Reproducibility of the Pink Esthetic Score—Rating Soft Tissue Esthetics Around Single-Implant Restorations with Regard to Dental Observer Specialization

PETER GEHRKE, DR. MED. DENT.*
MARKUS LOBERT, DR. MED. DENT.†
GUNTER DHOM, DR. MED. DENT.‡

ABSTRACT

Purpose: The pink esthetic score (PES) evaluates the esthetic outcome of soft tissue around implant-supported single crowns in the anterior zone by awarding seven points for the mesial and distal papilla, soft-tissue level, soft-tissue contour, soft-tissue color, soft-tissue texture, and alveolar process deficiency. The aim of this study was to measure the reproducibility of the PES and assess the influence exerted by the examiner’s degree of dental specialization.

Materials and Methods: Fifteen examiners (three general dentists, three oral maxillofacial surgeons, three orthodontists, three postgraduate students in implant dentistry, and three lay people) applied the PES to 30 implant-supported single restorations twice at an interval of 4 weeks. Using a 0–1–2 scoring system, 0 being the lowest, 2 being the highest value, the maximum achievable PES was 14. At the second assessment, the photographs were scored in reverse order. Differences between the two assessments were evaluated with the Spearman’s rank correlation coefficient ($R$). The Wilcoxon signed-rank test was used for comparisons of differences between the ratings. A significance level of $p < 0.05$ was chosen for both tests.

Results: Observer results indicated that the agreement between the first and second rating for all occupational groups was 70.5%, with a broad correlation between the two ratings and a high statistical significance (Spearman’s $R = 0.58$, $p = 0$; Wilcoxon $T = 163,182$, $Z = 3.383599$, $p = 0.000716$). The most agreement between the first and second rating was obtained by orthodontists with 73.5% ($R = 0.67$), and the least by lay people 65.9% ($R = 0.50$). Very poor and very esthetic restorations showed the smallest deviations. Orthodontists were found to have assigned significantly poorer ratings than any other group. The assessment of postgraduate students and laypersons were the most favorable.

CLINICAL SIGNIFICANCE

The PES allows for a more objective appraisal of the esthetic short- and long-term results of various surgical and prosthetic implant procedures. It reproducibly evaluates the peri-implant soft tissue around single-implant restorations and results in good intra-examiner agreement. However, an effect of observer specialization on rating soft-tissue esthetics can be shown.

INTRODUCTION

Ossseointegration is no longer a possibility but rather a given in implant dentistry today. However, with this predictability that one can expect has also come a shift into esthetic concerns. As a genuine treatment alternative, implant-supported restorations should conform to the good esthetic outcome of conventional crown and bridge technique or provide a better outcome. The condition of the peri-implant soft tissues appears to be a critical determinant.\(^1\)\(^-\)\(^3\) Although criteria concerning the functional assessment of implants (stability, radiographic bone loss, prosthetic complications, and peri-implant hygiene\(^4\)\(^-\)\(^9\)) are prevalently employed for the determination of implant success, the use of newer indices for objective evaluation of soft-tissue esthetics must, on the other hand, be checked for validity.\(^1\)\(^3\) Measured by the abundance of implant dentistry publications that are chiefly concerned with osseointegration processes, clinical success rates, and many examples of top-quality restorative techniques, few studies are concerned with the esthetic parameters of implant restorations.\(^1\)\(^0\),\(^1\)\(^1\) There is a lack of objective methods of measurement in order to assess esthetic quality. Parameter of anterior implant success, such as maintenance or reestablishment of harmoniously scalloped peri-implant mucosa lines and natural contours, should be included in future studies.\(^1\)\(^2\)

In order to evaluate and record esthetics, a fundamental distinction may be drawn between subjective and objective methods. One subjective method is for the patient to answer questionnaires in which he or she can express his or her satisfaction and any deficiencies that may exist. However, this subjective assessment is not suitable for evaluating any potential sources of error or scope for improvement in restoration.\(^1\)\(^3\) Objective methods by a professional examiner based on defined criteria are rare in the field of esthetic implant dentistry. Fürhauser and colleagues\(^1\)\(^4\) presented a rating matrix for evaluating the soft tissue around single implant restorations. This pink esthetic score (PES)\(^1\)\(^4\) evaluates the esthetic outcome of soft tissue around implant-supported single crowns in the anterior zone by awarding seven points for the mesial and distal papilla, soft-tissue level, soft-tissue contour, soft-tissue color, soft-tissue texture, and alveolar process deficiency (Figures 1 to 4). With the exception of papilla formation, the evaluation is performed by visually comparing reference teeth (i.e., with the contralateral tooth in the incisor zone and adjacent tooth in the premolar zone). For the mesial and distal papilla, the categories are complete, incomplete, and absent. For each criterion it is possible to award a score between two points (for a very good outcome) and no points (for a poor outcome). The maximum score that can be achieved, 14 points, indicates an outcome that reflects complete conformity between the soft tissue of the tooth being assessed and that of the reference tooth. The PES is thus designed to allow reproducible evaluation of soft tissue around single-tooth implants.\(^1\)\(^4\) Studies concerning the validity and reproducibility of the PES have to be performed. The question of whether assessment of the esthetic outcome of the peri-implant mucosa is subject to the subjective appraisal of the observer or whether there are objective, comprehensible rules of evaluation remains to be clarified.

The aim of this study is to measure the reproducibility of the PES and assess the influence exerted by the examiner’s degree of dental specialization.

MATERIALS AND METHODS

Standardized intraoral photographs of 30 patients with maxillary anterior implant-supported single crowns and adjacent peri-implant soft tissue were evaluated (XiVE®, Dentsply Friadent, Mannheim, Germany). All images were available as black-and-white and color...
prints. Seven parameters influencing the esthetic outcome of peri-implant mucosa were selected for questioning purposes: mesial and distal papilla, soft-tissue level, soft-tissue contour, soft-tissue color, soft-tissue texture, and alveolar process deficiency. Three general dentists, three oral maxillofacial surgeons, three orthodontists, three postgraduate students in implant dentistry, and three lay people evaluated the peri-implant mucosa of the 30 implant-supported single restorations by providing seven evaluation index points. The grading used the PES suggested by Fürhauser and colleagues \cite{14} and was performed twice by each examiner at an interval of 4 weeks. The contralateral tooth and the adjacent soft tissue served as reference. Using a 0–1–2 scoring system, with 0 being the lowest and 2 being the highest value, the maximum achievable PES was 14. At the second assessment,
the photographs were scored in reverse order.

**STATISTICAL ANALYSIS**

Differences between the two assessments were evaluated with the Spearman’s rank correlation coefficient ($R$). The Wilcoxon signed-rank test was used for comparisons of differences between the ratings. The significance level of both tests was set at $p < 0.05$. The statistical analyses were performed using the software package STATISTICA 7.1 (StatSoft, Inc., Tulsa, OK, USA).

**RESULTS**

The assessments of 30 images by 15 examiners applying seven evaluation criteria took place at two points in time. In isolated cases there were missing values because of a lack of assessments for individual criteria. Consequently, 3,150
single PES variables were available for analysis of the first assessment, and 2,940 for analysis of the second assessment. The mean PES was 9.28 for the first assessment and 9.57 for the second assessment. The difference of the two assessments was statistically significant (Wilcoxon: \( p = 0.0082 \)) as shown in Table 1.

**Agreement of First and Second Ratings**

Intra-observer results indicated that the agreement between the first and second rating for all occupational groups was 70.5%. An identical assessment for the first and second rating has been given in 2,073 cases (Table 2). A broad correlation between the two ratings (\( R = 0.58 \)) with high statistical significance (\( p = 0 \)) could be identified (Table 3). In 489 cases (16.6%) a higher PES score has been assigned in the second rating. A lower score was given in 378 (12.9%) of the cases. This difference was statistically significant according to the Wilcoxon signed-rank test (\( p = 0.000716 \)) (Table 4). The most agreement between the first and second rating was obtained by orthodontists with 73.5% (\( R = 0.67 \)), and the least by lay people 65.9% (\( R = 0.50 \)) (Table 3). An analysis of the agreement showed no statistically significant difference between the two assessments, except for the lay people (\( p < 0.0001 \)).

**Single-Variable Analysis**

The highest possible score of 2 was most assigned to a restored alveolar process (55.71 and 60% of cases), as shown in Table 5. Peri-implant soft-tissue color (34.52 and 33.81% of cases) and texture (38.57 and 40.48% of cases) were assigned significantly poorer ratings. Consequently, a score of 0 was most often assigned to the peri-implant soft-tissue texture (20 and 15.95% of cases) and soft-tissue color (19.76 and 15.71% of cases) as shown in Table 6. An analysis of all available PESs demonstrated that the highest achievable score of 14 was assigned 42 times in the first assessment (10%) versus 32 times in the second assessment (7.62%) (Table 7). The lowest score of 0 was assigned four times in the first assessment (0.95%) versus three times in the second assessment. Very poor and very esthetic restorations showed the smallest deviations. Orthodontists were found to have assigned significantly poorer ratings than any other group. The assessment of postgraduate students and laypersons were the most favorable (Table 8).

**DISCUSSION**

Although within the context of a rise in esthetic awareness and the resulting standards expected by patients and dental professionals, there is still a lack of comparative clinical studies on the long-term esthetic outcome of implant-supported restorations. The range of subjective opinions on the part of the observer when evaluating esthetics is known. Studies demonstrated the influence of the individual notion of esthetics in relation to the examiner’s degree of specialization and called for standardization criteria for evaluating the esthetics of single-tooth restorations.12,16,17 Objective rating systems for conventional crown and bridge prostheses cannot be applied to implant-supported dental restorations, as they incorporate factors that are of no relevance to implant dentistry.18–22 In the relevant guidelines concerning dental implants only functional parameters are used as assessment criteria.23
Chang and colleagues interviewed dentists and patients to evaluate the esthetic outcome of implant-supported single-tooth restorations using standardized questionnaires. The results confirm a high level of patient satisfaction with the outcome of treatment. The picture was contradictory in the assessment of these patient cases by prosthodontists, who assigned the treatment outcome a much lower rating than the patients. The author interpreted this result to the effect that the clinicians are either more critical or they apply different standards when assessing esthetic outcome from those applied by the patients involved themselves. In actual fact, a statistical analysis indicated that parameters such as crown shape, contact point position, color, and topography of the surrounding soft tissue had a significant influence on the rating of general satisfaction with appearance, although it was not

<table>
<thead>
<tr>
<th>TABLE 2. CORRELATION BETWEEN FIRST AND SECOND ASSESSMENT FOR ALL EXAMINERS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>First assessment</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>Ratio (%)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>Ratio (%)</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>Ratio (%)</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Ratio (%)</td>
</tr>
</tbody>
</table>

Observer results indicated that the agreement between the first and second ratings for all occupational groups was 70.5% (highlighted in gray).

<table>
<thead>
<tr>
<th>TABLE 3. SPEARMAN’S RANK CORRELATION COEFFICIENT FOR THE FIRST AND SECOND ASSESSMENTS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman’s rank correlation coefficient</td>
</tr>
<tr>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Ratings 1 and 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 4. WILCOXON SIGNED-RANK TEST FOR THE FIRST AND SECOND ASSESSMENTS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilcoxon signed-rank test</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>Ratings 1 and 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 5. SINGLE-VARIABLE ANALYSIS FOR THE HIGHEST POSSIBLE SCORE OF PINK ESTHETIC SCORE 2 (GOOD ESTHETIC RESULT).</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Assessment</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>N %</td>
</tr>
<tr>
<td>Alveolar process deficiency</td>
</tr>
<tr>
<td>Distal papilla</td>
</tr>
<tr>
<td>Soft-tissue contour</td>
</tr>
<tr>
<td>Soft-tissue level</td>
</tr>
<tr>
<td>Mesial papilla</td>
</tr>
<tr>
<td>Soft-tissue color</td>
</tr>
<tr>
<td>Soft-tissue texture</td>
</tr>
</tbody>
</table>

© 2008, COPYRIGHT THE AUTHORS
possible to detect any similar connections within patient questioning.\textsuperscript{25} Comparable differences from esthetic rating on the part of patients and clinicians have also been described in other studies.\textsuperscript{11,16} The majority show that the soft tissue between an implant-supported single-tooth reconstruction and the adjacent teeth has a substantial influence on esthetic outcome. Jemt introduced a score to be able to assess the papillary volume and the height of interproximal mucosa.\textsuperscript{26,27} However, this rating is restricted to criteria indicated and disregards the entire peri-implant tissue and the appearance of the dental restoration.

In 2005 two other rating scales in addition to the PES by Fürhauser and colleagues were presented for assessing the esthetics of implant-supported single-tooth restorations. These indices aim to allow objective appraisal of the esthetic short- and long-term results of various surgical and prosthetic implant records. Testori and colleagues\textsuperscript{28} published a case study on a patient with immediate implantation and immediate loading in the incisor zone, and in this context proposed an index for evaluating the esthetic outcome. The latter covers five parameters concerning the presence and stability of the mesiodistal papilla, buccopalatal alveolar ridge stability, the structure and color of peri-implant soft tissue, and gingival contour.\textsuperscript{28} Meijer and colleagues\textsuperscript{29} published a new index for the assessment of the esthetics of implant-supported single crowns. This Implant Crown Aesthetic Index (ICA) takes into account parameters for the evaluation of the implant crown and the surrounding peri-implant mucosa. A pilot study showed good intra-observer agreement with considerable discrepancies in the evaluation of occupational groups.\textsuperscript{29} In an evaluation of the influence of different bone augmentation techniques and materials on the esthetics of implant-supported

\begin{table}[h]
\centering
\caption{Single-variable analysis for the lowest possible score of pink esthetic score 0 (poor esthetic result).}
\begin{tabular}{|l|c|c|c|c|}
\hline
 & First assessment & Second assessment & Diff 1/2 \\
 & $N$ & $\%$ & $N$ & $\%$ & \\
\hline
Soft-tissue texture & 84 & 20 & 67 & 15.95 & $-4.05$ \\
Soft-tissue color & 83 & 19.76 & 66 & 15.71 & $-4.05$ \\
Mesial papilla & 54 & 12.86 & 37 & 8.81 & $-4.05$ \\
Alveolar process deficiency & 51 & 12.14 & 39 & 9.29 & $-2.86$ \\
Soft-tissue contour & 48 & 11.43 & 38 & 9.05 & $-2.38$ \\
Soft-tissue level & 41 & 9.76 & 31 & 7.38 & $-2.38$ \\
Distal papilla & 23 & 5.48 & 13 & 3.1 & $-2.38$ \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Summation scores for assessments 1 and 2.}
\begin{tabular}{|l|c|c|c|c|}
\hline
Pink esthetic score & Assessment 1 & Assessment 2 & Total \\
 & No. & $\%$ & No. & $\%$ & No. & $\%$ \\
\hline
0 & 4 & 0.95 & 3 & 0.71 & 7 & 0.83 \\
1 & 3 & 0.71 & 4 & 0.95 & 7 & 0.83 \\
2 & 8 & 1.9 & 3 & 0.71 & 11 & 1.31 \\
3 & 8 & 1.9 & 8 & 1.9 & 16 & 1.9 \\
4 & 4 & 0.95 & 5 & 1.19 & 9 & 1.07 \\
5 & 27 & 6.43 & 18 & 4.29 & 45 & 5.36 \\
6 & 34 & 8.1 & 22 & 5.24 & 56 & 6.67 \\
7 & 33 & 7.86 & 34 & 8.1 & 67 & 7.98 \\
8 & 44 & 10.48 & 54 & 12.86 & 98 & 11.67 \\
9 & 46 & 10.95 & 38 & 9.05 & 84 & 10 \\
10 & 35 & 8.33 & 48 & 11.43 & 83 & 9.88 \\
11 & 50 & 11.9 & 49 & 11.67 & 99 & 11.79 \\
12 & 53 & 12.62 & 60 & 14.29 & 113 & 13.45 \\
13 & 29 & 6.9 & 42 & 10 & 71 & 8.45 \\
14 & 42 & 10 & 32 & 7.62 & 74 & 8.81 \\
\hline
Total & 420 & 100 & 420 & 100 & 840 & 100 \\
\hline
\end{tabular}
\end{table}
restorations in the anterior maxilla, the ICA was deemed objective. These positive findings are inconsistent with current study results demonstrating an insufficient objective assessment of implant-supported single-tooth restorations utilizing the ICA. Deficiencies arise particularly with regard to the validity and reproducibility of results.

The PES seems to be a suitable instrument for reproducibility evaluating the soft tissue around single-implant restorations. In the present study a high agreement between two ratings for all occupational groups could be shown (70.5%). The most agreement was obtained by orthodontists, and the least by lay people. In addition, an influence of the individual perception of esthetics in relation to the examiner’s degree of dental specialization can be demonstrated. Orthodontists were found to have assigned significantly poorer ratings than any other group. The texture and color of the soft-tissue fared worst among all ratings. These variables should, therefore, be given more attention when aiming for esthetic quality. The present results correspond with the findings of Fürhauser and colleagues. A simple rating by three categories (i.e., 0–1–2) seems to be less sensitive to misjudgments than a more detailed rating, which can be problematic in the intermediate category. Although the PES was designed to assess an objective outcome of different surgical and prosthetic protocols, its rating is restricted to criteria of the peri-implant mucosa and disregards the appearance of the superstructure. However, in the relevant literature, both the appearance of the peri-implant soft tissue and the dental restoration are regarded as differentiating between a successful esthetic outcome and an unsuccessful one. Further studies comparing different rating systems and encompassing larger numbers of examiners from every dental occupational group could help in identifying objective methods of measurement in order to assess esthetic quality.

CONCLUSIONS
In the present study, the PES reproducibly evaluated the peri-implant soft tissue around single-implant restorations and resulted in good intra-examiner agreement. However, an effect of observer specialization on rating soft-tissue esthetics can be shown.

DISCLOSURE AND ACKNOWLEDGMENTS
The authors wish to express special thanks to Dr. Wolfgang Reimers for the statistical analysis of the results and to Dr. Marco Degidi for providing clinical pictures for esthetic evaluation. Dr. Gehrke occasionally lectures for Dentsply Friadent and Degudent, for which he receives honoraria.

<table>
<thead>
<tr>
<th>Occupational group</th>
<th>Agreement between first and second rating</th>
<th>Spearman’s rank correlation coefficient (R)</th>
<th>Statistical significance (p values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postgraduate students</td>
<td>72.4%</td>
<td>0.52</td>
<td>0.35</td>
</tr>
<tr>
<td>Orthodontists</td>
<td>73.5%</td>
<td>0.67</td>
<td>0.79</td>
</tr>
<tr>
<td>Lay people</td>
<td>65.9%</td>
<td>0.50</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Oral-maxillofacial surgeons</td>
<td>72.1%</td>
<td>0.59</td>
<td>0.58</td>
</tr>
<tr>
<td>General dentist</td>
<td>69.4%</td>
<td>0.62</td>
<td>0.17</td>
</tr>
</tbody>
</table>
REFERENCES


Reprint requests: Peter Gehrke, Dr. Med. Dent. Bismarckstrasse 27, 67059 Ludwigshafen, Germany; Tel.: +49(0)621-68 12 44 44; email: dr-gehrke@prof-dhom.de