European Journal of Orthodontics 29 (2007) 95–99
doi:10.1093/ejo/cjl060

Influence of panel composition on aesthetic evaluation of adolescent faces

R.M.A. Kiekens*, M.A. van 't Hof**, H Straatman***, Anne M. Kuijpers-Jagtman* and J.C. Maltha*

Departments of *Orthodontics and Oral Biology, **Preventive and Curative Dentistry, and ***Epidemiology and Biostatistics, Radboud University Nijmegen Medical Centre, The Netherlands

SUMMARY The objective of this study was to evaluate the influence of professional background, age, gender, and geographical region of panel members on their evaluation of the facial aesthetics of adolescents, and to assess the optimal panel size for epidemiological studies on facial aesthetics.

A panel of 76 adult laymen from two different regions (Belgium and The Netherlands) and a panel of 89 orthodontists from the same two regions, evaluated photographic sets (one frontal, one three-quarter smiling, and one lateral view) of 64 adolescents (32 boys, 32 girls) on a visual analogue scale (VAS) in relation to a reference set of photographs. The effects of the characteristics of the panel members on the VAS scores for boys and girls separately, as well as their interactions, were evaluated by multilevel models. The adolescents entered the model as a random effect and four characteristics of the panel members were included in the model as fixed effects.

The multilevel model with main effects and first-order interactions revealed that laymen rated adolescents as more attractive than orthodontists. This finding was significant for all laymen, except for older males, and Belgian laymen, when rating girls. Older panel members rated boys significantly more attractive than younger panel members. Males rated adolescents more attractive than females. The latter was significant for all male subgroups, except for the lay male subgroup. There were regional differences.

Based on the intraclass correlation coefficient, a panel of seven randomly selected laymen and/or orthodontists is sufficient to obtain reliable results in the aesthetic evaluation of adolescent faces, using photographs and a VAS.

Introduction

In investigations of facial aesthetics, judgements of panels have often been compared, but conflicting results have been reported. Differences in study design may, to a large extent, be responsible for these conflicting results. In addition, factors related to the individual characteristics of the panel members such as professional background, age, gender, and geographical region may also influence the ratings. Although high correlations have been reported between professionals and laymen (Peerlings *et al.*, 1995; Spyropoulos and Halazonetis, 2001; Kiekens *et al.*, 2005), some investigations have shown that professionals are more critical than laymen (e.g. Kerr and O'Donnell, 1990; Kiekens *et al.*, 2005), while other studies found the opposite (e.g. Tedesco *et al.*, 1983a; Phillips *et al.*, 1992a,b; Giddon *et al.*, 1996; Spyropoulos and Halazonetis, 2001). Differences in panel composition concerning age and gender can be confounders in this respect (Spyropoulos and Halazonetis, 2001). The age of panel members was not found to be influential on their ratings of facial aesthetics (Cross and Cross, 1971; Howells and Shaw, 1985).

The influence of gender of panel members on their ratings of facial aesthetics is not clear. Some studies indicated that the gender of panel members was not decisive for their ratings (De Smit and Dermaut, 1984; Howells and Shaw, 1985). Other studies, however, suggest that females are less critical than males (Tedesco *et al.*, 1983b). Cross and Cross (1971) found that female laymen rated female faces as more attractive than male laymen, while female and male laymen rated male faces the same.

Limited research has been performed on the effect of the geographical region of panel members on the appreciation of facial aesthetics. Udry (1965) reported on different preferences of feminine beauty in Britain and the United States; it appears that no publications are available on regional differences of panel members, assessing facial aesthetics in adolescents. Such regional differences, however, may be of interest for orthodontists moving to another region. They can benefit from this knowledge in their discussions on treatment expectations.

Panel size is another issue that should be taken into consideration. The literature shows a wide range in panel size (e.g. Phillips *et al.*, 1992a,b; Kiekens *et al.*, 2005). Howells and Shaw (1985) stated that for evaluation of facial aesthetics, a panel of two persons can give acceptable reliability, but for improvement, they advocated a further increase in panel size. However, the optimal size of such a panel has never been established.

The aim of this study was twofold: to evaluate the influence of, and the possible interactions between, professional background, age, gender, and geographical region of panel members on their ratings of facial aesthetics in adolescents, and to find indications for optimal panel size for epidemiological investigations of facial aesthetics.

Materials and methods

The 1990–2000 files of the Department of Orthodontics and Oral Biology of the Radboud University Nijmegen Medical Centre, The Netherlands, were searched for pre-treatment sets of three photographs (one frontal, one three-quarter smiling, and one lateral) of healthy Caucasian adolescents. The inclusion criteria were aged between 10 and 16 years, not wearing glasses, and without dental or facial trauma or known congenital defects. From this group, 64 subjects were selected, using randomized stratification for Angle Class and gender. Angle classifications were defined as follows: Angle Class I: neutro-occlusion and neutro-relationship of the jaws; Class II division 1: disto-occlusion and disto-relationship of the jaws, with proclined upper incisors; Class II division 2: disto-occlusion and disto-relationship of the jaws, with retroclined upper incisors; and Class III: mesio-occlusion and mesio-relationship of the jaws. This stratification was carried out to have approximately eight boys and eight girls for each of the Classes, in order to have a wide range of dental/skeletal variation.

A panel of 78 laymen, with a relatively high socio-economic status, from Flanders (the northern, Dutch-speaking part of Belgium) and The Netherlands, and a panel of 89 orthodontists (85 orthodontists and four postgraduates) from the same geographical regions evaluated the photographic sets. Distribution of geographical region, gender, and age for both panels is shown in Table 1. Ratings of facial aesthetics were performed on a visual analogue scale (VAS) in relation to reference sets of photographs, one for the boys and one for the girls, on which the VAS scores were indicated. Each set of photographs of one individual, together with the appropriate reference set, was shown for 15 seconds. The panel members were asked to assess facial aesthetics of the individual on a VAS from 0 (very unattractive) to 100 (very attractive). This method has been shown to yield reproducible and valid results (Kiekens *et al.*, 2005).

Table 1 Distribution of geographical region, gender, and age (years) of laymen and orthodontists in the panels.

Panel members	*n*	Age		
		Mean ± SD	Median	Range
Laymen				
Dutch	42 (26 m, 16 f)	53.2 ± 9.5	52	30–74
Belgian	36 (12 m, 24 f)	48.3 ± 10.7	47	28–76
Orthodontists				
Dutch	47 (29 m, 18 f)	46.5 ± 8.5	45	31–65
Belgian	42 (9 m, 33 f)	37.3 ± 6.4	38	25–53

SD, standard deviation; m, male; f, female.

Statistics

Statistical analysis was performed on the ratings of a final panel of 76 laymen and 89 orthodontists. The ratings of two laymen were not taken into account because of missing data.

Means and standard deviations (SDs) of the ratings for each set of photographs were calculated for each panel, age, gender, and geographical region. The fixed effect for age was dichotomized at 46 years of age, which was the median age of the panel members, with 46 years and older = old and under 46 years = young.

The influence of professional (orthodontic) background, age, gender, and geographical region on the VAS scores for the boys and the girls separately and their possible interactions was tested within the framework of multilevel models. Second-order and higher order interactions are difficult to interpret; therefore, only the model with the four main effects and first-order interactions were presented. Analyses were performed with the procedure 'Proc Mixed' in the statistical package SAS 8.0. (SAS® Software, SAS Institute Inc., Cary, North Carolina, USA).

The subjects were entered as a random effect in the analysis. Variance of the random effects, Vb, is the between-subjects variance and reflects the variability of the VAS score between subjects. The within-subject variance, Vw, reflects the variability of the panel members over the same subject. The intraclass correlation coefficient (ICC) is then given by $V\text{b}/(V\text{b} + V\text{w})$, which can be interpreted as the mean correlation of randomly selected pairs of single panel members. The ICC is 1 when all panel members agree perfectly on all subjects. When the within-subject variance is large (raters substantially disagree on the same subject) compared with the between-subjects variance, the ICC is close to 0. The VAS can be considered to be a reliable measure if the ICC is above 0.80. When the VAS panel score is based on the average VAS scores of N randomly selected raters, the ICC for pairs of panels is $\text{ICC}(N) = N \times \text{ICC}(1)/[1 + (N - 1) \times \text{ICC}(1)]$.

The optimal panel size was found by choosing the smallest value of N where ICC(N) was substantial above 0.80 for girls as well as for boys.

Results

VAS means and SD of the aesthetic scores for boys, girls, and boys and girls together, for each given panel, age, gender, and geographical region, were calculated and are shown in Table 2.

Table 2 Visual analogue scale (VAS) means and standard deviation (SD) of the aesthetic scores for the photographs of boys, girls, and boys and girls taken together, given by laymen and orthodontists, young (under 46 years) and old (46 years and over) panel members, males and females, and Dutch and Belgian panel members.

Panel members	*n*	Mean VAS scores ± SD		
		Boys	Girls	Both
Laymen	76	55.3 ± 5.8	52.6 ± 6.3	53.9 ± 5.7
Orthodontists	89	49.4 ± 6.8	50.6 ± 6.1	50.0 ± 6.2
Young	88	50.1 ± 6.9	50.9 ± 6.0	50.5 ± 6.1
Old	77	54.4 ± 6.3	52.2 ± 6.6	53.3 ± 6.1
Males	75	53.5 ± 6.0	52.7 ± 6.1	53.1 ± 5.7
Females	90	50.9 ± 7.5	50.6 ± 6.4	50.8 ± 6.6
Dutch	87	52.1 ± 6.9	52.0 ± 6.0	52.1 ± 6.2
Belgian	78	52.1 ± 7.1	51.0 ± 6.5	51.6 ± 6.4

The differences in VAS score are presented in Table 3 for professional background within subgroups, young, old, males, females, Belgian, and Dutch separately for boys and for girls. Laymen rated boys significantly more attractive than orthodontists. Young laymen, female laymen, and Dutch laymen also rated the girls significantly more attractive than young, female, and Dutch orthodontists.

Differences in VAS score are shown in Table 3 for age within subgroups, laymen, orthodontists, males, females, Belgian, and Dutch separately for boys and girls. Older panel members rated boys significantly more attractive than younger panel members. There was no statistical difference between the ratings of older and younger panel members for the girls.

Table 3 shows the differences in VAS score for gender within subgroups, laymen, orthodontists, young, old, Belgian, and Dutch separately for boys and for girls. Males rated the boys and girls more attractive than females. This finding was significant for all males, except for lay males. There was no significant difference between the ratings of male and female laymen.

Differences in VAS score are presented in Table 3 for geographical region within subgroups, laymen, orthodontists, young, old, males and, females separately for boys and for girls. Belgian laymen rated the girls significantly less attractive than Dutch laymen. Belgian orthodontists rated the boys and girls significantly more attractive than Dutch orthodontists. Older and female Belgian panel members rated the boys significantly more attractive than older and female Dutch panel members.

Significant first-order interactions were seen between gender and professional background ($P < 0.01$ for boys and girls) and between region and professional background ($P < 0.01$ for boys and girls). These interactions revealed that the difference between laymen and orthodontists is different for females versus males and for Belgian versus Dutch panel members.

Table 3 Difference in visual analogue scale (VAS) means in a multilevel model with all main effects. The *P*-values are corrected for multiple testing (Tukey–Kramer).

	Boys		Girls	
	Difference in VAS	*P*-value	Difference in VAS	*P*-value
Laymen–orthodontists				
Young	5.1	<0.01	1.8	<0.01
Old	4.0	<0.01	1.4	0.10
Males	2.6	<0.01	0.2	0.97
Females	6.5	<0.01	3.0	<0.01
Belgian	2.2	<0.01	−1.3	0.21
Dutch	6.9	<0.01	4.5	<0.01
Old–young				
Laymen	2.0	<0.01	−0.1	0.99
Orthodontists	3.1	<0.01	0.3	0.94
Males	2.9	<0.01	0.2	0.99
Females	2.2	<0.01	0.1	0.99
Belgian	3.1	<0.01	−0.7	0.75
Dutch	2.0	<0.01	0.9	0.31
Males–females				
Laymen	0.2	0.97	0.4	0.89
Orthodontists	4.2	<0.01	3.2	<0.01
Young	1.9	<0.01	1.8	0.01
Old	2.6	<0.01	1.8	0.02
Belgian	2.1	<0.01	1.8	0.02
Dutch	2.4	<0.01	1.7	<0.01
Belgian–Dutch				
Laymen	−0.7	0.60	−3.4	<0.01
Orthodontists	4.0	<0.01	2.4	<0.01
Young	1.1	0.15	0.3	0.9
Old	2.2	0.01	−1.3	0.2
Male	1.5	0.09	−0.5	0.9
Female	1.9	<0.01	−0.6	0.7

As the between-subjects variance (Vb) for the orthodontists was larger than for the laymen, the orthodontists used a larger part of the VAS for judging the group of adolescents. The within-subject variance (Vw) for the orthodontists was also larger than for the laymen. As a consequence, the orthodontists disagreed more than the laymen (Table 4).

In Table 5, the ICCs for panels of varying sizes from 1 to 10 are given. A random selection of seven panel members from a total of 165 led to an ICC of 0.80 and 0.85 for boys and girls, respectively. The ICC for the mean VAS score of seven randomly selected adult laymen from a total of 76 was 0.82 for boys and girls. A random selection of six orthodontists from a total of 89 resulted in an ICC of 0.82 and 0.86 for boys and girls, respectively. Based on the ICC, a panel of seven randomly selected laymen and/or orthodontists is sufficient to obtain reliable results in the aesthetic evaluation of adolescent faces, using photographs and a VAS.

Discussion

The lay panel was composed of males and females with a relatively high socio-economic status. This was justified

Table 4 Between-subjects variance (*V*b) and within-subject variance (*V*w) for laymen, orthodontists, and a mixed panel in the evaluation of the facial aesthetics of boys and girls.

	*V*b	*V*w
Boys		
Mixed	94.10	163.89
Laymen	82.93	130.75
Orthodontists	127.44	169.74
Girls		
Mixed	126.28	156.28
Laymen	86.72	133.83
Orthodontists	172.55	163.20

since orthodontic treatment demand is higher in groups with a high socio-economic status than in those with a lower status, whereas the objective treatment need is similar in both groups (Wheeler *et al.*, 1994). Therefore, the lay panel can be assumed to be representative of that part of the general public assessing orthodontic treatment demand.

Since significant interactions were found, the results have to be considered separately for the different descriptors of the panel members.

Laymen rated the adolescents, especially boys, more attractive than orthodontists. As laymen are the end-users of orthodontic services, the opinion of laymen may have the most value in determining the appropriateness of aesthetic results (Bowman and Johnston, 2001). Orthodontists should be aware of the fact that they are probably more critical about facial aesthetics than patients and their parents. They can use this information in their clinical practice and in communication with their patients on the treatment expectations.

In the present study, older panel members rated boys as more attractive than younger panel members. This suggests an 'age effect', meaning that, as people become older, they become less critical in judging facial aesthetics of boys. A 'birth year effect' is another possibility. This means that the older panel members were already less critical of boys when they were younger, as in the past, aesthetics in boys was considered less important than nowadays.

Male judges rated the adolescents more attractive than the female judges. Tedesco *et al.* (1983b) found that female laymen gave higher aesthetic scores than male laymen, but the panel used in that study consisted of only 12 college freshmen (three black females, three black males, three white females, and three white males). Therefore, their panel is probably not representative of the general public and conclusions on the influence of gender differences should be considered with caution.

Although Belgium and The Netherlands are neighbouring countries, and the panel members in this study speak the same language, several differences were found between the ratings of the Belgian and the Dutch sub-panels. Orthodontists working abroad must be aware of the fact that colleagues and patients from other countries might have a different perception of facial aesthetics.

The fact that differences were found between orthodontists and laymen, between older and younger panel members, between males and females, and between panel members from different countries does not mean that they do not agree on ranking facial aesthetics or on who is more beautiful and who is less. It simply means that some groups are more critical than others in the evaluation of facial aesthetics. In fact, in a previous study, high correlations were found between the aesthetic scores of the same laymen and orthodontists incorporated in the present study (Kiekens *et al.*, 2005). Orthodontists used a larger part of the VAS than laymen, and their VAS scores within the same subject differed more than those of laymen. In calculating the ICC, the difference in scoring between laymen and orthodontists was insignificant— laymen do not see much difference between the subjects but they agree more than orthodontists.

A panel of seven randomly selected laymen and/or orthodontists (males and/or females) is sufficient to yield reliable results, using the VAS as the outcome measure in clinical and epidemiological studies of facial aesthetics in adolescents. The use of smaller panels will lead to less reliable results, while the use of larger panels is unnecessary, more time-consuming, and more expensive. However, panel characteristics have an influence on aesthetic evaluation. This means that for comparison of facial aesthetics in different groups of adolescents (e.g. different centres), the same panel composition should be used, and this is also true when comparing the facial aesthetics of boys and girls.

Table 5 Intraclass correlation coefficient for panels of size 1 to 10, separate for boys and girls, and within gender of the subjects separate for laymen, orthodontists, and mixed panels.

	1	2	3	4	5	6	7	8	9	10
Boys										
Mixed	0.36	0.53	0.63	0.70	0.74	0.78	0.80	0.82	0.84	0.85
Laymen	0.39	0.56	0.66	0.72	0.76	0.79	0.82	0.84	0.85	0.86
Orthodontists	0.43	0.60	0.69	0.75	0.79	0.82	0.84	0.86	0.87	0.88
Girls										
Mixed	0.45	0.62	0.71	0.76	0.80	0.83	0.85	0.87	0.88	0.89
Laymen	0.39	0.56	0.66	0.72	0.76	0.80	0.82	0.84	0.85	0.87
Orthodontists	0.51	0.68	0.76	0.81	0.84	0.86	0.88	0.89	0.90	0.91

Conclusions

The composition of a panel has a large impact on the aesthetic evaluation of adolescent faces, using photographs and a VAS. A panel of seven randomly selected laymen and/or orthodontists is sufficient to obtain reliable measurements of facial aesthetics.

Address for correspondence

Professor Anne M. Kuijpers-Jagtman
Department of Orthodontics and Oral Biology
Radboud University Nijmegen Medical Centre
309 Tandheelkunde
Postbus 9101
NL 6500 HB Nijmegen
The Netherlands
E-mail:a.kuijpers-jagtman@dent.umcn.nl

Acknowledgements

The authors would like to thank L. Wetzels, F. Rangel, O. Van Vlijmen, K. Windels, V. Akkerman, H. Leeuwenburg, and H. van der Laan for producing the computer images and slide show and for their contribution in the data collection. We would also like to thank the members of the Dutch Association of Orthodontists, the Belgian Orthodontic Society, and the members and their partners of the Rotary club Oranjewoud, Heerenveen, The Netherlands, and the Professional Women's Association Rollarienses of Roeselare Belgium for their kind co-operation as panel members.

References

Bowman S J, Johnston L E 2001 Much about facial esthetics. In: McNamara JA (ed.) Treatment timing: orthodontics in four dimensions. Monograph No. 39, Craniofacial Growth Series. Center for Human Growth and Development, University of Michigan, Ann Arbor, pp. 199–217

Cross J F, Cross J 1971 Age, sex, race, and the perception of facial beauty. Developmental Psychology 5: 433–439

De Smit A, Dermaut L 1984 Soft-tissue profile preference. American Journal of Orthodontics 86: 67–73

Giddon D B, Sconzo R, Kinchen J A Evans C A 1996 Quantitative comparison of computerized discrete and animated profile preferences. Angle Orthodontist 66: 441–448

Howells D J, Shaw W C 1985 The validity and reliability of ratings of dental and facial attractiveness for epidemiological use. American Journal of Orthodontics 88: 402–408

Kerr W J S, O'Donnell J M 1990 Panel perception of facial attractiveness. British Journal of Orthodontics 17: 299–304

Kiekens R M A, Maltha J C, van 't Hof M A, Kuijpers-Jagtman A M 2005 A measuring system for facial aesthetics in Caucasian adolescents: reproducibility and validity. European Journal of Orthodontics 27: 579–584

Peerlings R H J, Kuijpers-Jagtman A M, Hoeksma J B 1995 A photographic scale to measure facial aesthetics. European Journal of Orthodontics 17: 101–109

Phillips C, Trentini C J, Douvartzidis N 1992a The effect of treatment on facial attractiveness. Journal of Oral and Maxillofacial Surgery 50: 590–594

Phillips C, Tulloch C, Dann C 1992b Rating of facial attractiveness. Community Dentistry and Oral Epidemiology 20: 214–220

Spyropoulos M N, Halazonetis D J 2001 Significance of the soft tissue profile on facial esthetics. American Journal of Orthodontics and Dentofacial Orthopedics 119: 464–471

Tedesco L A, Albino J E, Cunat J J, Green L J, Lewis E A, Slakter M J 1983a A dental-facial attractiveness scale. Part I. Reliability and validity. American Journal of Orthodontics 83: 38–43

Tedesco L A, Albino J E, Cunat J J, Slakter M J, Waltz K J 1983b A dental-facial attractiveness scale. Part II. Consistency and perception. American Journal of Orthodontics 83: 44–46

Udry J R 1965 Structural correlates of feminine beauty preferences in Britain and the United States: a comparison. Sociology and Social Research 49: 330–342

Wheeler T T, McGorray S P, Yurkiewicz L, Keeling S D, King G J 1994 Orthodontic treatment demand and need in third and fourth grade school children. American Journal of Orthodontics and Dentofacial Orthopedics 106: 22–33