Introduction

Esthetic concern is always first priority when patient considering an orthodontic treatment. Even in primary and secondary school children, 85% of them recognized the importance of well-aligned teeth for overall facial appearance [1-5]. An understanding of facial attractive perception is essential for orthodontists to address patients’ needs for better esthetic improvement. To evaluate the facial beauty, many characteristics including facial proportions and several cephalometric normal values have been proposed from anthropometric or cephalometric measurements. The neoclassical facial-proportion canons, formulated by the Renaissance scholars and artists Dürer, Alberti, Cousin, Audran, Francesca, Pacioli, Cennini, Savonarla and da Vinci, would be one of interest when concerning about facial attractive analysis [6-8]. The validity of neoclassical canons of facial proportion has been tested among North American Caucasians [6,9,10], Chinese [10-12], African Americans [13-15], Vietnamese [10], Thais [10], Turkish [16], Greece [17] and Korean [18]. They found only 16.7% of vertical facial proportion and 51.5% of horizontal facial

Female Facial Attractiveness Assessed by 2D Photography

Abstract

Background: Esthetic concern is always first priority when patient considering an orthodontic treatment. The aim of this study was to evaluate whether the perception of female facial attractiveness is consistency across gender, age and professional background.

Materials and methods: A series of 100 sets female 2D photos were projected on a screen. Each set was consisted one frontal and two laterals right and left views and was shown for 5 seconds. Raters should mark their impression of facial perception of female facial attractiveness is consistency across gender, age and professional background.

Results: High internal consistency of rating female facial attractiveness was achieved by evaluator, no matter of gender, age, or professional background. Every evaluation show central tendency and unimodal distribution regardless of the attractiveness or rater’s background. More consistency is found for the evaluation of unattractive faces than attractive faces by both hospital staff and laypeople. In the evaluation of 2D photos, females give higher score than males and the significant different was found among laypeople (p=0.011). No significant different between the rating of senior and junior raters (p=0.457 and 0.781 for hospital staff and laypeople). Hospital staff rated significant higher score than laypeople (p=0.005).

Conclusion: The Likert’s rating of 2D female facial attractiveness had central tendency and unimodal distribution regardless of attractiveness or rater’s background. The ratings for unattractive faces were more consistent than that of attractive faces. Laypeople were more critical than hospital staff in the evaluation of female facial attractiveness, especially for male laypeople.

Keywords: Facial attractiveness; Likert scale; 2D photography

Introduction

Esthetic concern is always first priority when patient considering an orthodontic treatment. Even in primary and secondary school children, 85% of them recognized the importance of well-aligned teeth for overall facial appearance [1-5]. An understanding of facial attractive perception is essential for orthodontists to address patients’ needs for better esthetic improvement. To evaluate the facial beauty, many characteristics including facial proportions and several cephalometric normal values have been proposed from anthropometric or cephalometric measurements. The neoclassical facial-proportion canons, formulated by the Renaissance scholars and artists Dürer, Alberti, Cousin, Audran, Francesca, Pacioli, Cennini, Savonarla and da Vinci, would be one of interest when concerning about facial attractive analysis [6-8]. The validity of neoclassical canons of facial proportion has been tested among North American Caucasians [6,9,10], Chinese [10-12], African Americans [13-15], Vietnamese [10], Thais [10], Turkish [16], Greece [17] and Korean [18]. They found only 16.7% of vertical facial proportion and 51.5% of horizontal facial
proportion fitted to the tested neoclassical canons respectively. This indicated that neoclassical canons were not generally applicable to the human faces.

Golden ratios are commonly implied for facial attractiveness after Ricketts [19] found 6 vertical and 5 horizontal facial proportions equaling to golden ratio [20-22]. Moss et al. showed none of the facial proportions measured from attractive models matched the golden mean [23] and Kiekens et al. found only 4 out of 19 measured facial proportions to be negative correlated to golden ratio with r less than -0.36 [24]. Even Kawakami who supported the used of golden ratio as a guide for maxillofacial surgery of Caucasians, found all of the 7 measured vertical facial proportions deviated from golden proportion in Japanese subjects [25]. Another study in Japanese population, Mizumoto et al. found while the models generally had more balanced faces, their facial measurements showed more deviated from the golden proportion compared with averaged young women [26]. Moreover, case-controlled studies did not advocate the use of golden ratio as facial attractive indicator [24,27-29].

For cephalometric measurements such as Ricketts’ E-plane, there was ethnic diversity and conflicting results. For example, while the distance of lower lips to E plane were, on average, no significant difference in the attractive profile (2.96 ± 1.89 mm) when compare with normal profile (2.73 ± 1.82 mm) in the female Italian samples [30] nor attractive female Turkish profile (-1.00 ± 2.17 mm) in comparison with unattractive samples (-3.55 ± 3.67 mm) [31], this distance is significantly larger in attractive Japanese profile (1.09 ± 1.59 mm) than in normal profile (-0.13 ± 2.51 mm) [32]. And while Oh et al. showed lower lip to E plane were negative correlated to the esthetic rating in 45 American samples (-2.9 ± 3.2 mm), the correlation was not strong (r=0.29) and there was no statistically correlated in 48 Chinese samples (0.9 ± 2.4 mm). Therefore, using Rickett’s E plane to define the attractive position of lips must be careful. As shown, most of the facial characteristics derived from anthropometric and cephalometric facial measurements cannot provide accurate indicators for facial esthetics; other method should be used when considering the facial attractive evaluation.

Langlois and Roggman rediscovered Galton’s finding in 1878 by creating the averaged composites of male and female faces with computerized method [33]. They proved that averaged composites were generally gained higher attractiveness rating score than their original individual faces. Later other researchers proved that people perceived averaged faces as attractive faces by many other different ways such as facial measurements, facial manipulation by moving the landmark points toward averaged faces, morphing the created averaged facial shape, morphing the faces through inter-pupillary distance and quantitative facial analysis [34-41]. This indicated that facial attractiveness could be sufficiently ensured by facial averageness [7,42].

Many evidences proved that some standard of beauty was set by nature [43] included infants preferring to look at faces that adults find attractive [44-47], people from different cultural background showed high agreement on which faces are attractive and which are not [38,48,49] and experimental studies proved that the time to perceived the facial attractiveness could be as short as 100 ms [50,51].

Up to present, most of anthropometric or cephalometric studies using neoclassical canons, golden ratio, or esthetic lines have been tried hard to define facial attractiveness in two dimensions (2D) but in vain. The perception of facial attractiveness should be in three-dimensional (3D). Therefore a serial 3D analysis on facial attractiveness has been carried out in the Craniofacial Center, Chang Gung Hospital, Taipei, Taiwan. However, in order to classify the 3D facial attractiveness, the validity and reliability of 2D perception of facial attractiveness should be set up first using the conventional evaluation method. This is the first part of these series to evaluate the consistency of 2D perception in female facial attractiveness according to professional background, gender, age.

**Material and Methods**

**Obtained two dimensional photos and three dimensional images**

Sets of 2D facial photos (one frontal, one right and one left lateral views) and 3D facial images in rest position were collected from female subjects at Chang Gung Memorial Hospital, Taipei, Taiwan, from 2009-2010. The 2D facial photos were taken with Nikon D300 camera (Nikon Corporation, Tokyo, Japan) with single 105 mm macro lens with an aperture of F14 speed 1/125 second from a standard distance of 1.5 meters. The background was in light blue color. Two umbrella flashes were synchronized with camera flash to reduce the background shadow. The subjects were in standing position with eyes looking forward and face in relaxed and rest position. The 3D full facial images were taken by the 3dMD cranial system (3dMD Inc., Atlanta, GA, USA) in sitting position with eyes looking forward and face in relax and rest position. The capture speed was 1.5 milliseconds per surface image.

The inclusion criteria of the samples were female, age between 20-30 year old, Chinese background, no craniofacial anomalies and no history of facial trauma. This study was concentrated on 2D facial photos. Rates were divided into groups of hospital staff and laypeople. Hospital staff was plastic surgeons, orthodontists and research assistants who work in craniofacial center, Chang Gung Memorial hospital, Taipei, Taiwan. Laypeople were non-medical students and the staff from Chang Gung University, Taoyuan, Taiwan.

During each viewing session, raters were sitting in a classroom with a big screen at front. No other specific instruction was given except to evaluate the facial esthetics. Each set of female color photos (one frontal, one right and one left lateral views), were projected on a screen by PowerPoint for 5 seconds. Total 100 sets of 2D facial photos were randomly arranged without any order of attractiveness. In the next 3 seconds, the photos disappeared from the screen and the raters marked their impression of facial attractiveness on a 5-point Likert scale varied from the most unattractive as 1 to the most attractive as 5. All raters had to turn off their cell phone and computer while rating so that the
whole session took 13 minutes 20 seconds to complete without any interruption. Different 100 sets of photos were separately evaluated by hospital staff and laypeople. However, 54 photos were evaluated by both hospital staff and laypeople. To evaluate the intra-rater reliability, 1 set and 6 sets of photographs were duplicated in evaluation by hospital staff and laypeople respectively and raters were not told that there were duplicate images during the evaluation.

**Statistics**

**Outliers:** The outliers were removed before data analysis was performed. The criteria of the outliers were:

1. The raters using 1 or 2 scale interval throughout the whole evaluation will be entirely deleted.
2. Any score which was very different from the overall mean scores more than mean ± 3SD would be deleted.

From the first criteria, the 6 hospital staff and 6 laypeople were entirely deleted from 43 hospital staff and 48 laypeople respectively. And from the second criteria, 6 scores and 12 scores were deleted from total 370 and 420 scores evaluated by hospital staff and laypeople respectively.

**Consistency and reliability:** After removing duplicated photos, there were 99 and 94 photos evaluated by hospital staff and laypeople respectively. The internal consistency and inter-rater reliability were calculated from the evaluation of these photos. To assess the internal consistency of the composed scores within each panel, Cronbach’s Alpha coefficient was separately calculated from 99 and 94 photos evaluated by hospital staff and laypeople. To assess the inter-rater reliability, Intra-class Correlation Coefficient (ICC) was calculated from 99 and 94 photos evaluated by hospital staff and laypeople. To assess the intra-rater reliability, paired t test was used to compare the mean attractive scores of first and second time rating of 1 and 6 duplicated photos evaluated by hospital staff and laypeople respectively. Pearson’s correlation coefficient was used to test the correlation between first and second time rating of those duplicated photos.

**Agreement of facial attractive perception:** The overall facial attractive modes, means and standard deviations for each set of photographs were calculated. The photos were ranged from the most unattractive face to the most attractive face according to their overall mean attractive scores.

To illustrate the distribution pattern of 99 and 94 rating scores evaluated by hospital staff and laypeople respectively, mean percent of raters were calculated from number of raters rating most common used scale for both one and two scale range. To determine the different between perception of attractive and unattractive faces, the 30 lowest mean attractive scored and 30 highest mean attractive scored photos were used as samples representing unattractive and attractive faces respectively.

To evaluate whether raters agree more in judging attractive faces attractive or judging unattractive faces unattractive, an independent T test was used to compare the mean percent of raters rating one scale range and two scale range.

**Influence of gender, age and professional background on facial attractive evaluation:** After ranging the photos from the most unattractive face to the most attractive face according to their overall mean attractive scores, scatter diagrams of mean facial attractive scores given by subdivided groups of raters according to gender, age and professional background were created. In order to demonstrate the tendency of each evaluation, the polynomial or curvilinear trend line were made from Microsoft Excel 2010, using the following equation to calculate the least squares fit through points: $y = b + c_1 x + c_2 x^2 + c_3 x^3$ where $b$ and $c_1$, $c_2$, $c_3$ are constants.

To assess the influence of gender and age on facial attractive perception, means and standard deviations of the facial attractive evaluation for set of 99 and 94 photos evaluated by hospital staff and laypeople respectively were calculated according to gender and age of raters. An independent t test was used to compare the facial attractive scores between male and female, old and young raters. In addition because the median age of every rater was equal to 30 years old; therefore, we used 30 years old as a reference point. Raters with 30 years and younger were assigned in the young group and raters who are older than 30 years were assigned in the old group.

To assess the influence of professional background on facial attractive perception, means and SDs of the facial attractive evaluation for set of 54 photos which has been evaluated by both hospital staff and laypeople were calculated. An independent t test was used to compare the facial attractive scores between professional backgrounds.

The hospital staff and laypeople were also subdivided regarding their gender and age. One-way Analysis of Variance (ANOVA) was employed to compare the facial attractive perception between four groups of evaluators regarding their gender and age. Post hoc testing was done with the Tukey HSD method for multiple comparisons.

All statistical analysis was performed with software (Statistical package for Social Sciences, Version 19.0, SPSS Inc., Chicago, Illinois, USA) and the statistical significant was set at $p \leq 0.05$ for all analyses.

**Results**

**Raters and mean facial attractive scores**

After removed the outliers there were total 37 hospital staff and 42 laypeople included in this study. The distribution of gender and mean age of raters, means and standard deviations of attractive scores given by hospital staff and laypeople were shown in Table 1. The overall mean attractive scores of 99 and 94 photos non-duplicated photos evaluated by hospital staff and laypeople equal to 2.39 ± 0.68 and 2.31 ± 0.61 respectively.

**Consistency and reliability**

From evaluation of 99 and 94 photos by hospital staff and laypeople respectively, Cronbach’s Alpha showed excellent internal consistency of facial attractive perception in both
hospital staff (ρ=0.999) and laypeople (ρ=0.989) The ICC showed the inter-rater reliability of hospital staff and laypeople equal to 0.953 and 0.686 respectively.

Mean and standard deviation of attractive scores and comparison between first and second evaluations of duplicated photos were shown in Table 2. Mean attractive score of all 6 pair duplicated photos given by laypeople were higher for the second duplicated photos than the first duplicated photos. But paired t test showed no significant different of mean facial attractive rating scores for each duplicated photos.

Pearson’s correlation coefficients for the association between each duplicated photos.

No significant different of mean facial attractive rating scores for first and second evaluation of duplicated photos evaluated by photos than the first duplicated photos. But paired t test showed those duplicated photos evaluated by laypeople hospital staff equal to 0.352 and range from 0.356 to 0.765 for photos given by laypeople were higher for the second duplicated photos than the first (Table 2). Every tested correlation were significant (ρ<0.05).

**Agreement of facial attractive perception**

Although, 5-point Likert scale was used, every photo was rated with central tendency and unimodal distribution (Figure 1). This distribution was consistently formed for all score distribution no matter of attractiveness of photos or rater’s background. Considering only the most common used of one-scale interval rating each photo, there were mean percent of 54.3 ± 8.4% hospital staff and 52.7 ± 8.3% laypeople. Considering the most common used of 2-scale interval rating each photo, there were mean percent of 86.3 ± 6.4% hospital staff and 84.9 ± 6.8% of laypeople. There was no significant different between the percentage of hospital staff and laypeople rated with the most common scale either with one or two scale interval (ρ=0.193 and 0.126 respectively) (Table 3) (Figure 2).

The difference in rating attractive and unattractive faces by both hospital staff and laypeople was also revealed (Figures 3 and 4). For the most common used of one scale range, 56.8 ± 8.4% and 54.8 ± 9.3% of hospital staff rated 30 most unattractive and 30 most attractive faces respectively and 54.9 ± 10.2% and 51.0 ± 6.5% of laypeople rated 30 most unattractive and 30 most attractive faces respectively. The different between mean percent of raters rating unattractive and attractive with one scale range was no significant (ρ=0.387 and ρ=0.083 for hospital staff and laypeople respectively). For the most common used of two scale range, 87.8 ± 6.9% and 83.5 ± 5.2% of hospital staff rated 30 most unattractive and 30 most attractive faces respectively and 90.4 ± 6.6% and 81.6 ± 5.6% of laypeople rated 30 most unattractive and 30 most attractive faces respectively. The difference between mean percent of rating unattractive and attractive was significant (ρ=0.010 and ρ=0.000 for hospital staff and laypeople respectively). All raters had more consistent in rating unattractiveness than attractiveness.

**Comparison of facial attractive perception according to gender**

Overall mean attractive score given by male and female raters were shown in Table 1. Mean attractive scores of 99 and 94

**Table 1 Distribution of raters’ gender, age, and professional background and mean attractive scores of each evaluation.**

<table>
<thead>
<tr>
<th></th>
<th>Hospital staff (n=37)</th>
<th>Laypeople (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Age</td>
<td>Mean attractive score</td>
</tr>
<tr>
<td>Male</td>
<td>15</td>
<td>36.93 ± 12.52</td>
</tr>
<tr>
<td>Female</td>
<td>22</td>
<td>32.50 ± 8.76</td>
</tr>
<tr>
<td>Old</td>
<td>17</td>
<td>42.71 ± 9.43</td>
</tr>
<tr>
<td>Young</td>
<td>20</td>
<td>26.63 ± 2.06</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>34.22 ± 10.45</td>
</tr>
</tbody>
</table>

**Table 2 Mean and standard deviation of mean attractive scores of duplicated photos, statistics of differences (ρ value) and correlation (r) between attractive scores of hospital staffs and laypeople.**

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>Photo</th>
<th>Mean ± SD</th>
<th>p Value (paired t)</th>
<th>Pearson’s correlation coefficient (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital staffs</td>
<td>#88</td>
<td>2.24 ± 0.55</td>
<td>0.800</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>#98</td>
<td>2.22 ± 0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#4</td>
<td>2.17 ± 0.76</td>
<td>0.472</td>
<td>0.356</td>
</tr>
<tr>
<td></td>
<td>#42</td>
<td>2.26 ± 0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#10</td>
<td>3.29 ± 0.67</td>
<td>0.439</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td>#59</td>
<td>3.38 ± 0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#22</td>
<td>1.79 ± 0.81</td>
<td>0.073</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>#88</td>
<td>1.98 ± 0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#23</td>
<td>2.60 ± 0.80</td>
<td>0.498</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>#65</td>
<td>2.67 ± 0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#27</td>
<td>2.02 ± 0.72</td>
<td>0.133</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>#70</td>
<td>2.14 ± 0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#30</td>
<td>1.20 ± 0.41</td>
<td>0.058</td>
<td>0.587</td>
</tr>
<tr>
<td></td>
<td>#79</td>
<td>1.34 ± 0.48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 Mean percent of raters rating with most common scale.

<table>
<thead>
<tr>
<th>Raters</th>
<th>Scale</th>
<th>Mean percent of raters (range)</th>
<th>Overall photos</th>
<th>30 Most unattractive</th>
<th>30 Most attractive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital staff</td>
<td>1 scale</td>
<td>54.3 ± 8.4%</td>
<td>56.8 ± 8.4%</td>
<td>54.8 ± 9.3%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 scale</td>
<td>86.3 ± 6.4%</td>
<td>87.8 ± 6.9%</td>
<td>83.5 ± 5.2%</td>
<td></td>
</tr>
<tr>
<td>Laypeople</td>
<td>1 scale</td>
<td>52.7 ± 8.3%</td>
<td>54.9 ± 10.2%</td>
<td>51.0 ± 6.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 scale</td>
<td>84.9 ± 6.8%</td>
<td>90.4 ± 6.6%</td>
<td>81.6 ± 5.6%</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 Frequency of 5-point Likert scale rating of most unattractive (left), average (middle), most attractive photographs (right) evaluated by hospital staff (upper row) and laypeople (Lower row). The unimodal distribution of the evaluating scores was shown in all evaluations.

Figure 2 Scattergram of mean percent of hospital staff and laypeople using most common scale in both 1 and 2 scale range rated facial attractiveness of 99 and 94 photos respectively.

Comparison of facial attractive perception according to age

Overall mean attractive score given by old and young raters were shown in Table 1. Mean attractive scores of 99 and 94 photos given by hospital staff and laypeople according to age were shown in Figures 7 and 8. There were no significant difference in the facial attractive evaluation between old and young raters in
both hospital staff and laypeople (p=0.457 for hospital staff and p=0.781 laypeople).

Comparison of facial attractive perception according to professional background

Figure 9 showed mean facial attractive scores of 54 photographs given by both hospital staff and laypeople. The overall mean attractive score of 54 photographs evaluated by hospital staff and laypeople equal to 2.29 ± 0.34 and 2.04 ± 0.41 respectively. The t test revealed hospital staff gave significant higher scores than laypeople (p=0.005).

Figure 10 showed mean facial attractive scores for gender difference between hospital staff and laypeople (ANOVA, all panels F=5.302, p=0.002). The same trend of facial attractive evaluation was found. In each group female raters gave better score than male raters and hospital staff gave better score than laypeople. Tukey HSD revealed male laypeople gave significantly lower scores than other evaluators. (p=0.010 and p=0.002 for comparison between male laypeople versus male and female hospital staff respectively).

Figure 11 showed mean facial attractive scores for age difference between hospital staff and laypeople (ANOVA, all panels F=3.878, p=0.010). The same trend of facial attractive evaluation was found. In each group young raters gave better score than old raters and hospital staff gave better score than laypeople. Tukey HSD revealed young hospital staff gave significantly higher scores than old laypeople (p=0.012).
Figure 5: Scatter diagram with polynomial trend lines shows high agreement of facial attractive evaluation of 99 photos by female and male hospital staff. Although, female hospital staff tended to give a better score than males, no significant different of mean facial attractive scores rated by female and male hospital staff (p=0.710, t-test).

Figure 6: Scatter diagram with polynomial trend lines shows high agreement of 94 facial attractive evaluation by female and male laypeople with female laypeople rated significant higher scores than male laypeople (p=0.011, t-test).

Figure 7: Scatter diagram with polynomial trend lines shows high agreement of facial attractive evaluation of 99 photos by old and young hospital staff with no significant different of mean facial attractive scores rated by old and young hospital staff (p=0.457, t-test).
**Figure 8** Scatter diagram with polynomial trend lines shows high agreement of facial attractive evaluation of 94 photos by old and young laypeople with no significant different of mean facial attractive scores rated by old and young laypeople (p=0.781, t-test).

**Figure 9** Scatter diagram with polynomial trend lines shows high agreement of 54 facial attractive evaluation by both hospital staff and laypeople with hospital staff rated significant higher scores than laypeople (p=0.005, t-test).

**Figure 10** Scatter gram with trend lines shows mean facial attractive scores of 54 photos evaluated by male hospital staff, female hospital staff, male laypeople and female laypeople (ANOVA, all panels F=5.302, p=0.002).
Discussion

Raters

In this study, the hospital staff was recruited from people working in the craniofacial center. They represented wide age range and various professional experiences and they were assumed to be representative of people who involved in daily aesthetic assessment and treatment. The laypeople were non-medical university students or teaching staff. None of them was trained in medical, dentistry or the facial art. They were representative of people who appreciated daily for facial esthetics among their own people.

Consistency and reliability

Previous studies allowed time limitation as 10 seconds or 15 seconds or without time limitation [52-64]. However, experimental studies showed that attractiveness could be rapidly and accurately extracted within 1000 ms [50] or 100 ms [51] of viewing time. Therefore, we allowed 5 seconds for evaluators to view the photos and 3 seconds to mark the attractive score on 5-point Likert scale. We found not only the raters made their decision less than 5 seconds, Cronbach’s alpha revealed very high internal consistency of female facial attractive perception within groups of hospital staff and laypeople. Moreover, they did not feel any constrain during the evaluation. This study could be the first published evidence to support a person could make the facial attractive judgment within 5 seconds.

The correlation of first and second time attractive evaluation for one week interval was 0.75-0.92 [57] and for two-week interval was 0.23-0.91. However, there was significant difference of ranking score of the duplicated female profile images (p<0.01) [65]. In another study, they showed the Pearson correlation coefficient of the immediate evaluation between first and second time ranged from 0.40-0.87. These showed that although there were difference opinions within individuals, the consensus of the overall evaluators still high enough [55]. In this study, although paired t test showed no significant different among the facial attractive evaluation of all raters, the mean score of second viewing always higher than the first viewing for all 6 duplicated photos evaluated by laypeople. This phenomenon could explain the more you see someone, the more you like them. Moreover, positive correlation of first and second time evaluation for all of duplicated photos were significant different. Although one duplicated photo evaluated by hospital staff ($\bar{x}_{#1}=2.24 \pm 0.55$, $\bar{x}_{#5}=2.22 \pm 0.59$; $r=0.0352$) and 2 out of 6 duplicated photos evaluated by laypeople ($\bar{x}_{#1}=2.17 \pm 0.76$, $\bar{x}_{#2}=2.26 \pm 0.73$; $r=0.356$ and $\bar{x}_{#5}=3.29 \pm 0.67$, $\bar{x}_{#6}=3.38 \pm 0.76$; $r=0.40$) were not high, the correlations of other 4 duplicated photos were moderate to good ($r=0.587-0.765$).

Agreement of facial attractive perception

The present study also clearly showed that any raters regardless the gender, age or professional background agreed to judge unattractive faces with low mean attractive scores and agreed to judge attractive faces with high mean attractive scores. These evaluations were distributed with central distribution. As shown in Figure 1, although the attractive scales varied from 1 to 5, no matter of unattractive, average or attractive photos the rating scores by hospital staff or laypeople would be concentrated within one to two scales with only single mode distribution. This was well-supported by Figures 4 and 5 that each photo was well-uniformed rated within 2 scale by majority of evaluators (86.3% of hospital staff and 84.9% of laypeople). Similar to other study, which showed a well-formed mode centered around 5 for the mean attractive score of 5.15 ± 1.76 in Likert scale from 1 to 9. This indicated a strong central tendency exists even the raters are from different population [66].

Moreover, the consistency of agreement in judging unattractive faces unattractive was higher than the consistency of agreement in judging attractive faces attractive because evaluators used within 2 to 3 scales to evaluate the 10 most unattractive photos while used 3 or more scales to evaluate the 10 most attractive photos. In other words, people showed more variety of opinions in judging the attractive faces attractive than judged unattractive faces unattractive.
Effect of gender on female facial attractive perception of hospital staff

For the assessing of the effect of gender on facial attractive perception among groups of hospital staff, there was no significant different between male and female raters though female raters tend to give higher score. Similar result was found that dentists and orthodontists showed perfect agreement in their profile evaluation [56]. These might indicated that trained people such as orthodontist and plastic surgeon used similar standard to evaluate the female facial attractiveness. The finding results were different from Kieken et al. who found male orthodontist rated the female adolescents more attractive than the female orthodontists [67]. These might due to different groups of evaluators and different sets of photographs used in their study especially the three-quarter smiling view included. The smiling view might impact on the evaluation [68]. It was also found that female orthodontists detected significant differences of smile arc and buccal corridors width while male orthodontist did not [69].

Effect of gender on female facial attractive perception of laypeople

In laypeople, however, male laypeople rated female faces with significant lower scores than female laypeople did. While the female attractive perception among female and male laypeople most of the male laypeople were still single at the time of rating. This suggested that male laypeople were more critical in female attractive evaluation than female laypeople. Similar result was found in other study that male raters rated attractiveness of the female patients who presenting at dermatology lower scores than female raters [70]. In contrast, other found male laypeople tend to give higher scores than female laypeople. These findings supported an idea that different standard might existed for facial attractiveness between male and female evaluators as male and female raters might perceive female facial attractiveness differently [71]. For the female profile preferences, it was found that male evaluators more preferred to the convex profiles while female evaluators more preferred to the concave profiles [56] or while both male and female evaluators are in agree in preference of lip position that is more protrusive than Rickett’s standard, female prefer a fuller lip position than males in both female and male stimulus faces [72]. However, some other studies found there was no significant different between the female attractiveness evaluation by male and female laypeople [38,62,65,67].

Effect of age on female facial attractive perception of hospital staff and laypeople

For an effect of age, we found both young hospital staff and laypeople evaluated the female facial attractiveness similarly with old hospital staff and laypeople respectively. Kieken RMA et al set the effect for age by using dichotomized at 46 years of age and also found the same result that “age effect” was not found in the female attractive perception of orthodontists and laypeople [67].

Effect of professional background between hospital staff and laypeople on female facial attractive perception

Although many studies agreed that dental professions especially orthodontists are more critical and more sensitive in judging the esthetic of teeth and smile [73-76], the controversy whether the professional are more or less critical than the laypeople when judging the beauty of the face existed. While some studies revealed that professionals were less critical [52,53,59,65], others found there were significant in facial esthetic evaluation between professional and laypeople [55,68] or even found that laypeople were less critical [57,67,77]. This study found while hospital staff and laypeople generally agreed in their perceptions of facial attractiveness, significant different was observed as laypeople are more critical than the hospital staff. It supported an idea that different standard existed for facial attractiveness between professional and laypeople such as orthodontists preferred more forward profile than laypeople do [78] or different attractive rating score will be given to different facial profile between laypeople and orthodontists [56,60,65].

Previous studies proposed many factors yielding differences of facial attractive perception between hospital staff and laypeople. For example socioeconomic and societal factors, lower socioeconomic raters tended to score less favorably than those in higher groups [79]. Another study suggested that the level of dental education or training experience has a significant effect on facial attractive evaluation because they found that orthodontic residents consistently rating patients as more attractive than dental students and laypeople [53]. In addition, it was found that the laypeople took lesser time to complete the attractive evaluation than orthodontists and oral surgeons; therefore, it might be concluded that laypeople tended to rate the profiles on their initial evaluations while orthodontists and oral surgeons might over evaluated each profiles [64]. Also, laypeople might make their evaluations based on the entire face while orthodontists and oral surgeons tend to direct their attention to certain portion of facial profile such as the dentoalveolar region and concentrate on a specific area of the profile. Other study also supported that orthodontists were more focused on and influenced by the profile than on the whole face [59]. This study, however, showed that differences between gender and age of raters were factors caused the different of attractive perception between professional and laypeople. For gender, male laypeople were the most critical group when judging female facial attractiveness. Therefore hospital staff should be aware of their less critical facial attractive perception than the patients especially compare to the male patients.

For age, although there was no significant different within hospital staff and laypeople, but there was significant different between young hospital staff and old laypeople. Young hospital staff was the less critical group and old laypeople were the most critical group when judging female facial attractiveness. Therefore, young hospital staff especially the one who involving craniofacial patients should be aware of the fact that they might have a different perception of facial attractiveness especially with the old laypeople.
Limitations
The hospital staff in our study was staff working in craniofacial center who involved with craniofacial patients. They do not exactly represent the general hospital staff. And laypeople were limited only to university students and staff considering as well-educated young adults. Further studies should expand more to other populations. Based on 2 dimensional photographs, we have proved that the consistency of the female facial attractive evaluation was very high; however, factors of gender and professional background might affect the opinions. With the advancement of 3D technology improvement we should find new ways to be able to assess the attractiveness of the face more accurate. The finding of this study are important to remind that the decision to perform treatment enhance facial beauty should not be based only on professional preference, but also on the patient’s perception of facial esthetics. Moreover, the facial evaluation in this study only focused on female facial attractiveness, further studies should proceed with more focusing on the male facial attractiveness.

Conclusions
1. High internal consistency of female facial attractive perception is achieved by evaluators, no matter of gender, age, or professional background.
2. Inter-rater reliability is high for hospital staff, but moderate for laypeople.
3. Laypeople tend to give higher score for the second time viewing of all 6 duplicated photos.
4. Every evaluation show central tendency and unimodal distribution regardless of the attractiveness or rater’s background.
5. More consistency is found for the evaluation of unattractive faces than attractive faces by both hospital staff and laypeople.
6. In hospital staff, factors of gender and age would not influence the female facial attractiveness evaluation.
7. In laypeople, male evaluators were more critical than female evaluators in the evaluation of female facial attractiveness.
8. In laypeople, the factors of age would not influence the evaluation of female facial attractiveness.
9. Laypeople were more critical than hospital staff in the evaluation of female facial attractiveness.
10. While other raters gave similar trend for female facial attractiveness, male laypeople was the most critical.

Funding
This project was supported by Chang Gung Memorial Hospital (CRRPG5C0263, CRRPG5C0223) and Ministry of Science and Technology, Taiwan (103-2314-B-182-042-MY2).

References


