INTRODUCTION

Prosthodontic patients constitute a unique cohort within dental care necessitate tailored treatment plans to address complex dental and facial concerns effectively. Understanding the specific needs and intricacies of prosthodontic patients is fundamental in delivering effective and personalized care within this specialized domain of dentistry. The significance of esthetics in prosthodontic care is closely linked to an individual's self-confidence, psychological outlook, and, ultimately, their overall quality of life. As a result, assessing esthetics during oral reconstruction should not solely rely on individual preferences, whether from the clinician or patient.

The patient’s subjective assessment of their oral health, including their personal perception of esthetics, holds considerable importance in evaluating treatment effectiveness. This subjective aspect poses a challenge in assessing esthetics because the ultimate outcome needs to align with the patient’s specific esthetic preferences, which should be one of the treatment outcomes...
measured in prosthodontic care. To streamline and enhance this process, there is a critical need for valid and reliable tools to standardize and expedite the evaluation of esthetics. Therefore, Mursid et al., conducted a scoping review to assess various instruments used for esthetics and concluded that OES is the most widely used instrument for self-evaluation in orofacial esthetics research.4

The Orofacial Esthetic Scale (OES) is a validated tool used to assess and measure the esthetic perceptions of individuals regarding various aspects of the oral and facial region. The 8-item instrument have seven questions directly asking on esthetics pertaining to face, facial profile, mouth, rows of teeth, tooth shape/form, tooth color, gums and one global assessment question about the overall perception towards the face, mouth, and teeth.5 It measures the individual’s orofacial esthetic perception using an 11-point Likert scale (0-70). The first seven items are combined into a summary score ranging from 0 (worst score, patients are very dissatisfied with all esthetic items) to 70 (best score, patients are very satisfied with all esthetic items).6

The questionnaire was developed based on applying broad esthetic principles to the stomatognathic system. OES, which was originally developed for prosthodontic patients7 was further tested on general population by OES, which was originally developed for prosthodontic patients. Numerous authors have translated OES into several languages, and some have used it as an intervention tool. The absence of validated orofacial esthetics assessment tools in prosthodontic care impedes communication and personalized treatment planning. Thus, this study aimed to evaluate the psychometric properties of the Orofacial Esthetic Scale (OES) within prosthodontic patients, ensuring its reliability and applicability for quantifying subjective esthetic perceptions in this clinical context.

METHODS

The methodology was based on the guidance by PRISMA and Cochrane Handbook for Systematic Review.8 The focus question was formulated to determine the overall assessment score in prosthodontic patients based on “PICO” as follow:

"Is Orofacial Esthetic Scale a valid and reliable instrument to use for prosthodontic patients?"

P = prosthodontic patients
I = OES questionnaire
C = psychometric scores of OES
O = measure validity and reliability of OES

The authors extracted articles from the PubMed/ Medline, Scopus, Web of Science, Science Direct databases using the meSH terms and filtered them from 2010 to 2023. The detailed search strategy for PubMed/ Medline database in advanced search was as follows: 1 = “Orofacial esthetic” OR “Orofacial aesthetic”, 2 = “Scale” OR “Questionnaire” OR “Survey” OR “Development” OR “Validation” OR “Psychometric”, and then combining #1 AND #2, using the filters “2010-2023” and “English”. Each database was last searched on 30 October 2023.

After initial identification of 2044 related publications through databases, duplicates were removed using EndNote (version X9; Thomson Reuters, New York, NY). References were then manually cross-reviewed for each title and abstract to identify studies fulfilling the set inclusion/exclusion criterion by two authors. In addition to database searches, a thorough grey literature search was conducted. Authors also scanned reference lists in articles and reviews, and related database entries.

Inclusion criteria:

• Studies that have used OES questionnaire in prosthodontic patients.
• Studies that assessed reliability and validity of OES in prosthodontic patients.
• All study designs were included.
• Studies published since 2010.
• Studies written in English.
• Studies that have been published.

Exclusion criteria:

• Studies that compared the validity of OES with a distinct tool that measured orofacial esthetics.
• Review articles, proposal papers and articles whose full texts were not available.
• Reliability and validity values were digitized. Internal consistency was assessed using Cronbach’s alpha, while test-retest reliability employed interclass correlation (ICC). Convergent validity was determined through correlation coefficients with OES comparator measures, and construct validity was analyzed via factor analysis, KMO values, and the number of factors.

Assessment of the quality and bias risk: Two authors independently assessed bias using the Joanna Briggs Institute (JBI) critical appraisal checklists. Specifically, the JBI checklist for case control studies was employed. Items were rated as “yes,” “no,” “unclear,” or “inapplicable.” Studies scoring over 70% “yes” were deemed low bias risk, 50-69% moderate, and below 49% high.9

Statistical analysis of the included studies: Meta-analysis was conducted on Cronbach’s alpha and ICC reliability scores. Convergent validity couldn’t undergo meta-analysis due to differing survey items. Construct validity was qualitatively assessed due to lack of recognized methodology. Content and discriminative analysis were performed using MedCalc-version 19.5.3 which calculates the weighted summary correlation coefficient under the random effects model using the Hedges-Olkin method and a Fisher Z transformation of the correlation coefficients. Cochrane guidelines interpret heterogeneity as follows: 0-40% not significant, 30-60% moderate, 50-90% substantial, 75-100% considerable.10

RESULTS

A flowchart describing the systematic review search results is presented in Fig.1. Of the 2044 articles, seven peer-reviewed ones were selected, all with a case-control design. They involved 1144 prosthodontic
Fig.1: Flowchart of the selection of studies for the review.
patients. For the psychometric evaluation, seven articles tested for Cronbach alpha, six for test-retest, five for construct validity, seven for convergent validity, six for discriminant validity and two for content validity.

For convergent validity five studies compared three items of OHIP questionnaire with $r = 0.43$ to $0.72$, $0.73$ to $0.81$, $0.83$ to $0.70$, $0.65$ to $0.88$ and $0.67$ to $0.81$, respectively and reported significant p value. The study by Reissmann et al.$^{15}$ calculated the correlation between the instrument’s summary score with the patients’ global assessment of their esthetics ($r = 0.87$) and external ratings of an expert group ($r = 0.55$). The study by Edoardo et al.$^{16}$ calculated the correlation between OES scores of prosthodontic patients and clinicians with $r = 0.93$. Construct validity was calculated by factor analysis; one study$^{12}$ used both confirmatory factor analysis (CFA) and exploratory factor analysis (EFA), with the factor loading from $0.892$ to $0.933$. Three studies used EFA only, with factor loading of $0.45$ to $0.80$, $0.74$ to $0.94$ respectively and the values for one article$^{13}$ were not given, while one study$^{16}$ used CFA only. All the studies revealed unidimensional

Table-I: Validity parameters of Orofacial Esthetic Scale tool administered to identify patient dental esthetics needs.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Content validity</th>
<th>Concurrent validity</th>
<th>Construct validity</th>
<th>Convergent validity</th>
<th>Discriminative validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larsson, P (2010)$^{10}$</td>
<td>119</td>
<td>NA</td>
<td>NA</td>
<td>Spearman rank correlation</td>
<td>t-test</td>
</tr>
<tr>
<td>Persic, S. (2011)$^{18}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Spearman rank correlation</td>
<td>one-way ANOVA &amp; Sheffe post hoc test</td>
</tr>
<tr>
<td>Zhao, Y. (2013)$^{19}$</td>
<td>4 experts</td>
<td>NA</td>
<td>CFA, EFA KMO=0.939, one factor model</td>
<td>Pearson correlation coefficient</td>
<td>one-way ANOVA &amp; Sheffe post hoc test</td>
</tr>
<tr>
<td>Bimbashi, V. (2015)$^{20}$</td>
<td>NA</td>
<td>NA</td>
<td>EFA, KMO=0.921, one factor model</td>
<td>Spearman rank correlation</td>
<td>One way ANOVA</td>
</tr>
<tr>
<td>Reissmann, DR. (2015)$^{21}$</td>
<td>NA</td>
<td>NA</td>
<td>EFA one factor model</td>
<td>Pearson correlation coefficient</td>
<td>two sample t tests Cohen’s d, a standardized effect size</td>
</tr>
<tr>
<td>Alhajj, M.N. (2017)$^{22}$</td>
<td>NA</td>
<td>NA</td>
<td>EFA One factor model, KMO=0.87</td>
<td>Spearman’s rank correlations</td>
<td>one-way ANOVA &amp; Sheffe post hoc test</td>
</tr>
<tr>
<td>Edoardo, R. (2023)$^{24}$</td>
<td>NA</td>
<td>NA</td>
<td>CFA</td>
<td>Pearson correlation coefficient</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA – not available.

Table-II: Risk of bias using Joanna Briggs Institute Critical Assessment Checklist for Case control studies.

<table>
<thead>
<tr>
<th>STUDY ID</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larsson, P (2010)$^{10}$</td>
<td>70.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persic, S. (2011)$^{18}$</td>
<td>70.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao, Y. (2013)$^{19}$</td>
<td>70.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bimbashi, V. (2015)$^{20}$</td>
<td>40.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reissmann, D. R. (2015)$^{21}$</td>
<td>60.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alhajj, M.N. (2017)$^{22}$</td>
<td>60.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edoardo, R. (2023)$^{24}$</td>
<td>60.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q1. Were the groups comparable other than the presence of disease in cases or the absence of disease in controls?
Q2. Were cases and controls matched appropriately?
Q3. Were the same criteria used for identification of cases and controls?
Q4. Was exposure measured in a standard, valid and reliable way?
Q5. Was exposure measured in the same way for cases and controls?
Q6. Were confounding factors identified?
Q7. Were strategies to deal with confounding factors stated?
Q8. Were outcomes assessed in a standard, valid and reliable way for cases and controls?
Q9. Was the exposure period of interest long enough to be meaningful?
Q10. Was appropriate statistical analysis used?
structure. Discriminant validity was calculated in four studies\textsuperscript{11,14} using one-way analysis of variance (ANOVA) and Sheffe post hoc test while two studies\textsuperscript{6,15} calculated by two sample t-test. Only two studies calculated content validity: one study\textsuperscript{6} with 119 respondents and the other\textsuperscript{12} with a sample size of four experts. Table-I summarized the types of validity calculated in each article.

The JBI assessment checklist (Table-II) shows that three studies\textsuperscript{6,11,12} were ranked as low risk of bias, three\textsuperscript{6,14-16} were moderate and only one study\textsuperscript{13} was high. The domains with the greatest number of methodological flaws were related to the control of confounding factors and strategies to deal with them. However, almost all the studies used appropriate statistical analysis according to their study requirements.

Among the seven studies that calculated psychometric properties of the OES on the prosthodontic patients, only six (of which one study calculated two values of ICC) with the same study design provided information on the ICC scores with overall ICC of 0.95 (95\% CI: 0.94 to 0.96) and seven studies reported Cronbach alpha values with overall value of 0.94 (95\% CI: 0.92 to 0.96), indicating excellent internal consistency. Heterogeneity was significant in both analyses (p = <.0001), with OES score variations of 76.32\% for ICC group and 88.41\% for Cronbach alpha group respectively. Bias due to small

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{forest_plot.png}
\caption{Forest plot depicting the results of meta-analysis for, (A) studies including ICC values. (B) studies including Cronbach alpha values.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{funnel_plot.png}
\caption{Funnel plot for evaluation of publication bias for, (A) studies with ICC values. (B) Studies with Cronbach alpha values.}
\end{figure}
study effect was tested using Begg’s test and Egger’s test. The two tests showed no evidence of bias (P = 0.60 and 0.36 for the ICC group and 0.51 and 0.51 for the Cronbach alpha group). Funnel plots showed no publication bias in both meta-analyses (Fig 3).

**DISCUSSION**

This systematic review and meta-analysis thoroughly evaluated the OES among prosthodontic patients. Seven articles, all with a case-control design, were included, comprising 1144 patients. Psychometric evaluation covered Cronbach’s alpha, test-retest reliability, construct validity, convergent validity, discriminant validity, and content validity.

This study revealed that four studies were rated as low risk of bias, indicating high-quality, while three were moderate, and only one recorded with the lowest score of 40%. Overall, evidence supports OES validity and reliability in assessing orofacial esthetics among prosthodontic patients. This review qualitatively analyzed OES validity methods across articles, confirming construct validity through structural, convergent, and known-groups validity. All items were good indicators for the measured construct, correlated well with other measures of the same construct, and differentiated well between subjects with known differences.

Additionally, these publications supported OES unidimensionality, adhering to traditional methods of sample size involving 5-20 respondents per item for the factor analysis. The pooled ICC and Cronbach fell within the recommended ranges 0.70 – 0.95 values, signify substantial agreement and coherence within the scale items, validating its application in clinical and research settings. The meta-analyses indicated significant heterogeneity among studies, reflecting variations across population and countries. This underscores the need for cross-cultural adaptation. Moreover, absence of publication bias enhances credibility of findings within the published literature.

**Limitation:** The limitation of this review includes reliance on a restricted number of studies for meta-analysis, necessitating updates with more data in the future. Additionally, English-only search criteria may introduce language bias, potentially omitting other relevant evidence. The strength of this review lies in it being the first meta-analysis to explore orofacial esthetics perception globally among prosthodontic patients using the OES questionnaire. Dentists can utilize this tool for better treatment planning and understanding patient needs and perceptions.

**CONCLUSION**

This comprehensive analysis consolidates the evidence supporting the OES as a valid and reliable instrument for assessing orofacial esthetics in prosthodontic patients. The robust psychometric properties, highlighted by the high ICC and Cronbach’s alpha values, signify its potential utility in both research and clinical settings. Future studies might benefit from exploring the sources of heterogeneity observed and validating the scale in diverse populations to enhance its applicability and generalizability. The findings also provide valuable insights into the assessment of orofacial esthetics and lay a strong foundation for employing the OES as a standardized tool in prosthodontic care, facilitating better treatment planning and patient-centered care.

**Competing financial interests statement:** None.

**Funding:** None.

**REFERENCES**


**Authors’ Contribution:**

ZA: Conceived and designed the study, collected the data and performed data analysis, prepared the manuscript.

Jj, MD & ZY: Conceived and designed the study, helped in analyzing the data, reviewed and provided final approval.

All authors are responsible for the accuracy and integrity of the manuscript.