Psychometric Properties of the Croatian Version of the Orofacial Esthetic Scale and Suggestions for Modification

Sanja Persic, Dr Med Dent[a]/Sladjana Milardovic, Dr Med Dent[a]/Ketij Mehulic, Dr Med Dent, MSc, PhD[b]/
Asja Celebic, Dr Med Dent, MSc, PhD[b]

Purpose: The aims of this study were to develop a Croatian version of the Orofacial Esthetic Scale (OES) and to test its psychometric properties. ***Materials and Methods:*** The English version of the OES was translated into Croatian (forward-backward method). The original 11-point scale as well as a 5-point scale (1 = unsatisfactory, 5 = excellent) were used. Convergent validity was tested on 126 subjects, and discriminative validity was tested on the same subjects divided into four groups: esthetically normal patients (n = 25), esthetically impaired patients (n = 42), esthetically normal controls (n = 37), and esthetically impaired controls (n = 22). Test-retest reliability was tested on 43 subjects. Responsiveness was tested on 32 esthetically impaired patients who received prosthodontic treatment. ***Results:*** An additional explanation was added to the first two items of the OES. Convergent validity was confirmed by the association between OES scores and self-reported oral esthetics and three questions from the Oral Health Impact Profile related to esthetics (correlation coefficients ranged from 0.734 to 0.811, $P < .001$). Discriminative validity showed the results as predicted. Test-retest reliability showed high intraclass correlation (0.79 to 0.95) and no significant differences between the two administrations of the 5-point OES scale ($P > .05$). The 11-point OES scale showed significant differences for questions 3 and 8 ($P < .01$). Internal consistency showed high Cronbach α values (0.802 to 0.962). Responsiveness was confirmed by a significant difference between baseline and follow-up ($P < .001$) and a high effect size. ***Conclusion:*** Psychometric properties of the Croatian version of the OES render the instrument suitable for the assessment of esthetics in Croatia. The authors recommend changing the first two items by adding the explanation that the questions are related to the lower third of the face and using the 5-point scale for rating. *Int J Prosthodont 2011;24:523–533.*

It has been proven that dental appearance influences not only other peoples' judgment of a person's facial attractiveness, but also personal characteristics.[1,2] The importance of dentofacial attractiveness to the psychosocial well-being of an individual has been well established.[3]

Improvement of esthetic appearance is one of the main reasons why patients seek dental treatment. Moreover, esthetics is strongly correlated with patients' satisfaction with any restorative or prosthodontic treatment.[4–8] Many authors found that quality of life and self-confidence were higher in patients who were satisfied with their dental restorations; thus, dentofacial esthetics is not only important itself, it is also associated with many other general concepts of welfare.[7,8]

However, little information is available regarding dental patients' perceptions of a pleasing esthetic appearance. Some studies show that dentists are far more critical in their esthetic perceptions than patients or laypeople in general.[9–16] Therefore, expert-based assessments have shifted to patient-based approaches through the use of questionnaires. Still, assessment is challenging because the matter is neither directly observable nor measurable, and several factors, such as culture, environment, social norms, age, sex, and level of education, affect a patient's esthetic perceptions.[9,14–16] Certainly, orofacial esthetics is influenced by the position, shape, size, and shade of the teeth; the position, texture, color, and lines of the gingiva and lips; and the shape of the jaws.[17–23] Therefore, those factors should be included in measuring instruments. Although there have been increased demands for esthetic procedures during

[a]*Assistant, Department of Prosthodontics, School of Dental Medicine, University of Zagreb, Zagreb, Croatia.*
[b]*Professor, Department of Prosthodontics, School of Dental Medicine, University of Zagreb, Zagreb, Croatia.*

Correspondence to: *Dr Asja Celebic, Department of Prosthodontics, School of Dental Medicine, University of Zagreb, Gunduliceva 5, 10000 Zagreb, Croatia. Fax: +38514802159. Email: celebic@sfzg.hr*

prosthodontic treatment, only few self-reporting instruments are currently available to measure and describe how prosthodontic patients perceive the appearance of their orofacial esthetics.[24] The Oral Health Impact Profile (OHIP)[25] is a very popular instrument for measuring oral health–related quality of life, but it does not evaluate dental appearance sufficiently by itself.[26] Recently, Larsson et al[17,27,28] introduced the Orofacial Esthetic Scale (OES), an 8-item instrument for measuring self-reported orofacial esthetics in patients with prosthodontic concerns. The questionnaire was designed to be used as a stand-alone instrument to measure direct esthetic impacts. However, it can also be used together with broader instruments, such as the OHIP questionnaire, that cover indirect esthetic impacts comprehensively. The OES questionnaire showed good reliability and validity.[28] Therefore, it seems that the OES could be widely used.

Cross-cultural adaptation of an instrument is not a simple issue. Instruments must be translated and adapted to languages other than the source language, and the psychometric properties of the translated instruments must be appropriate in the new culture and new country.[29–31]

The aim of this study was to test the psychometric properties of the Croatian translation of the OES questionnaire.

Materials and Methods

OES Translation

The English version of the OES was translated into Croatian according to accepted methods.[27–31] The translation was done jointly by a professional translator familiar with dental vocabulary and two dentists with excellent knowledge of the English language who had spent at least 6 months in the United States for educational purposes. The translation was reviewed by two Croatian dentists from the Department of Prosthodontics, School of Dental Medicine, University of Zagreb, Croatia, who had excellent knowledge of the English language. The translators and language reviewers worked independently. The translations were merged into one final version. The final Croatian version of the OES (OES-CRO) was then back-translated into English by another professional translator also working with a dentist who had excellent knowledge of the English language and who had attended his or her postdoctoral studies in English-speaking countries. The back-translation was then evaluated by two native English speakers, who compared it with the original English version.

Afterward, a pilot study was performed with a group of 10 participants (age range: 23 to 62 years) to test the clarity of the questions. Prior to distribution of the questionnaire to the participants, 10 individuals with a pleasing orofacial appearance were chosen, and they completed the OES questionnaire. Two specialists discussed each item of the assessment with them. It was surprising that some older participants with more than satisfactory orofacial esthetics gave lower ratings on the first two OES items than on the next six items. They explained that they graded their facial appearance and facial profiles with lower scores because of the wrinkles on their foreheads and around their eyes. One person rated her facial profile with low grades because of an imperfection of her nose. Therefore, the authors decided to reformulate the first two items. The first item, "How do you feel about your facial appearance?," was changed to "How do you feel about the appearance of the lower third of your face?," and the second item was changed from "How do you feel about your profile facial appearance?" to "How do you feel about the profile appearance of the lower third of your face?" Afterward, 8 new older patients (age range: 58 to 70 years) with a pleasing orofacial appearance filled out the OES questionnaire, and the two prosthodontists discussed their ratings with them. There were no further misunderstandings or errors in comprehension considering the first two items. Patients involved in testing the clarity of the translated OES questionnaire were not included in further research because they already knew the questions, which could affect the results.

Subjects

The study was approved by the institutional ethics committee. One hundred twenty-six subjects participated in the study and were divided into four groups that included two patient groups (n = 67) and two control groups (n = 59), similar as in the original study.[27,28] Patients were further divided into two groups: esthetically normal but functionally impaired patients (P-EN) (n = 25) and esthetically impaired patients (P-EI) (n = 42). Patients in the P-EN group had complete dentures (CDs) (n =14) or removable partial dentures (RPDs) replacing posterior teeth attached to metal-ceramic anterior dentition by precise attachments (n = 8). The remaining 3 patients had fixed partial dentures (FPDs). P-EN patients visited a dentist annually and were satisfied with their esthetic appearance. Also, four prosthodontic specialists agreed that subjects in the P-EN group had satisfactory orofacial esthetics. The P-EI group

comprised patients seeking treatment at the department of prosthodontics. Some of them had very old and unsatisfactory CDs (n = 13) with highly decreased vertical dimensions of the lower third of the face and denture teeth showing a high degree of wear or staining, while others (n = 29) had some anterior teeth missing and needed either FPDs or FPDs in combination with RPDs.

Controls were also divided into two groups: esthetically normal controls (C-EN) (n = 37) and esthetically impaired controls (C-EI) (n = 22). The C-EN group included subjects with natural teeth, no orthodontic or dentofacial anomalies, and no need for any dental treatment. The C-EI group included subjects with natural teeth but with orthodontic anomalies (large midline diastemata or anterior teeth showing considerable crowding or a large amount of gingival display during smiling [gummy smile]). The sampling strategy was similar to that used in the evaluation of the psychometric properties of the original OES questionnaire.[27,28]

Each participant received a thorough verbal explanation of the aims of the study. Only those who provided verbal informed consent were included. Participants completed the original OES questionnaire (8 items) using the original 11-point scale (0 to 10). They also answered one question considering their general satisfaction with their orofacial esthetics (scale: 1 to 5) and three questions from the Croatian version of the OHIP-49 related to orofacial esthetics (items 3, 22, and 31).[30] Participants also completed the OES questionnaire (same 8 items) but using a scale from 1 to 5 (1 = unsatisfactory, 5 = excellent). This was done because in Croatia, the system of validation in primary schools, high schools, and universities is based on a 5-point Likert scale (1 = unsatisfactory, 5 = excellent).

Statistical Analysis

Statistical analysis was performed using SPSS version 17.0 for Windows (IBM) and Microsoft Office Excel 2003 (Microsoft). Descriptive statistics were calculated for all groups. Coefficients of variation (CVs) were also calculated.

Validity

To test the psychometric properties of the OES-CRO, two types of validation were conducted: discriminative and convergent validity. Convergent validity was determined from the association between self-reported general satisfaction with orofacial esthetics and the OES (0 to 10) and OES (1 to 5) summary scores of the first 7 items by using Spearman rank correlation.[32] The association was also determined between self-reported general satisfaction with orofacial esthetics and the average OES (0 to 10) and OES (1 to 5) scores. The OES (0 to 10) the OES (1 to 5) summary scores and the OES (0 to 10) and OES (1 to 5) average scores were also related with the OHIP summary score for the three items related to esthetics (items 3, 22, and 31) and with the OHIP average score for the same three items.

Discriminative validity was tested comparing the OES (0 to 10) and OES (1 to 5) average and summary scores between the two patient and two control groups (one-way analysis of variance [ANOVA] and Sheffé post hoc test). Relevant and statistically significant differences were expected between the two patient and two control groups. No significant difference was expected between the P-EN and C-EN groups. However, statistically significant differences were expected between the P-EI and C-EI groups, since the P-EI group had some missing anterior teeth and, therefore, more pressing esthetic disorders than the C-EI group, consisting of individuals who still had all of their natural teeth.

Reliability

Two types of reliability were assessed: test-retest reliability and internal consistency. Test-retest reliability was assessed by calculating the intraclass correlation coefficients (ICCs) based on one-way repeated-measures ANOVA, using both the OES (0 to 10) and OES (1 to 5) questionnaires, including the average and summary scores from each questionnaire. A total of 43 subjects were included: 16 subjects from the P-EN group and 27 subjects from the C-EN group. Each subject completed the same questionnaire twice within a 2-week time interval. No subject received any treatment during the observed period. It was predicted that the OES scores would not change during the 2-week period without any oral treatment. ICCs were calculated according to the Shrout and Fleiss method.[33] The standard error of measurement was also calculated using the following formula:

$$SEM = SD*SQRT(1-r^2)$$

where SD is the standard deviation of the baseline scores of the test-retest subjects, r is a test-retest reliability coefficient, and SQRT is the square root. Internal consistency was assessed by calculating the Cronbach α coefficient and the average interitem correlation for the OES scores.[34]

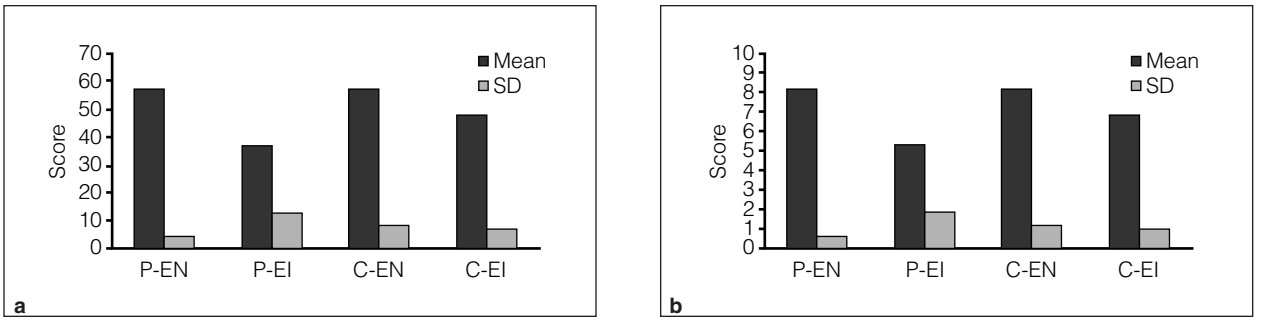

Fig 1 **(a)** Mean summary scores and standard deviations and **(b)** mean average scores and standard deviations for the OES (0 to 10) questionnaire.

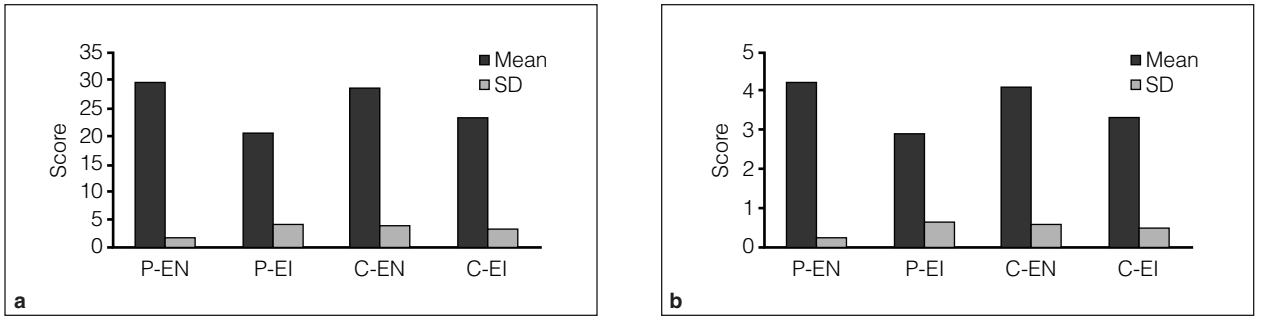

Fig 2 **(a)** Mean summary scores and standard deviations and **(b)** mean average scores and standard deviations for the OES (1 to 5) questionnaire.

Responsiveness

Responsiveness was tested for the OES (0 to 10) and OES (1 to 5) questionnaires. For that purpose, 32 esthetically impaired patients were included (P-EI group). They completed the OES questionnaires twice (once before treatment and again 2 weeks after prosthodontic treatment). A total of 13 patients received new CDs; 15 patients received FPDs in the maxillary anterior region, mostly composed of metal-ceramic (only 3 subjects received all-ceramic FPDs); and 4 patients received RPDs replacing posterior teeth attached to an FPD in the anterior region (metal-ceramic) with precision attachments. It was assumed that the OES scores would improve after treatment compared to the status before treatment. The significance of the difference of the OES (0 to 10) and OES (1 to 5) questionnaires, including the average and summary scores, between baseline and follow-up was tested using paired *t* tests and by calculating the standardized effect size. The standardized effect size was calculated according to Allen et al[35] as follows:

Mean (baseline OES score – follow up-OES score) / standard deviation of the baseline OES score

Results

Subjects

A total of 126 individuals were included in the study, divided into two patient and two control groups. Subjects in the two patient groups (P-EN, mean age: 63.6 ± 11.16 years; P-EI, mean age: 54.12 ± 18.70 years) were older than subjects in the two control groups (C-EN, mean age: 39.19 ± 14.46 years; C-EI, mean age: 35.68 ± 10.19 years). All four groups examined had more women than men. The P-EN group had 76% women vs 24% men, P-EI had 69% women vs 31% men, C-EN had 62% women vs 38% men, and C-EI had 59% women vs 41% men.

Mean summary scores and standard deviations and mean average scores and standard deviations of the OES (0 to 10) are presented in Figs 1a and 1b, respectively. Mean summary scores and standard deviations and mean average scores and standard deviations of the OES (1 to 5) are presented in Figs 2a and 2b, respectively. CVs for the OES (1 to 10) average scores were higher than CVs of the OES (1 to 5) average scores: P-EN = 8.2% vs 6.11%, P-EI = 34.34% vs 21.6%, C-EN = 15.68% vs 14.36%, and C-EI = 14.7%

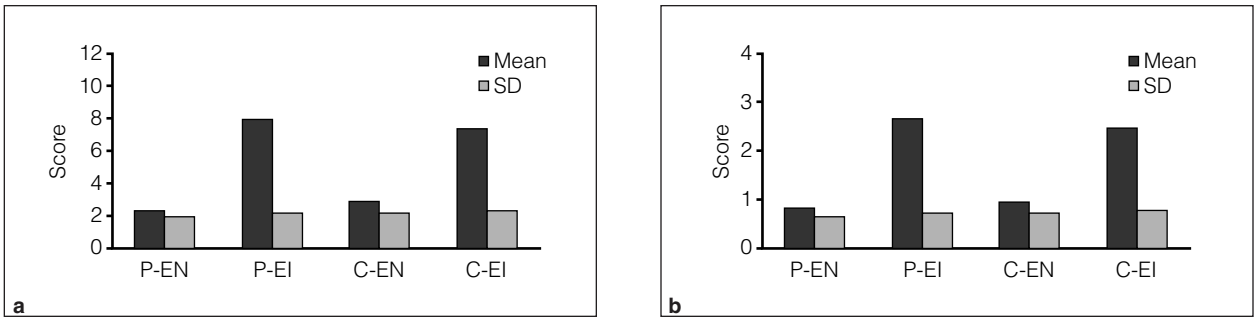

Fig 3 **(a)** Mean summary scores and standard deviations and **(b)** mean average scores and standard deviations for the OHIP 3-item questionnaire.

Table 1 Cronbach α Values for the Original OES (0 to 10) and the OES (1 to 5) Questionnaires

		OES (0 to 10)		OES (1 to 5)	
	n	Cronbach α	Mean ICC	Cronbach α	Mean ICC
All examinees	126	0.962	0.766	0.941	0.677
P-EN	25	0.868	0.502	0.802	0.404
P-EI	42	0.948	0.704	0.912	0.576
C-EN	37	0.940	0.679	0.916	0.568
C-EI	22	0.916	0.595	0.915	0.593

ICC = intraclass correlation coefficient.

vs 13.69%, respectively. Similar results were obtained for the CVs of the OES summary scores, which were higher for the OES (1 to 10) than for the OES (1 to 5): P-EN = 8.03% vs 5.87%, P-EI = 34.43 vs 21.74%, C-EN = 15.6% vs 14.22%, and C-EI = 15.01% vs 14.06%, respectively.

Mean summary scores and standard deviations and mean average scores and standard deviations for the three questions of the OHIP are presented in Figs 3a and 3b, respectively. The P-EN and C-EN groups had the highest OES scores, followed by the scores for the C-EI. The P-EI had the lowest scores. The OHIP scores were the highest in the P-EI and C-EI groups and lowest in the C-EN and P-EN groups.

Reliability

Two types of reliability were assessed: test-retest reliability and internal consistency. To assess internal consistency, the Cronbach α coefficient and average interitem correlation were calculated for both OES questionnaires (0 to 10 and 1 to 5). All subjects were included in the analysis (n = 126). The results are presented in Table 1. Both OES questionnaires (0 to 10 and 1 to 5) showed more than satisfactory Cronbach α coefficients. Test-retest reliability was evaluated in 43 subjects who completed the questionnaires twice in a 2-week period without any dental treatment. The 95% confidence intervals of the means were computed. Mean differences between the first and second OES questionnaires, ICCs, 95% confidence intervals of the mean, and the levels of significance are presented in Tables 2 (OES 0 to 10) and 3 (OES 1 to 5). The differences were small. For the OES (0 to 10) questionnaire, the greatest difference was −0.35 for question 3 and −0.91 for the OES summary score. For the OES (1 to 5) questionnaire, the greatest difference was −0.14 for question 8 and −0.16 for the OES summary score. However, questions 3 and 8 of the OES (0 to 10) showed statistically significant differences between the test and retest ($P < .05$). The OES (1 to 5) showed no significant difference for each item or for the average and summary scores ($P > .05$). ICCs were satisfactory for both questionnaires. The standard error of measurement was 3.532 points for the OES (0 to 10) questionnaire and 1.554 points for the OES (1 to 5) questionnaire.

Table 2 Test-Retest Reliability for the OES (0 to 10) Questionnaire

Question	ICC	Mean difference	95% CI	*P*
1	0.92	−0.16	−0.37 ± 0.05	0.128 NS
2	0.89	−0.05	−0.29 ± 0.20	0.700 NS
3	0.88	−0.35	−0.62 ± -0.07	0.014**
4	0.92	−0.14	−0.40 ± 0.12	0.278 NS
5	0.91	−0.02	−0.27 ± 0.22	0.850 NS
6	0.81	−0.30	−0.61 ± 0.01	0.057 NS
7	0.93	0.12	−0.10 ± 0.33	0.280 NS
8	0.92	−0.26	−0.46 ± 0.05	0.015**
Average score	0.94	−0.15	−0.31 ± 0.02	0.080 NS
Summary score	0.94	−0.91	−2.06 ± 0.24	0.119 NS

ICC = intraclass correlation coefficient; CI = confidence interval; NS = not significant.
**$P < .05$.

Table 3 Test-Retest Reliability for the OES (1 to 5) Questionnaire

Question	ICC	Mean difference	95% CI	*P*
1	0.84	−0.05	−0.21 ± 0.12	0.570 NS
2	0.85	−0.05	−0.20 ± 0.10	0.533 NS
3	0.82	−0.07	−0.25 ± 0.11	0.445 NS
4	0.79	0.02	−0.16 ± 0.21	0.800 NS
5	0.89	−0.07	−0.19 ± 0.05	0.262 NS
6	0.82	−0.07	−0.21 ± 0.07	0.323 NS
7	0.87	0.09	−0.04 ± 0.22	0.160 NS
8	0.84	−0.14	−0.28 ± 0.00	0.057 NS
Average score	0.94	−0.04	−0.11 ± 0.04	0.323 NS
Summary score	0.95	−0.16	−0.68 ± 0.35	0.527 NS

ICC = intraclass correlation coefficient; CI = confidence interval; NS = not significant.

Table 4 Spearman Rank Correlation Between Variables for Convergent Validity***

Variable	OHIP summary score	OHIP average score	General satisfaction with orofacial esthetics (1 to 5)
OES (0 to 10) average	−0.734	−0.734	0.786
OES (0 to 10) summary	−0.736	−0.737	0.783
OES (1 to 5) average	−0.795	−0.796	0.811
OES (1 to 5) summary	−0.792	−0.793	0.804

***$P < .001$.

Validity

Convergent validity was verified by a significant positive association ($P < .001$) between the self-reported satisfaction with orofacial esthetics and the OES (0 to 10) and OES (1 to 5) average and summary scores. It was also verified by a negative significant association between the OHIP summary scores and the OES (0 to 10) and OES (1 to 5) average and summary scores, as well as by a negative significant association between the OHIP average scores and the OES (0 to 10) and OES (1 to 5) average and summary scores (Table 4). Discriminative validity was tested comparing the OES (0 to 10) and OES (1 to 5) average and summary scores between the two patient and two control groups (one-way ANOVA and Sheffé post hoc test). The results are presented in Table 5. No significant difference was found between the P-EN and C-EN groups for all OES scores. Statistically significant differences were found between the two patient and two control groups ($P < .01$), as expected. However, statistically significant differences were also found between the P-EI and C-EI groups.

Responsiveness

To test the responsiveness of the OES (0 to 10) and OES (1 to 5) questionnaires, 32 patients requiring prosthodontic treatment were included. For that purpose, the questionnaires were filled out twice: once prior to treatment and 2 weeks after treatment. The results are presented in Tables 6 and 7 for the OES (0 to 10) and OES (1 to 5) questionnaires, respectively. Mean change in score after treatment for the summary score was 17.28 ($P < .001$) for the OES (0 to 10) questionnaire and 7.56 ($P < .001$) for the OES (1 to 5) questionnaire. The standardized effect size was 1.26 for the OES (0 to 10) summary score and 1.23 for the OES (0 to 10) average score. The standardized effect size was 1.55 for the OES (1 to 5) summary score and 1.57 for the OES (1 to 5) average score.

Table 5 Significance of Differences Between Two Patient and Two Control Groups (One-Way ANOVA)

	Sheffé post hoc test					
Variable	P-EN	P-EI	C-EN	C-EI	F	*P*
OES (0 to 10) average score					39.11	< .01
P-EN		*	NS	*		
P-EI	*		*	*		
C-EN	NS	*		*		
C-EI	*	*	*			
OES (0 to 10) summary score					40.10	< .01
P-EN		*	NS	*		
P-EI	*		*	*		
C-EN	NS	*		*		
C-EI	*	*	*			
OES (1 to 5) average score					54.12	< .01
P-EN		*	NS	*		
P-EI	*		*	*		
C-EN	NS	*		*		
C-EI	*	*	*			
OES (1 to 5) summary score					54.48	< .01
P-EN		*	NS	*		
P-EI	*		*	*		
C-EN	NS	*		*		
C-EI	*	*	*			

NS = not significant.
*P < .01.

Table 6 Responsiveness of the OES (0 to 10) Questionnaire

Question	Mean difference	95% CI	*P*
1	−2.41	−3.19 ± −1.62	< .01
2	−2.19	−2.95 ± −1.42	< .01
3	−2.72	−3.38 ± −2.06	< .01
4	−2.69	−3.46 ± −1.92	< .01
5	−2.50	−3.25 ± −1.74	< .01
6	−2.87	−3.59 ± −2.15	< .01
7	−1.97	−2.72 ± −1.22	< .01
8	−2.37	−3.01 ± −1.74	< .01
Average score	−2.46	−3.11 ± −1.80	< .01
Summary score	−17.28	−21.94 ± −12.62	< .01

CI = confidence interaval.

Table 7 Responsiveness of the OES (0 to 10) Questionnaire

Question	Mean difference	95% CI	*P*
1	−1.03	−1.31 ± −0.75	< .01
2	−1.06	−1.34 ± −0.79	< .01
3	−1.22	−1.52 ± −0.92	< .01
4	−1.09	−1.46 ± −0.72	< .01
5	−1.06	−1.37 ± −0.76	< .01
6	−1.28	−1.60 ± −0.96	< .01
7	−0.81	−1.15 ± −0.48	< .01
8	−1.06	−1.34 ± −0.79	< .01
Average score	−1.08	−1.33 ± 0.83	< .01
Summary score	−7.56	−9.30 ± −5.82	< .01

CI = confidence interval.

Discussion

The OES questionnaire was developed to assess the direct or primary impacts of impaired orofacial esthetics on prosthodontic patients.[17,27,28] It was constructed to reflect patients' perceived esthetic values to not mix esthetic items with psychosocial items.[26–28,36,37] Questions related to psychosocial consequences of impaired oral esthetics are contained in both the OHIP-14 and the OHIP-49 questionnaires.[25,31] The OES questionnaire was developed based on patients' opinions together with input from dental professionals. The OES questionnaire consists of eight items. The first seven items create the OES summary score, and the eighth item is a global assessment of overall impact. The authors recommend the use of the OES questionnaire in daily practice as well as in research studies alone or together with other questionnaires. As of recently, only Swedish and English versions of the OES questionnaire were available.[17,27,28]

There was no Croatian version of the OES questionnaire available. Therefore, the aim of the study was to develop the Croatian version of the OES questionnaire (the OES-CRO) and to test the psychometric properties of the new instrument (validity, reliability, and responsiveness). To achieve this goal, the English OES questionnaire had to be adapted into the Croatian cultural environment. The original and

English versions of the OES questionnaire used the 11-point scale (0 to 10).[17,27,28] However, psychometric properties of a questionnaire may depend on the scale format (number of scale categories). Scales in primary and high schools and universities in Croatia traditionally range from 1 (unsatisfactory) to 5 (excellent). In cross-cultural assessment, the first step is to determine whether scales are actually measuring the same concepts in all countries. Specifically, construct equivalence or absence of construct bias has to be established. Therefore, it was decided that subjects included in the study would rate their esthetics twice, the first time using the original 11-point OES scale and the second time using the 5-point Likert scale (1 = unsatisfactory, 5 = excellent). Subjects included in the study also had to answer one general question related to esthetics: "How do you assess your orofacial esthetics in general?" and three questions from the OHIP-CRO questionnaire related to esthetics.[30,31] The same three OHIP questions were used as in the original study.[17,27,28] The OHIP scores were rated as 0 = no problems and 4 = the most severe problems.[30,31] Subjects either filled out the OES 5-point or OES 11-point scale first, then answered the three OHIP questions related to esthetics, and then answered the one general question related to orofacial esthetics. In the end, they again filled out the OES questionnaire but using a different rating scale than previously used. Which OES scale would be answered first was determined randomly but in a way so that one half of the individuals filled out the OES (0 to 10) and the other half filled out the OES (1 to 5).

In general, there are eight key attributes of instrument assessment: conceptual and measurement model, reliability, validity, responsiveness, interpretability, respondent and administrative burden, alternative forms, and cultural and language adaptations.[37] The methods of assessing reliability of a questionnaire involve determining the extent to which the test produces consistent results on retesting (test-retest), the relative accuracy of a test at a given time (alternate forms), the internal consistency of the items (split half), and the degree of agreement between two examiners (inter-examiner agreement).[38] Internal consistency examines whether several items that measure the same general construct produce similar scores. Cronbach α is a summary statistic that captures the extent of agreement between all possible subsets of questions. This study tested the reliability of the questionnaires using the Cronbach α (126 individuals) and the test-retest approach (43 individuals). Cronbach α values > 0.80 indicate a reliable scale, although values > 0.70 are also acceptable.[34] The Cronbach α coefficients were satisfactory in this study. Both OES questionnaires (0 to 10 and 1 to 5) showed Cronbach α values > 0.80. Average interitem correlations also confirmed satisfactory reliability of both OES questionnaires (see Table 1).

Test-retest reliability was also tested. ICCs varied between 0.81 and 0.94 for the OES (0 to 10) and between 0.79 and 0.95 for the OES (1 to 5) questionnaires (Tables 2 and 3). The OES questionnaire using the scale of 1 to 5 showed no significant differences between the two administrations of the same test either for each item or for the OES average or summary scores ($P > .05$). Contrary to that, the OES (0 to 10) questionnaire showed statistically significant differences for the third and eighth questions between the test and retest. Therefore, the OES (1 to 5) questionnaire showed better test-retest properties than the OES (0 to 10) questionnaire. This may be because Croatian people are used to the Likert scale (1 to 5) for rating.

Scales with more categories may result in greater data dispersion. The CVs obtained in this study for the OES (0 to 10) average and summary scores were higher in all groups than those for the OES (1 to 5) average and summary scores. Higher CVs also indicate less representative mean values. Formally, when introducing more scale categories, the statistical value of the obtained results may decrease. The 11-point scale is probably too broad for exact retesting and causes more confusion. If the scale is too broad and difficult to use, respondents may become demotivated and the quality of their responses may decrease.

Another important psychometric property of a questionnaire is its validity. Construct validity defines the ability to seek agreement between a theoretical concept and a specific measuring procedure. Convergent validity investigates how closely the new scale is related to other measures of the same construct. Therefore, the results of the OES (1 to 5) and OES (0 to 10) questionnaires were compared with the test measuring similar clinical properties (esthetics): one general question regarding overall orofacial esthetics and three questions from the OHIP related to orofacial esthetics. The same three questions from the OHIP were used as in the original study.[17,27,28] There was a significant positive association between self-reported satisfaction with orofacial esthetics and the OES (0 to 10) and OES (1 to 5) average and summary scores ($P < .001$). Moreover, there was also a negative significant association between the OHIP summary scores and the OES (0 to 10) and OES (1 to 5) average and summary scores, as well as between the OHIP average scores and the OES (0 to 10) and OES (1 to 5) average and summary scores. The Spearman coefficient of correlation varied between 0.734 and 0.811, respectively (see Table 4).

Discriminative validity is one way to measure construct validity, and it describes the ability to discriminate between groups with different treatment needs.[37] The authors tested the discriminative validity between all four included groups using one-way ANOVA and the Sheffé post hoc test. They predicted that both esthetically normal groups (P-EN and C-EN) would have significantly higher OES scores than the esthetically impaired groups (P-EI and C-EI). They also predicted that the esthetically impaired patient group (P-EI) would have lower scores than the esthetically impaired control group (C-EI) because patients in the P-EI group had some missing anterior teeth or had old CDs with highly reduced vertical dimensions of the lower third of the face and stained and worn anterior denture teeth, while the C-EI group had all natural anterior teeth; however, the teeth were either stained, crowded, or with midline diastemata or a gummy smile. It was assumed that maxillary anterior tooth loss had a higher impact on orofacial esthetics than orthodontic anomalies. The Sheffé post hoc test confirmed this assumption, showing significant differences between the esthetically normal and impaired groups and between esthetically impaired patients and controls. However, there was no significant difference between the esthetically normal patient and control groups. This was the result of the one-way ANOVA for both the OES (0 to 10) and OES (1 to 5) summary and average scores.

Responsiveness measures the response between two administrations of the same test, for example, a change caused by a treatment procedure. The authors assumed that after prosthodontic treatment, the OES scores would improve significantly. For that purpose, 32 esthetically impaired patients requiring prosthodontic treatment were included. Most of these patients had some missing anterior teeth and they received an FPD in the maxillary anterior region, mostly composed of metal-ceramic (all-ceramic FPDs were fabricated for only 3 subjects), or RPDs replacing posterior teeth attached to an FPD in the anterior region (metal-ceramic) with precision attachments (19 patients). Thirteen patients received new CDs. The results confirmed the satisfactory responsiveness to appropriate treatment for both OES scales (0 to 10 and 1 to 5) (see Tables 6 and 7). All questions showed significant improvement in comparison to pretreatment results ($P < .01$). According to Cohen,[32] an effect size of 0.20 is considered small, 0.50 moderate, and 0.80 large. The effect size was large for both OES (0 to 10 or 1 to 5) summary and average scores. Responsiveness was not tested for the English OES version, so the results cannot be compared to the original version.[27,28] However, the results for responsiveness obtained in this study indicate that the OES questionnaire is suitable for monitoring the esthetic aspects of the success of prosthodontic therapy.

The results for the psychometric properties of the OES-CRO questionnaires (one using the 0 to 10 scale and another using the 1 to 5 scale) proved satisfactory psychometric properties, but the OES (1 to 5) showed better test-retest results and smaller CVs.

In cross-cultural studies, psychometric properties of a questionnaire have to be satisfactory. However, psychometric properties may depend on the scale and number of categories being used for the assessment.[39–46] The first step is to determine whether the scales are actually measuring the same concepts in all countries. One type of bias, namely response bias, may affect answers, particularly when using different rating scales.[39] For example, a respondent may tend to answer on the positive (or negative) side of a rating scale when assessing items. Although a great deal of research has been devoted to the effects of variations in rating scale formats, including differences in numbers of response categories, the issue of the optimal number of response categories in rating scales is still unresolved.[40]

Some studies found relatively constant test-retest reliabilities over scales with 2, 3, 5, and 7 response categories, relatively constant inter-rater reliability over scales with 3, 5, and 7 response categories, and a decrease in reliability for 11-point scales.[40,42] High reliability has been reported for 5-point scales.[43] It has been shown that reliability (internal consistency) and validity are improved by using 5- to 7-point scales rather than coarser ones, but more finely graded scales did not improve reliability and validity further.[44] It is also known from the literature that scales with few response categories (2 or 3) performed worse considering validity and discriminating power. A scale with relatively few response categories (2 or 3) tends to generate scores with comparatively little variance, limiting the magnitude of correlations with other scales.[40,44]

One study on respondents' preference ratings also showed substantial differences between scales. Scales with 5, 7, and 10 response categories were rated as relatively easy to use, while shorter scales with 2, 3, or even 4 response categories were rated as relatively quick to use, but they were rated extremely unfavorable on the extent to which they allowed the respondents to express their feelings adequately.[40]

In terms of the interface between the respondent and interviewer in a telephone or personal survey, there are some advantages and disadvantages of each scale format. With a 5-point scale, it is quite simple for the interviewer to read the complete list of

scale descriptors (1 = strongly disagree, 2 = disagree, etc) out loud. This clarification is lengthier for the 7-point format. Such a verbal clarification becomes quite impractical for a 10-point or greater format as the gradations of agreement become too fine to easily express in words.[44]

Finely graded scales, by definition, provide more options for the respondent. Therefore, finer scales could result in a greater spread of data.[45] This would result in a larger variance and increased CV, which is also evident from the results of this study. If scales with different numbers of scale categories are used, there are some straightforward methods by which each scale can be standardized into the other scale format. One is based on a formula used by Preston and Colman[40]:

(rating – 1) / (number of response categories – 1) × 100

Another method is that employed by Dawes,[44] whereby the scale end points for the 5- and 7-point versions are anchored to the end points of the 10-point scale. The intervening scale values are inserted at equal numeric intervals. For example, to rescale the 5-point scale to 10 points, 1 remains as 1, 5 is rescaled to 10, the mid-point of 3 on the 5-point scale is adjusted to be as per the mid-point between 1 and 10 (namely 5.5), and so on. However, rescaling may also result in a slight biasing effect.[40–44] Many textbooks suggest that the most commonly used scale is the 5-point Likert scale.[44,46]

The OES (1 to 5) is more suitable in Croatia. The first two items were modified in the Croatian translation to allow for better comprehension. It was added that the first two questions were related to the lower third of the face. The authors also recommend such a change for the English version of the OES.

The authors also recommend the OES (1 to 5) scale for international use, since it is the most used scale worldwide. Moreover, for many countries, the 11-point scale might be too broad, producing more variability and bias rather than improving discriminating power. It is also completely impractical for personal and telephone interviews.

Conclusion

The Croatian version of the OES questionnaire showed good psychometric properties. The first two items had to be modified for better understanding by adding a remark that they were related to the lower third of the face, which should also be done in the English version. The test-retest reliability was better using the 5-point scale in comparison to the 11-point scale. Because of the ease of use and good psychometric properties, the authors recommend the 5-point scale for international use. The results for responsiveness render the OES questionnaire useful in monitoring the esthetic impacts of prosthodontic therapy.

Acknowledgments

This study was supported by grant no. 065-0650446-0420 and grant no. 065-0650446-0435 of the Croatian Ministry of Science, Education, and Sports.

References

1. York J, Holtzman J. Facial attractiveness and the aged. Spec Care Dentits 1999;19:84–88.
2. Newton JT, Prabhu N, Robinson PG. The impact of dental appearance on the appraisal of personal characteristics. Int J Prosthodont 2003;16:429–434.
3. Davis LG, Ashworth PD, Spriggs LS. Psychosocial effects of aesthetic dental treatment. J Dent 1998;26:547–554.
4. Celebić A, Knezović-Zlatarić D, Papić M, Carek V, Baucić I, Stipetić J. Factors related to patient satisfaction with complete denture therapy. J Gerontol A Biol Sci Med Sci 2003;58: M948–953.
5. Marachlioglou CR, Dos Santos JF, Cunha VP, Marchini L. Expectations and final evaluation of complete dentures by patients, dentist and dental technician. J Oral Rehabil 2010; 37:518–524.
6. Zlatarić DK, Celebić A. Factors related to patients' general satisfaction with removable partial dentures: A stepwise multiple regression analysis. Int J Prosthodont 2008;21:86–88.
7. Zlatarić DK, Kristek E, Celebić A. Analysis of width/length ratios of normal clinical crowns of the maxillary anterior dentition: Correlation between dental proportions and facial measurements. Int J Prosthodont 2007;20:313–315.
8. Ellis JS, Pelekis ND, Thomason JM. Conventional rehabilitation of edentulous patients: The impact on oral health-related quality of life and patient satisfaction. J Prosthodont 2007;16:37–42.
9. Brisman A. Esthetics: A comparison of dentists' and patients' concepts. J Am Dent Assoc 1980;100:345–352.
10. Kokich VO, Kokich VG, Kiyak HA. Perceptions of dental professionals and laypersons to altered dental esthetics: Asymmetric and symmetric situations. Am J Orthod Dentofacial Orthop 2006;130:141–151.
11. Kokich VO Jr, Kiyak HA, Shapiro PA. Comparing the perceptions of dentists and lay people to altered dental esthetics. J Esthet Dent 1999;11:311–324.
12. LaVacca MI, Tarnow DP, Cisneros GJ. Interdental papilla length and the perception of aesthetics. Pract Proced Aesthet Dent 2005;17:405–412.
13. Gracco A, Cozzani M, D'Elia L, Manfrini M, Peverada C, Siciliani G. The smile buccal corridors: Aesthetic values for dentists and laypersons. Prog Orthod 2006;7:56–65.
14. Celebić A, Knezović-Zlatarić D. A comparison of patient's satisfaction between complete and partial removable denture wearers. J Dent 2003;31:445–451.
15. Montero J, López, JF, Galindo MP, Vicente P, Bravo M. Impact of prosthodontic status on oral wellbeing: A cross-sectional cohort study. J Oral Rehabil 2009;36:592–600.
16. Turker SB, Sener ID, Ozkan YK. Satisfaction of the complete denture wearers related to various factors. Arch Gerontol Geriatr 2009;49:e126–e129.

17. Larsson P. Methodological studies of orofacial aesthetics, orofacial function and oral health-related quality of life. Swed Dent J Suppl 2010;(204):11–98.
18. Schlosser JB, Preston CB, Lampasso J. The effects of computer-aided anteroposterior maxillary incisor movement on ratings of facial attractiveness. Am J Orthod Dentofacial Orthop 2005;127:17–24.
19. Dong JK, Rashid RG, Rosenstiel SF. Smile arcs of Caucasian and Korean youth. Int J Prosthodont 2009;22:290–292.
20. Ibrahimagić L, Jerolimov V, Celebić A, Carek V, Baucić I, Zlatarić DK. Relationship between the face and the tooth form. Coll Antropol 2001;25:619–626.
21. Petricević N, Stipetić J, Antonić R, et al. Relations between anterior permanent teeth, dental arches and hard palate. Coll Antropol 2008;32:1099–1104.
22. Petricević N, Katunarić M, Mehulić K, Simeon P, Rener-Sitar K, Celebić A. Selection of appropriate artificial frontal teeth size using dimensions of hard palate. Coll Antropol 2006; 30:573–577.
23. Celebić A, Stipetić J, Nola P, Petricević N, Papić M. Use of digital photographs for artificial tooth selection. Coll Antropol 2004;28:857–863.
24. Priest G, Priest J. Promoting esthetic procedures in the prosthodontic practice. J Prosthodont 2004;13:111–117.
25. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. Community Dent Health 1994;11:3–11.
26. Mehl C, Kern M, Freita-Wolf S, Wolfart M, Brunzel S, Wolfart S. Does the Oral Health Impact Profile questionnaire meassure dental appearance? Int J Prosthodont 2009;22:87–93.
27. Larsson P, John MT, Nilner K, Bondemark L, List T. Development of an Orofacial Esthetic Scale in prosthodontic patients. Int J Prosthodont 2010;23:249–256.
28. Larsson P, John MT, Nilner K, List T. Reliability and validity of the Orofacial Esthetic Scale in prosthodontic patients. Int J Prosthodont 2010;23:257–262.
29. Rener-Sitar K, Celebić A, Petricević N, et al. The Slovenian version of the Oral Health Impact Profile Questionnaire (OHIP-SVN): Translation and psychometric properties. Coll Antropol 2009;33:1177–1183.
30. Petricević N, Celebić A, Papić M, Rener-Sitar K. The Croatian version of the Oral Health Impact Profile Questionnaire. Coll Antropol 2009;33:841–847.
31. Rener-Sitar K, Petricević N, Celebić A, Marion L. Psychometric properties of Croatian and Slovenian short form of oral health impact profile questionnaires. Croat Med J 2008;49:536–544.
32. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NY: Lawrence Erlbaum, 1988.
33. Shrout PE, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.
35. Allen PF, McMillan AS, Locker D. An assessment of sensitivity to change of the Oral Health Impact Profile in a clinical trial. Community Dent Oral Epidemiol 2001;29:175–182.
36. Wolfart S, Thormann H, Freitag S, Kern M. Assessment of dental appearance following changes in incisor proportions. Eur J Oral Sci 2005;113:159–165.
37. Streiner D, Norman G. Health Measurement Scales—A Practical Guide to Their Development and Use. Oxford: Oxford University Press, 2008.
38. Groth-Marnat G. Handbook of psychological assessment, ed 4. Hoboken, New Jersey: Wiley & Sons, 2003.
39. Van Herk H, Poortinga Y, Verhallen TMM. Response styles in rating scales: Evidence of method bias in data from six EU countries. J Cross-Cult Psychol 2004;35:346–360.
40. Preston CC, Colman AM. Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst) 2000;104:1–15.
41. Bendig AW. Reliability and the number of rating scale categories. J Appl Psychol 1954;38:38–40.
42. Brown G, Wilding RE, Coulter RL. Customer evaluation of retail salespeople using the SOCO scale: A replication, extension, and application. J Acad Market Sci 1991;9:347–351.
43. Lissitz RW, Green SB. Effect of the number of scale points on reliability: A Monte Carlo approach. J Appl Psychol 1975;60:10–13.
44. Dawes J. Do data characteristics change according to the number of scale points used? Int J Market Res 2008;50:61–77.
45. Finn RH. Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. Educ Psychol Meas 1972;32:255–265.
46. Malhotra N, Peterson M. Basic Marketing Research: A Decision-Making Approach, ed 2. New Jersey: Prentice Hall, 2006.

Literature Abstract

Estimating bisphosphonate use and fracture reduction among US women aged 45 years and older, 2001–2008

This study estimated the number of postmenopausal women treated and fractures reduced with oral-administered bisphosphonates in the United States during the time period from 2001 to 2008. Oral bisphosphonates are a first-line therapy for prevention and treatment of postmenopausal osteoporosis. No studies have determined the number of women treated and fractures prevented over the years, although bisphosphonate treatment has been shown to reduce fractures in randomized controlled trials. Two medical claims databases from the years 2001 to 2008 were combined to determine the number of women aged 45 years and older filling prescriptions for bisphosphonates by time-dependent medication possession ratios (MPRs): < 50%, 50% to 79%, and ≥ 80%. Fracture incidence was compared for each cohort by MPR category relative to the referent (untreated) cohort with MPR < 50%. Extrapolation to the US female population treated with oral bisphosphonates on fracture rates was applied by MPR category over the 8-year period. The results showed that 460,584 women aged 45 years and older in the databases received treatment with oral bisphosphonates from 2001 to 2008, with a mean follow-up time of 2.4 years. Overall, fracture rates declined with improved MPR from 1.52% to 1.18% for age 45 to 64 and from 5.12% to 3.75% for age 65 and older. Extrapolating to the US population of female bisphosphonate users, over 27.9 million person-years of bisphosphonate treatment with 50% or greater MPR were estimated, and 144,671 fractures were prevented. The authors concluded that treatment with oral bisphosphonates has prevented a substantial number of fractures. Risks of fractures would have been reduced if identification, treatment, and compliance were improved.

Siris ES, Pasquale MK, Wang Y, Watts NB. *J Bone Miner Res* 2011;26:3–11. **References:** 18. **Reprints:** Dr Ethel Siris, Toni Stabile Osteoporosis Center, Department of Medicine, Columbia University Medical Center, 180 Fort Washinington Ave 9-964, New York, NY 10032—*Arthur S. Sham, Hong Kong*