	62	
	Quite a bit	154	58	10	1	56	60	14	100	
	Very much	99	78	14	1	75	81	40	100	
	A little bit	94	71	17	2	67	74	35	100	
	Quite a bit	103	76	16	2	73	79	33	100	
	Very much	50	76	19	3	70	81	27	100	

Table 7. Descriptive statistics by severity of jaw difference

Scale	Jaw difference	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	None	169	63	15	1	60	65	14	100	<0.001
	Mild	224	57	14	1	55	59	14	100	
	Major	38	48	14	2	44	53	20	76	
Jaws	None	169	69	21	2	66	72	17	100	<0.001
	Mild	220	59	20	1	57	62	0	100	
	Major	38	44	19	3	37	50	0	100	

Table 8. Descriptive statistics by severity of smile difference

Scale	Smile difference	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	None	157	64	15	1	61	66	28	100	p<0.001
	Mild	195	57	14	1	55	59	14	100	
	Major	79	51	16	2	48	55	14	100	
Smile	None	158	72	18	1	69	74	23	100	p<0.001
	Mild	193	53	18	1	50	55	0	100	
	Major	79	43	18	2	38	47	0	91	
Teeth	None	158	70	15	1	68	73	34	100	p<0.001
	Mild	194	48	15	1	46	50	8	100	
	Major	77	35	16	2	32	39	0	78	
	Mild	155	73	17	1	71	76	27	100	
	Major	67	69	16	2	65	73	42	100	

Table 9. Descriptive Statistics by severity of teeth difference

Scale	Teeth difference	N	Mean	SD	SE	95% CI		Min	Max	p-value
						LB	UB			
Face	None	145	65	15	1	62	67	31	100	p<0.001
	Mild	151	57	14	1	55	60	14	100	
	Major	135	53	15	1	50	55	14	100	
Smile	None	145	73	18	2	70	76	23	100	p<0.001
	Mild	151	55	18	1	52	58	0	100	
	Major	134	45	18	2	42	48	0	91	
Teeth	None	144	71	15	1	69	74	40	100	p<0.001
	Mild	152	51	16	1	48	53	8	100	
	Major	133	39	15	1	36	42	0	78	

Table 10. Descriptive Statistics by status of treatment

Scale	Treatment status	N	Mean	SD	SE	p-value
Face	Pre	249	55	15	1	p<0.001
	Post	182	63	15	1	
Jaws	Pre	246	58	21	1	p<0.001
	Post	181	66	22	2	
Smile	Pre	248	49	18	1	p<0.001
	Post	182	70	19	1	
Teeth	Pre	248	43	16	1	p<0.001
	Post	181	69	15	1	

Table 11: Pearson correlations between the FACE-Q Dental Module scales

Scales		**r**	**n**
Face	Jaws	.659**	426
	Smile	.747**	428
	Teeth	.663**	427
Jaws	Face	.659**	426
	Smile	.554**	424
	Teeth	.534**	424
Smile	Face	.747**	428
	Jaws	.554**	424
	Teeth	.826**	426
Teeth	Face	.663**	427
	Jaws	.534**	424
	Smile	.826**	426

**p≤0.001; criteria: similar constructs ≥0.50; related but dissimilar constructs 0.30-0.50; unrelated constructs <0.30