

A Web-Based Method for Rating Facial Attractiveness

Zlatko Devcic, BS; Koohyar Karimi, BS; Natalie Popenko; Brian J.F. Wong, MD, PhD

Objectives/Hypothesis: To determine if facial attractiveness scores from a novel Internet-based facial attractiveness rating method correlate with those from the traditional focus group rating method, and whether this method can be an alternative for rating and evaluating facial attractiveness.

Study Design: Basic research study.

Methods: Eighty facial portraits were posted on a commercial Internet-based facial rating website to obtain facial attractiveness scores. These scores were correlated and compared with traditional focus group scores.

Results: In 21 days an average sample size of 857 raters were recruited and the attractiveness scores reached a stable mean on the Web. There was a strong correlation (0.90) and attractiveness score quartile match between the Internet-based and traditional focus group scores, with the most attractive and unattractive faces having the highest correlation and quartile match. The inter-rater variability of the Internet-based method was low ($P = .82$).

Conclusions: The Internet-based method can be an effective alternative to the traditional live focus group method of evaluating facial attractiveness. It also has five main advantages: 1) profoundly increases rater count; 2) increases rate of data accrual and analysis; 3) results are reproducible; 4) eliminates logistical and monetary obstacles; and 5) enables the experimenter to sweep broad demographics, acquire background data from raters, and locate raters with specific expertise.

Key Words: Facial beauty, Internet rating, focus groups, plastic surgery, attractiveness, facial beauty rating.

Level of Evidence: 5.

Laryngoscope, 120:902–906, 2010

INTRODUCTION

The study of facial beauty is an important field of academic study, in which there have been over 2,000 publications on this specific subject since 1970.¹ The study of beauty is obviously relevant to the practice of facial plastic surgery, as the neoclassical canons established by Da Vinci, Durer, and others still guide clinical practice to this day.¹ Although these seminal works still have a profound influence in the contemporary medical, sociological, and artistic literature, more modern approaches in the study of facial beauty have used anthropometric analysis and computer software to digitally modify facial portraits and present them to focus groups of raters. The focus groups are experts or lay evaluators who rate the facial portraits on an attractiveness ranking scale, such as 1 to 10, with 1 being very unattractive and 10 being very attractive. These scores are then correlated with, for example, various angular or linear facial measurements or digital modifications.^{2,3} This approach has been extensively used by Farkas⁴ and other research groups, such as the Johnston team,⁵ to tease out associations between attractiveness scores and anthropometrics, averageness,⁶ symmetry,⁷ and sexual dimorphism⁸ of the human face.

Despite the extensive number of facial beauty studies, research has been unable to define facial beauty with more specific parameters. A significant limitation is the tedium and difficulty of obtaining large focus groups, as the current method of recruitment is by nature inefficient and a key target for improvement. First, it is organizationally and temporally complex to recruit a large focus group of raters, manually display facial portraits to them, and then collect and analyze their evaluations for each face. Furthermore, focus groups are often too small and biased with regard to social, economic, geographic, age, cultural, and educational factors. Although it is possible to recruit large focus groups with broad demographics, these types of studies are often

From the Division of Facial Plastic Surgery, Department of Otolaryngology–Head and Neck Surgery (Z.D., K.K., N.P., B.J.F.W.); the Beckman Laser Institute and Medical Clinic (B.J.F.W.), and the Department of Biomedical Engineering (B.J.F.W.), University of California Irvine, Irvine, California, U.S.A.

Editor's Note: This Manuscript was accepted for publication January 11, 2010.

Presented as a poster at the Annual Meeting of the American Academy of Otolaryngology–Head and Neck Surgery, Washington, DC, U.S.A., September 18, 2007.

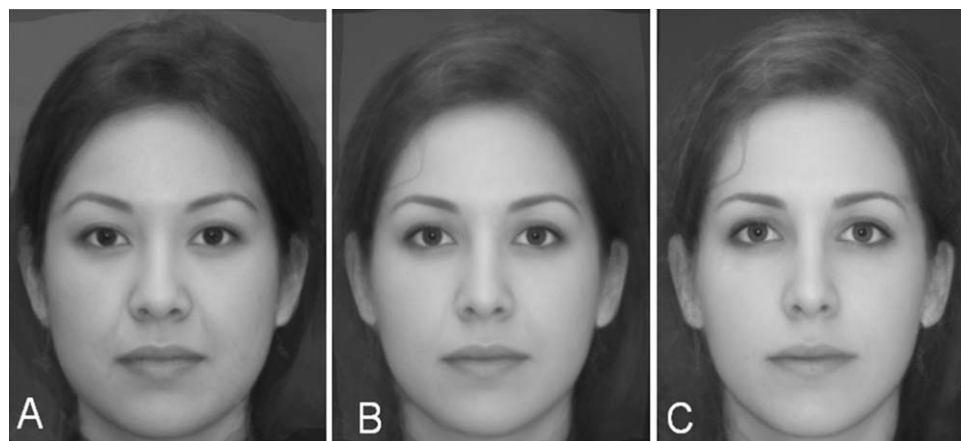
Zlatko Devcic was supported by the Undergraduate Research Opportunities Program at the University of California, Irvine. The authors have no other funding, financial relationships, or conflicts of interest to disclose.

This work was done at the Beckman Laser Institute, University of California, Irvine, California, U.S.A.

Send correspondence to Brian J.F. Wong, Beckman Laser Institute, University of California Irvine, 1002 Health Sciences Rd., Irvine, CA 92612. E-mail: bjwong@uci.edu

DOI: 10.1002/lary.20857

Fig. 1. Two facial images (A and C) are morphed together to create a synthetic face (B) that is a 50:50 average of the original two.



logistically complex and expensive. Due to these inefficiencies, facial beauty studies may accumulate data slowly, and the scope and accuracy of their results may be limited. To more rigorously study facial beauty, it would be valuable to create an innovative method that is less labor intensive, more time efficient, provides massive numbers of raters, and sweeps across vast samples and population characteristics. Based on these requirements, we investigated an Internet-based method for facial attractiveness rating.

Although the Internet is a largely untapped resource by facial beauty studies, preliminary investigation of an Internet-based method revealed that there are numerous commercial websites dedicated to rating facial portraits (i.e., digital photographs). These commercial websites provide a virtual environment for volunteers to post their facial portrait and have it evaluated by others on a standardized rating scale. As other people access the website and rate the posted face, its attractiveness score is continuously updated. Some of these websites, such as Hotornot.com (Eight Days, Inc., San Francisco, CA), are immensely popular and have received billions of votes. The benefits of using these commercially available websites are that they routinely receive heavy traffic, are designed for user friendly accessibility, are updated continuously, and provide automated data collection. This Internet-based method is essentially a virtual focus group that can efficiently rate facial attractiveness without the limitations of traditional focus group encounters.

In this study, we used Internet-based facial attractiveness rating to evaluate synthetic facial portraits created through a previous facial beauty study by our team.⁹ The objective of this study was to determine if facial attractiveness scores obtained from Internet-based rating correlate with those accrued from the traditional focus group method, and whether this method can be used as an alternative for evaluating facial attractiveness.

MATERIALS AND METHODS

Our Broader Body of Work

All facial portraits in this study were used with the approval of the institutional review board at the University of

California, Irvine. This study is part of a broader pilot investigation developing novel methods to more quantitatively and rigorously define facial beauty. Through this pilot investigation our group pioneered a method of combining a genetic algorithm with facial morphing software to evolve facial attractiveness.⁹ First, the morphing software (Morphman 2000; STOIK Imaging, Moscow, Russia) takes two facial portraits and breeds them together to create a realistic and natural-looking synthetic face that is an average of the original two (Fig. 1). Second, the genetic algorithm, which biases more attractive faces, selectively pairs faces from an initial pool of portraits, and evolves facial attractiveness through several generations of breeding facial portraits.

Focus Group Rating: The Traditional Method

One hundred twenty synthetic images were created in the pilot investigation and evaluated for facial attractiveness using the traditional focus group method.⁹ All of the synthetic portraits were created from standardized portraits of females representing various ethnicities between the ages of 18 and 25 years. Traditional focus group rating was conducted as described in our broader body of work.⁹ In short, each individual in each focus group independently rated all 120 faces on an attractiveness scale of 1 to 10. The first focus group consisted of beauticians ($n = 44$) trained in media concepts of facial beauty, the second of otolaryngology attendings and residents ($n = 12$), selected as a complement for their clinical knowledge of facial analysis, and the third consisted of undergraduate controls ($n = 44$) without specific education in the field. The attractiveness scores were averaged within each focus group and also combined so that each facial portrait had one average attractiveness score from each group and one combined score from all three groups.

Internet-Based Rating: The New Approach

This study utilized the facial attractiveness rating website Hotornot.com to post the synthetic facial portraits on the Web. According to institutional review board regulations at our institution concerning privacy, all of the facial portraits posted on the Web needed to be synthetic, as using real faces would require an extensive informed consent process profoundly limiting our long-term effort of building a database of over 3,000 faces. Hotornot.com was selected because it receives heavy traffic, automatically updates the attractiveness scores, has a user-friendly interface, and is free. When visitors access the website, it presents them with a randomly selected facial portrait so visitors cannot repeatedly rate their own portrait, and a 1 to 10

TABLE I.
Internet Attractiveness Score Data as a Function of Rater Count
and Time Since Posting for All 80 Faces.

Average Number of Raters	Time Since Posting (d)	Average Change in Internet Attractiveness Score (Attractiveness Points)
132	2	± 0.81
411	7	± 0.42
493	12	± 0.37
673	17	± 0.263
857	21	± 0.163

attractiveness scale to rate the portrait. As each facial portrait is cycled through the website and evaluated by raters, the website continuously updates the facial attractiveness score with two sets of data. The first is a reported attractiveness score that accounts for factors such as raters that consistently rate all faces using only 1 through 5 by normalizing their score to the 1 through 10 scale. The second set of data is a histogram distribution of the real attractiveness scores assigned by the raters. From the histogram distribution the real attractiveness score, which is not normalized, can be extracted. In fact, our study found there is a very strong linear correlation between the extracted histogram scores and the reported normalized scores ($R = 0.982$). Because the histogram-derived score is the true attractiveness score of the facial portraits on Hotornot.com, all references to Internet-based scores in this article refer to the histogram score.

Using Hotornot.com, 80 synthetic facial portraits created in the pilot investigation were selected at random and posted online. An individual account was created for each face, taking approximately 2 minutes per face. Preliminary investigation revealed that faces obtained more ratings if their accounts were frequently logged in. Thus, each account was accessed three to four times per day for a period of 3 weeks. At each login the facial attractiveness score, rater count, and time elapsed since posting were recorded, which took about 30 seconds. A screenshot of the histogram, which can only be accessed by the account holder, was also taken at the final login. Histogram distributions of the scores were manually analyzed in Photoshop (Adobe Systems, Inc. San Jose, CA) using the measure tool to determine the numerical average attractiveness score. Two weeks after this initial 3-week period, the same 80 faces were posted on the website again, and data were collected under the same conditions to measure the variability of this method.

Statistical Methods

Internet-based attractiveness scores were analyzed as a function of time and rater count. The 80 test faces were posted on the Web at two separate times, and the results were compared using a paired t test. Correlation analysis was then used to compare the Internet-based scores to the traditional focus group scores, and these two sets of scores were plotted against each other to assess their relationship graphically. Finally, we divided the Internet and focus group ratings into quartiles of attractiveness, and compared ratings across these quartiles.

RESULTS

The Internet-based method continuously updates attractiveness scores, and as the number of raters increases the attractiveness scores fluctuate less and reach a stable mean (Table I). Table I also reveals that

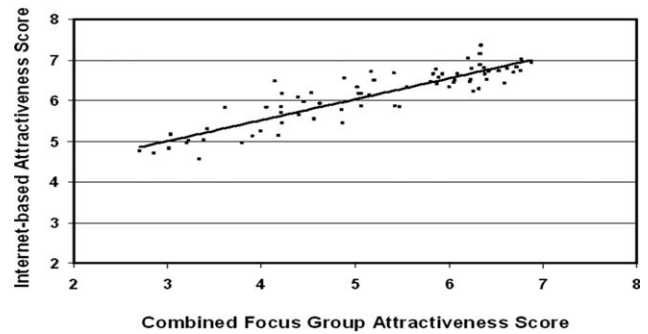


Fig. 2. Correlation between the Internet attractiveness scores and combined focus group scores for all 80 faces ($R = 0.90$).

an average of 857 raters can be recruited for each face after only 21 days.

Next, to determine if there is variability between Internet raters and to verify the accuracy of the original scores, all 80 facial portraits were posted online a second time under identical conditions. Data analysis revealed that there was no significant difference between the Internet attractiveness scores from the first and second trial of reposting all 80 facial portraits ($P = .82$).

There is not only a strong and positive linear correlation (0.90) between the Internet and combined focus group attractiveness scores (Fig. 2), but also with each individual focus group (Table II). Further, when the attractiveness scores of the Internet and combined focus groups of raters are matched by quartiles of attractiveness (Fig. 3), it is evident that both groups strongly agree on which attractiveness quartiles portraits should be assigned.

DISCUSSION

This study investigated the use of an Internet-based method as a viable alternative to the traditional focus group approach for rating facial attractiveness. One of the advantages to the Internet approach is that the investigator can dynamically monitor attractiveness scores as evaluations are accrued. As more raters score the facial portraits and the sample size increases, the Internet scores reach an accurate and stable mean at 857 raters in 21 days (Table I). Of further importance is the fact that the Internet attractiveness scores were not significantly different between the initial trial and the second reposting, indicating the variability between the Internet-based raters is low and the results are consistent and reproducible. In comparison to the focus groups, the

TABLE II.
Facial Attractiveness Score Correlations Between the
Internet-Based Raters and Each of the Focus Groups.

	Groups Compared		
	Internet and Surgeon Attractiveness Scores	Internet and Beautician Attractiveness Scores	Internet and Undergraduate Attractiveness Scores
Correlation coefficient	.882	.890	.884

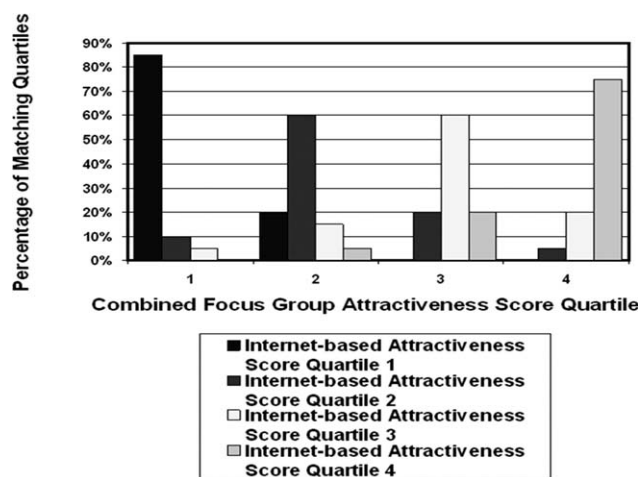


Fig. 3. Percentage of matching quartiles between Internet attractiveness scores and combined focus group scores for all 80 faces.

increase in rater count and rate of data accrual is tremendous. In our pilot investigation it took approximately 8 months and extensive logistical planning to yield only 100 raters from three focus groups. The time necessary to acquire live focus group ratings is rarely commented upon in the peer-reviewed literature, and it is generally acknowledged that this is a time consuming and laborious effort, particularly if one hopes to accrue large numbers of evaluators. With the Internet-based method we increased the rater count an average of almost nine-fold by simply creating an account for each face and logging on three times a day for 3 weeks. There were no logistical or monetary hurdles and very limited planning was required to obtain attractiveness scores. In short, the efficiency and rate of data aggregation with Internet-based rating is unparalleled to the traditional method, which is not surprising considering that such websites have received billions of votes.

Prior to discussing the strong correlation (Fig. 2) and attractiveness score quartiles match (Fig. 3) between the two rating methods, it is important to note that the Internet-based method was not expected to produce attractiveness scores that were equivalent to the traditional focus groups. Instead, the aim was to determine if the relative attractiveness scores are the same between the two methods. For example, if the focus group raters score face 1 higher than face 2, but lower than faces 3 and 4, it is critical that the Internet-based raters also score the faces in that order of attractiveness. This agreement on relative attractiveness scores is critical because investigators of facial beauty studies are interested in determining which faces are more or less attractive relative to each other to elucidate which variables increase or decrease facial beauty. Further, strictly equivalent scores were not expected because other studies have shown that the characteristics of the rating panel can impact the scores assigned by raters,¹⁰ and there is even inconsistency when scores are assigned by expert evaluator panels.¹¹ Therefore, in order for the Internet-based method to produce accurate data for fa-

cial beauty studies, it needs to generate similar relative attractiveness scores, not equivalent attractiveness scores.

The most critical point of this study is that there is a strong and positive linear correlation between the two rating methods (Fig. 2). This strong correlation indicates that both groups agree on which portraits are more or less attractive than others, and on the order in which the portraits increase in attractiveness. Similarly strong and positive linear correlations are also present when the Internet-based scores are plotted against each individual focus group (Table II). This finding suggests that the Internet method can not only serve as a proxy to traditional focus group encounters, but also to highly specialized subsets of evaluators. Further, the strong agreement in attractiveness score quartiles (Fig. 3) reinforces that both groups agree on which faces are the most attractive, least attractive, and which fall into quartiles 2 and 3. Similar to the correlation plot, this finding emphasizes that both groups of raters agree on the relative attractiveness scores assigned to the portraits. This strong agreement in relative attractiveness scores demonstrates that the Internet produces data that are as accurate and useful as those from traditional focus groups, and can be used as an effective alternative for studying facial beauty. Further, the Internet-derived scores are consistent, reproducible, and collected from a rater count nearly nine-fold greater, suggesting they are even more useful and accurate to the study of facial beauty than traditional focus group scores.

A closer inspection of the scatter plot and quartiles graph (Fig. 2 and Fig. 3, respectively) illustrates that there is a stronger correlation and quartile match in the higher and lower scoring faces than when compared to average scoring faces. Intuitively, there is a more universal concept of what defines very attractive and unattractive faces. Indeed, it has been proposed that there are intrinsic psychobiological factors related to reproductive strategy and health that drive us to rate certain faces as very attractive and others as unattractive.¹²⁻¹⁴ On the other hand, rating average faces is more subjective, because average facial cues may not elicit as strong of an intrinsic biological response as attractive and unattractive cues. Thus, the higher variability in rating average faces may not be due to the Internet-based method, but a consequence of rating facial attractiveness in general.

Although Internet-based facial attractiveness rating is a viable and efficient alternative to the traditional approach, it has some limitations that merit discussion. First, the Internet can potentially introduce new sampling biases, such as the gender, class, age, rating environment, and behavioral predispositions of those who have access to the Internet and those who visit this website specifically. However, the data suggests these potential biases are reduced by accruing a large sample size, and many of these biases are also present in the traditional method because focus groups are small and raters have behavioral predispositions in a controlled environment as well. Second, the website used in this study does not acquire rater demographics, and specific

facial beauty studies may need focus groups with known demographics. It is important to keep in mind, however, that focus groups with specific demographics are temporally and logistically tedious to recruit, and in the study of facial beauty it is often more critical to have a large sample size, which the Internet method provides. Third, at lower rater counts the Internet attractiveness scores will be less stabilized than at higher rater counts, especially for the average faces. As noted in the discussion, this may be due to higher score fluctuations at lower rater counts and greater subjectivity in rating average facial attractiveness. Last, the use of popular commercial websites subjects the investigator to the website's policies and regulations. Thus, future work will be aimed at creating an open website, specifically for research purposes, in which investigators can test for and control a wide range of variables, while simultaneously collecting anonymous background data on raters. The utility of such a website and resourcefulness of the Internet can enable the experimenter to potentially sweep very broad demographics, locating raters and virtual focus groups with specific expertise.

CONCLUSION

With the growing interest in the study of facial beauty, an alternative to the inefficient traditional focus group rating method is critical. Thus, we evaluated the effectiveness of Internet-based facial attractiveness rating. Our data indicates that the Internet-based method can produce accurate attractiveness scores that correlate very well with the traditional method. Equally as important, the Internet-based method has proven to be preferable because it has five main advantages: 1) the potential to massively increase the number of raters; 2) the results are consistent and reproducible; 3) it can be used quickly and efficiently while eliminating many logistical and monetary obstacles; 4) it can accelerate data accrual and analysis; and 5) the versatility of the Internet enables the experimenter to potentially sweep broad demographics, acquire background data from the raters, and locate raters with specific expertise. In short, prospective Internet-based methods can facilitate the execution of facial beauty research studies by efficiently recruiting a massive sample size of facial attractiveness scores that can be used as a proxy for traditional focus group scores. When specific demographics on the raters need to be collected, it will be valuable to have a website specifically for research purposes. Although the current

principles in facial beauty, as developed by Da Vinci and Durer, have withstood the test of time, more rigorous parameters defining facial beauty are necessary. Internet-based facial attractiveness rating can enhance the current study of beauty to deliver novel data to clinicians and improve the contemporary practice of facial plastic surgery.

Acknowledgment

The authors wish to thank Biostatistician Nancy K. Hills, PhD, for her assistance in the preparation of this manuscript.

BIBLIOGRAPHY

1. Bashour M. History and current concepts in the analysis of facial attractiveness. *Plast Reconstr Surg* 2006;118:741–756.
2. Honn M, Goz G. The ideal of facial beauty: a review. *J Orol-fac Orthop* 2007;68:6–16.
3. Edler R, Agarwal P, Wertheim D, Greenhill D. The use of anthropometric proportion indices in the measurement of facial attractiveness. *Eur J Orthod* 2006;28:274–281.
4. Farkas LG. *Anthropometry of the Head and Face*. New York, NY: Raven Press; 1994.
5. Johnston VS, Solomon CJ, Gibson SJ, Pallares-Bejarano A. Human facial beauty: current theories and methodologies. *Arch Facial Plast Surg* 2003;5:371–377.
6. Langlois JH, Roggman LA. Attractive faces are only average. *Psychol Sci* 1990;1:115–121.
7. Baudouin JY, Tiberghien G. Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychol* 2004;117:313–332.
8. Rhodes G. The evolutionary psychology of facial beauty. *Annu Rev Psychol* 2006;57:199–226.
9. Wong BJF, Koohyar K, Devcic Z, Christine ME, Wen-Pin C. Evolving attractive faces using morphing technology and a genetic algorithm: a new approach to determining ideal facial aesthetics. *Laryngoscope* 2008;118:962–974.
10. Spyropoulos MN, Halazonetis DJ. Significance of the soft tissue profile on facial esthetics. *Am J Orthod Dentofacial Orthop* 2001;119:464–471.
11. Kiekens RMA, van't Hof MA, Kuijpers-Jagtman AM, et al. Influence of panel composition on aesthetic evaluation of adolescent faces. *Eur J Orthod* 2007;29:95–99.
12. Smith MJ, Perrett DI, Jones BC, et al. Facial appearance is a cue to oestrogen levels in women. *Proc Biol Sci* 2006;273:135–140.
13. Cornwell RE, Law Smith MJ, Boothroyd LG, et al. Reproductive strategy, sexual development and attraction to facial characteristics. *Philos Trans R Soc Lond B Biol Sci* 2006;361:2143–2154.
14. Etcoff NL. *Survival of the Prettiest: The Science of Beauty*. New York, NY: Doubleday; 1999:325.