DOI: 10.1111/j.1365-263X.2010.01053.x

INTERNATIONAL JOURNAL OF PAEDIATRIC DENTISTRY

Development of an illustrated index of tooth appearance–perception-based quantification of tooth discolouration and surface defects

SOMA MODI[1], LUCY WILLIAMS[2], ROSEMARY GREENWOOD[3], NEIL DAVEY[4] & PETER CRAWFORD[5]

[1]Private Specialist Paediatric Dental Surgeon, Mumbai, India; [2]Specialist Registrar in Paediatric Dentistry, Bristol Dental Hospital; [3]Statistician, University Hospitals Bristol NHS Foundation Trust; [4]Helpdesk Advisor, Information Systems and Computing, University of Bristol, and [5]Consultant and Senior Lecturer in Paediatric Dentistry, University of Bristol Dental School

International Journal of Paediatric Dentistry 2010; 20: 293–304

Background. Existing indices to quantify tooth discolouration are mostly aetiology-specific. An index of tooth appearance (IOTA), derived from all types of tooth discolouration and surface defects, would allow the quantification of attractiveness for psychological assessment and treatment planning

Objective. To develop a perception based IOTA for quantification of all forms of tooth discolouration and surface defects.

Methods. One hundred images of discoloured teeth were twice ranked by a panel of judges according to perceived attractiveness. Mean image score was then used to arrange the images into a continuum of attractiveness and from these, ten images were selected, to constitute the illustrated IOTA. A second panel of judges assessed 35 clinical pictures using the IOTA, on two occasions.

Results. The first 100 images were assessed with a correlation of 0.79–0.81 between the two ranking sessions and with intra-group reproducibility of 0.8–0.94. The second panel of judges used the developed IOTA quickly, with an intra-judge correlation of 0.87 and inter-judge reliability of 0.72 and 0.74 for two sessions.

Conclusions. The IOTA could be used by clinicians as a tool for quantifying disfigurement in teeth, irrespective of aetiology or histology.

Introduction

Despite the phrase 'Beauty is only skin deep', it has long been recognized that the physical appearance of an individual has a profound psychosocial impact, and that appearance concerns affect a large percentage of the population[1]. Additionally, the mouth and teeth are crucial in verbal and nonverbal communication, and play a significant part in social interaction. The colour of teeth is perceived as critical in satisfaction with smile appearance[2], and noticeable discolouration of teeth can have a detrimental impact on a person's physical attractiveness, self image, self-confidence, and employability[3]. Participants in studies appear to appraise those with discoloured teeth as having poorer social competence, intellectual ability, psychological adjustment, and relationship status[4].

As dentists, we are becoming more aware of the prevalence of patients who perceive their teeth as less attractive. A UK study sought to determine the prevalence of perceived tooth discolouration and satisfaction with self tooth colour[5]. Half the study participants perceived themselves to have some form of tooth discolouration; satisfaction with tooth colour decreased with increasing perception of severity of discolouration. Modern dentistry has recognised aesthetics, and the resultant emotional and psychological effect of improving cosmetic appearance, as an integral part of dental health care. Thus in keeping with the WHO's broadened definition of health, several oral health quality of life indices have been developed to assess the psychosocial well being of an individual[6].

Correspondence to:
Peter Crawford, Bristol Dental School, Lower Maudlin Street, Bristol, BS1 2LY, Tel: 0117 3424 355, Fax: 0117 3428 998
E-mail: Peter.Crawford@Bristol.ac.uk

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

An altered dental appearance can have a significant impact on children as well. Dental features have been shown to be the fourth commonest cause for teasing among school children[7]. Additionally, children who report they have attractive teeth are more likely to believe themselves to have higher grades in school, be slimmer, have more friends, more money and better health than children reporting they have unattractive teeth[8], demonstrating an influence on self-esteem.

A recent study showed that parents rated their children's oral health more highly if their teeth were slightly whiter than 'normal' and if they had lower decay rates[9]. In fact, this preferred, increased whiteness could be attributed to fluorotic lesions – with a TF[10] (Thylstrup and Ferjeskov) score of 1, indicating mild fluorosis, *versus* a TF score of 0 indicating no abnormality. Thus an 'artificial' view of tooth colour was preferred.

Dentofacial attractiveness is a complex phenomenon, likely to be influenced by factors including age (of the observer and the observed), sex, dental training, socio-economic status, educational level, ethnicity, and cultural differences and the type of dentofacial anomaly. Therefore any assessment tool must be valid, usable, and reproducible between and within all demographic factors. In the field of orthodontics, several indices have been developed in an attempt to address the issue of perceived aesthetic impairment and the socio-psychological need for treatment. The SCAN index[11] is a 10-point index that has been used for many years and has proved invaluable for the aesthetic assessment of malocclusion. A comparable index, developed for assessing difference in tooth appearance, would be beneficial for dentists, psychologists, and epidemiologists alike. Although the need for understanding the importance of dental aesthetics has been emphasized[12], it is surprising that the psychosocial impact of tooth appearance is a largely unresearched field. This might either be due to the absence of validated, agreed measures and indices for classifying tooth appearance, or due to the fact that perception of aesthetics is an extremely subjective topic[5], influenced by socio-cultural values and individual preferences, making research difficult in the absence of objective markers.

To date, several indices have been used to categorize various enamel defects. Descriptive indices classify enamel defects based on their clinical appearance and include the Development Defects of Enamel Index[13], Jackson-Al-Alousi index[14], and Murray-Shaw index[15]. Fluorosis-specific indices such as Dean's index[16], Thylstrup and Ferjeskov (TF) index[10], Tooth Surface Index of Fluorosis[17], and Fluorosis Risk Index[18] deal exclusively with discolouration caused by excessive fluoride, and do not record other defects of enamel that may be present. There is, however, no index that takes into consideration all forms of tooth discolouration, including discolouration arising from dentine. Further, of the available indices, none addresses the issue of public perception and aesthetic impact or acceptability of these disfiguring defects. Research into the effect of tooth appearance on individuals is required to gauge the social burdens of tooth colour and surface alterations on the quality of life for those affected by it. An index to quantify tooth appearance would allow us to study the prevalence of such anomalies. It would also be a valuable tool for prioritising the need for cosmetic treatment in circumstances of limited resources. Additionally, such an index could also be used as a tool for patient counselling and to study treatment effectiveness. This paper therefore aims to:

- Develop an illustrated index for scoring all forms of tooth appearance based on a judging panel's perception of the attractiveness of variously affected dentitions;
- Test the index on a small sample of dentists for reproducibility and ease of use.

Materials and methods

Part 1

One hundred images of discoloured dentitions were selected from the photographs of patients who attended the Department of Paediatric Dentistry at Bristol Dental Hospital, UK. The criterion for selection was clarity of the image; tooth shape, and occlusion were

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

not considered. The selected source images included permanent and mixed dentitions affected by varying types, and severities of tooth discolouration and surface anomalies, including Amelogenesis and Dentinogenesis Imperfecta, fluorosis, isolated opacities, hypoplasia, loss of vitality, and idiopathic discolouration. Attempts were made to include as many variations of appearance as possible. The final sample included 98 digital photographs of dentitions with varying degrees of disfigurement, and two digital photographs of apparently unaffected dentitions.

Computer manipulation was used to eliminate the effect of tooth size, shape, alignment, occlusion, gingival colour, and contour on the perception of tooth colour and surface appearance (inspired by methods used by McKnight *et al.*[19] and Levy *et al.*[20]). A standard template of a well aligned 'ideal' dentition with the outline of six upper anterior teeth, adjacent healthy gingiva, part of the retracted upper lip, the tongue and 'shadow' of the teeth against the tongue was created digitally (see Fig. 1). The selected clinical photographs were then digitally superimposed on the standard template to prepare 100 standardized study images. Thus the only difference between the 100 study images was the degree and type of discolouration or surface defect.

The standard template and all study images were prepared using Jasc Paint Shop Pro® 8 running with Microsoft Windows® XP. A Hewlett Packard® deskjet 995C printer printed the images onto A4 sized Brilliant Hewlett Packard Premium Plus photo paper. Each image measured 9.5 cm × 15 cm surrounded by a 3 mm white margin. The 100 printed study images were hand shuffled for 2 min and numbered serially from 1 to 100. The number was written at the back of each image, to allow for data recording, and so that the images could be presented to each judge in the same shuffled order at each ranking session.

Fig. 1. Template for creating 100 standardised images.

The study images were tested on a group of 11 dentists and dental students at the Université Louis Pasteur, Strasbourg, and on five nondentally trained postgraduate students of The University of Bristol. The purpose of this test was to determine if the study images looked realistic, and if the judges were able to follow the instructions and rank them with some degree of reproducibility. (The instructions and ranking procedure were the same as used for the main study and will be described later). One of the images was reported to be blurry and was replaced with a new one before carrying out the main study. Analysis of data revealed good intrajudge reproducibility.

Twelve members of staff from Bristol Dental Hospital (six dentists, three male and three female; six lay staff, three of each sex, of matched ages to the dentists) were approached individually and requested to participate in the study. To improve levels of natural light when viewing the images, a table adjacent to a window was used to display them. The 100 study images were laid out serially according to their reference numbers, face up on the table, before the judges entered the room. The images were arranged in five rows with 20 images in each row, beginning with image number 1 placed at the top left hand corner, proceeding to image number 100 at the bottom right hand corner. The same room and table were used for each ranking session, and the images were presented to the judges in the same order each time. The methods described for ranking are similar to those used by Evans and Shaw in 1987 for the creation of the SCAN index[11].

Prior to beginning the first ranking session, each judge was tested to exclude colour blindness using five Ishihara's colour blindness test cards[21]. Details of the judge's age, sex and type of work were also noted. Judges were given written instructions asking them to lay out the images on the table in order,

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

according to their attractiveness as they perceived it. They were asked to place the most attractive image on the top left hand corner and proceed to arrange them in rows such that the least attractive one would be placed at the bottom right hand corner. They were given the option of allotting the same rank to two or more images if they found themselves unable to choose between them. Each judge ranked the images individually and at the end of each ranking session, the first author noted their final choices on a data recording sheet along with the time taken to complete the ranking.

To determine intraexaminer reproducibility, each judge was requested to perform the ranking on two separate occasions, separated by a week. Exactly the same procedure was followed for both sessions. After the second ranking session, the judges were asked to note down their reasons for ranking the images in their chosen order.

The mean rank and SD for each image over both ranking sessions for all 12 judges was calculated. The SPSS statistical package (SPSS Inc., Chicago, IL, USA) was then used to calculate the Pearson's correlation coefficients for reproducibility between each judge's first and second ranking session, as well as for interjudge agreement of scores to assess reliability.

The final rank order of the study images, from most attractive to least attractive, was obtained by arranging the images in ascending order according to their calculated mean position. In order to assess the stability of that mean rank, the SD of the positions were also calculated. To obtain the 10-point index, 10 images were selected, ostensibly separated by equal intervals in the final order of the study images. In order to ensure that the pictures used were those which were most consistently ranked by the judges, SDs of the selected images were compared to those of the images ranked one above or below the selected image. If the SD of the image ranked one above or below was lower than that of the selected image, the image with the lower standard deviation was chosen. Thus the chosen final ten images were those ranked at the 1st, 11th, 23rd, 34th, 46th, 57th, 68th, 78th, 88th and 99th positions. The images were arranged in a document in two columns in descending order with images one to five in column one, and images six to ten in column two. The images were numbered 1–10 and entitled 'IOTA', and are seen in Fig. 2.

Part 2

The second part of the study involved testing the 10-image index of tooth appearance (IOTA) for reproducibility and ease of use. For this purpose, it was decided to use the IOTA to score actual clinical photographs of patients with tooth discolouration and other surface defects. From the available scanned slides and digital photographs of patients attending the Department of Paediatric Dentistry at Bristol Dental School, 35 photographs of dentitions affected by tooth discolouration and surface defects were selected. The criterion for selection was a clear picture of at

Fig. 2. The Index of tooth appearance (IOTA) sheet.

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

least the upper anterior teeth and if possible, the entire dentition. Care was taken not to select the same source photographs from which the final 10 images of the IOTA were created. The photographs were zoomed into and cropped to give the appearance of the camera having captured the images from approximately the same distance.

The chosen 35 photographs included various forms of tooth discolouration and surface defects including Amelogenesis and Dentinogenesis imperfecta, fluorosis (of varying grades of severity), hypoplasia (chronological, idiopathic, and Turner teeth), nonvital teeth and isolated opacities. None were judged to be 'normal', and none appeared to have discolouration due to tetracycline staining.

Ten copies of the clinical photographs and the IOTA sheet were printed onto A4-sized Brilliant HP Premium Plus Photo paper using a Hewlett Packard® deskjet 995C printer. Each clinical photograph was 8 cm × 12 cm in size with a 3 mm white margin around it. After one set of 35 clinical photographs were printed and cut, they were hand shuffled for 2 min and each photograph was given a reference number from 1 to 35. All 10 copies of each photograph were then given the same reference number, which was written on the back of the photograph.

The IOTA research project was introduced to those attending the British Society of Paediatric Dentistry South West Branch study day held in April 2005. Attendees were given an information sheet outlining the study and were invited to take part in 'Part 2' of the research project. All attendees were qualified dentists, of whom 11 agreed to participate by filling in the reply slip accompanying the information sheet. From the received reply slips, one was eliminated as the handwriting was illegible. Each of the remaining 10 (henceforth judges) was sent a packet with an instruction sheet, a copy of the IOTA, 35 numbered clinical photographs, two scoring sheets, a personal details form, two envelopes for sealing the scoring sheets and a stamped self-addressed envelope.

The judges were asked to study the IOTA and the 35 clinical pictures carefully, in daylight, by a window; and to use the IOTA to score each photograph by selecting a representative image from the index, which they found most similar in attractiveness to the clinical photograph. It was emphasized that judges should determine overall attractiveness of the teeth and not look for a morphological similarity or to consider aetiology of the defect or discolouration *per se*. They were asked to use the reference numbers on the back of the photographs to note down their results on scoring sheet 1, along with the date and the time taken to complete the task. After completing scoring sheet 1, the judges were requested to seal it in the envelope labelled 'scoring sheet 1'. This was done to prevent them from referring to their scores from this session at the time of their second scoring session.

To evaluate the intrajudge reproducibility of the index, the judges were requested to repeat the scoring procedure after a 1-week interval, and to note down their results on scoring sheet 2. Scoring sheet 2 was to be sealed in the envelope labelled 'scoring sheet 2' at the end of the second scoring session. The judges were asked to return the two sealed envelopes in the self addressed and stamped envelope provided.

The mean score of each photograph at the first and the second scoring session were computed and the Kendall's Tau-b test was used to calculate the intra- and interjudge correlations as it is suitable for ordinal tied categorical data.

Ethical issues

In 'Part 1' the study participants were approached individually, the study was outlined to them and the task at hand (judging the 100 images) was explained. If they agreed to take part, they went ahead and ranked the images on two occasions. Their agreement to participate as judges in the study was taken to be their consent.

In terms of consent for the examiners (Part 2); these were all dentally qualified staff and at no risk of distress or concern at viewing such images. They were given written information sheets outlining the study with a reply slip. Only those who replied to us were included

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

as judges. Taking part in the study was taken as implied consent.

Within the standards of the time, consent was gained for the taking of the clinical photographs used in the study. The usefulness of images for record and study was explained. Patients, and the parents of children where appropriate, consented by permitting their photographs to be taken. Written records of that consent process were not made. The pictures used for the construction of the IOTA contained pixels captured from a large number of archival images, stretching back over many years. These were transferred onto an 'ideal' occlusion. The final images thus generated did not resemble individuals in any identifiable way.

Results

Part 1: construction of the IOTA

All 12 judges tested negative for colour blindness. Pearson's correlation coefficient for intraexaminer reproducibility ranged from 0.80 to 0.94. Demographics such as age, sex, whether or not dentally trained and level of dental training did not appear to be related to the consistency of rankings, shown in Table 1. Pearson's correlation coefficients for interjudge reliability were also calculated (Table 2). High correlations were found between the 12 judges for both their first and second ratings. The average correlation between all 12 judges for the first ranking session was 0.81, and for the second session was 0.79. For the dental judges alone, the average agreement within the group for the first and second ranking sessions was 0.82 and 0.77, respectively. In the nondental group the average agreement for the first and second ranking sessions was 0.83 and 0.84, respectively.

The average time taken by each judge for the first and second ranking sessions was 38.8 and 33.6 min, respectively. Judges responded to an open question asking what factors aided their decision making in the scoring process: reasons included intensity of discolouration, variation from 'normal' and symmetry of discolouration (Table 3). The SD and mean image rank of the 100 images is shown in Table 4. As is apparent, some images were consistently ranked at the same position by all the judges indicated by SDs as low as 0 to 1; whilst other images were ranked at

Table 1. Demographic details of the judges, time taken for the ranking sessions and Pearson's correlation coefficient for intrajudge reproducibility (Part 1).

Judge number	Sex	Age	Post held	Time (in min) taken for ranking session 1	Time (in min) taken for ranking session 2	Correlation between the scores given on first and second rankings
Dentally trained judges						
1	F	25	Senior house officer	15	15	0.83
2	M	55	Consultant orthodontist	55	52	0.94
3	F	36	Paediatric dentistry postgraduate student	22	20	0.93
4	M	34	Specialist registrar in orthodontics	31	20	0.90
5	M	40	Consultant in restorative dentistry	48	36	0.80
6	F	30	Paediatric dentistry postgraduate student	53	45	0.94
				Mean		0.89
Nondentally trained judges						
7	M	25	Office staff	39	34	0.92
8	F	34	Office staff	40	41	0.86
9	M	47	Office staff	49	45	0.94
10	F	23	Office staff	20	19	0.91
11	F	54	Office staff	43	26	0.80
12	M	47	Office staff	50	50	0.83
					Mean	0.88

Results for times taken to score the 100 photographs, and correlation between the judges' individual rankings over two sessions. Basic demographics are demonstrated to show range of ages, job title and sex of each judge.

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

Table 2. Pearson's correlation coefficients for interjudge reproducibility for the scores given on the first and second ranking sessions (Part 1).

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12
J1 × 1	1	0.82	0.68	0.80	0.80	0.81	0.77	0.78	0.70	0.78	0.65	0.68
J1 × 2	1	0.82	0.63	0.74	0.74	0.74	0.77	0.71	0.67	0.72	0.54	0.78
J2 × 1		1	0.71	0.88	0.89	0.86	0.83	0.77	0.69	0.80	0.61	0.64
J2 × 2		1	0.76	0.80	0.82	0.86	0.85	0.78	0.78	0.85	0.64	0.84
J3 × 1			1	0.82	0.78	0.82	0.87	0.84	0.88	0.90	0.74	0.80
J3 × 2			1	0.81	0.71	0.78	0.85	0.87	0.88	0.90	0.70	0.72
J4 × 1				1	0.88	0.86	0.87	0.85	0.82	0.90	0.69	0.76
J4 × 2				1	0.68	0.69	0.79	0.86	0.86	0.85	0.71	0.77
J5 × 1					1	0.85	0.86	0.81	0.78	0.85	0.66	0.71
J5 × 2					1	0.81	0.81	0.75	0.73	0.77	0.57	0.79
J6 × 1						1	0.90	0.82	0.77	0.87	0.65	0.69
J6 × 2						1	0.83	0.74	0.75	0.81	0.53	0.77
J7 × 1							1	0.87	0.87	0.92	0.72	0.76
J7 × 2							1	0.87	0.86	0.90	0.66	0.82
J8 × 1								1	0.83	0.88	0.70	0.74
J8 × 2								1	0.91	0.90	0.78	0.84
J9 × 1									1	0.89	0.82	0.84
J9 × 2									1	0.91	0.78	0.81
J10 × 1										1	0.74	0.82
J10 × 2										1	0.74	0.82
J11 × 1											1	0.83
J11 × 2											1	0.79
J12 × 1												1
J12 × 2												1

Average correlation for the first ranking session: (dental: 0.82; nondental: 0.83; dental and nondental: 0.81)
Average correlation for the second ranking session: (dental: 0.77; nondental: 0.84; dental and nondental: 0.79)
J1–J12 = judge numbers, × 1 = ranking session 1, × 2 = ranking session 2. As this represents interjudge reproducibility, the correlation for each judge compared with their own score = 1.0

Table 3. Judges' self-reported rationale for ranking the study images (Part 1).

Reason	Number of judges stating this reason
Intensity of discolouration, how far away from natural tooth colour	4
Colour of stains/patches–brown *versus* white	5
Contrast between colours	1
Area–small *versus* big areas, extent of discolouration on a single tooth, number of teeth affected	5
Pitting, surface irregularities	3
Symmetry, uniformity, isolated verses generalized discolouration	4

vastly different positions by different judges reflected by SDs as high as 25–29.

Part 2: preliminary evaluation of the IOTA

After reminder telephone calls, all 10 dentists chosen as judges for the second part of the research project returned the study packets with completed scoring sheets. The sample included three male and seven female judges with an average age of 44.3 years (range: 26–54 years).

The average time taken for the first scoring session was 15 min (range: 2–30 min). This reduced to 12.9 min (range: 3–30 min) for the second session. The average scoring time for each image for the first and second sessions in total was 25.7 and 22.1 s respectively. Table 5 demonstrates the demographic details and ranking time for each judge along with the correlations between the judges' first and second scores. Many of the judges were able to score the photographs at their first and second scoring attempts with a high degree of consistency. The correlations ranged from a high of 0.97 to a low of 0.73 with an average of 0.87.

The judges also demonstrated a fair amount of agreement with each other as is evident from an average correlation coefficient of 0.72 for the first scoring session and of 0.74 for the second session (Table 6).

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

Table 4. SD and mean image rank of the 100 study images arranged in final order (Part 1).

Image position in final order	Mean image rank	SD	Image position in final order	Mean image rank	SD
1	1	0	51	50.38	16.10
2	2.69	1.97	52	51.96	15.95
3	5.54	3.03	53	52.15	12.77
4	8.42	3.79	54	52.44	7.98
5	9	7.16	55	53.08	15.81
6	9.21	6.88	56	54.98	26.70
7	9.21	3.73	57	57.19	12.03
8	9.73	8.13	58	57.42	15.39
9	9.83	6.71	59	60.15	15.31
10	11.44	6.21	60	60.42	14.51
11	11.50	4.83	61	60.73	16.16
12	13.79	7.68	62	60.85	15.48
13	16.33	15.69	63	61.23	29.12
14	17.21	7.62	64	62.15	25.75
15	18.21	7.92	65	62.71	28.85
16	18.21	10.86	66	63.06	13.30
17	19.79	7.11	67	63.15	13.27
18	20.73	8.54	68	65.15	11.69
19	23.75	12.45	69	65.21	19.29
20	24.52	12.65	70	65.50	14.73
21	26.08	11.24	71	66.98	13.32
22	26.10	12.66	72	69.52	11.40
23	28.40	9.76	73	72.13	7.17
24	31.31	16.93	74	72.27	11.15
25	31.44	16.91	75	72.63	10.17
26	32.19	12.89	76	72.77	9.87
27	32.79	14.42	77	73	16.15
28	33.73	14.47	78	76.81	8.55
29	35.08	13.79	79	77.77	11.78
30	36.29	18.02	80	77.85	18.10
31	36.35	12.46	81	78.40	12.78
32	38	16.69	82	78.82	10.50
33	38.48	13.72	83	79.27	18.38
34	38.75	13.51	84	79.38	9.45
35	39.56	17.00	85	80.63	9.94
36	39.67	13.39	86	80.75	10.34
37	39.81	18.24	87	81.02	9.06
38	40.85	12.94	88	81.88	8.74
39	40.90	15.67	89	82.50	12.48
40	42.04	12.35	90	82.58	14.47
41	42.79	13.19	91	84.10	8.029
42	43.29	17.16	92	84.94	8.66
43	43.90	11.77	93	85.96	10.03
44	45.85	11.69	94	86.79	10.39
45	46.17	12.13	95	87.10	9.87
46	47.19	10.10	96	88.63	9.50
47	48.00	13.95	97	89.31	8.07
48	48.10	12.75	98	96.88	1.56
49	48.73	17.98	99	98.85	0.90
50	49.10	19.45	100	99.60	1.11

For each of the 35 clinical photographs, the IOTA scores allocated by all the judges at the first scoring session were summed and an average calculated (called the photograph mean score no. 1). The individual scores which had been allocated to each image by each of the 10 judges at the first session (called the photograph judge score no. 1) were then compared to the 'photograph mean score no. 1' to determine the deviation of each individual score from the panel's mean for the first session. The same analysis was repeated for the second scoring session (photograph mean score no. 2 was compared to photograph judge score no. 2). The scores allocated to each of the 35 photographs by all 10 judges over both scoring sessions were then summed to obtain an average score per photograph (the photograph mean score–average). An average for each judge's score for each photograph was obtained by calculating the mean of the two scores allocated by the judges to each image over both scoring sessions (the photograph judge score–average). The 'photograph mean score–average' and 'photograph judge score–average' for each photograph was then compared to determine the deviation of the individual judges' mean score from the panel's mean (Table 7). Most of the individual scores were within one score point of the panel's mean (82.6%, 84.6%, and 87.6% for the first session, second session and the average of the two sessions, respectively).

Discussion

This study sought to develop an index for quantifying all forms of tooth discolouration and associated surface defects. The basis for developing this index was the perceived attractiveness of various tooth colour and surface anomalies and not the aetiology, histology, or morphology of these defects. The IOTA was thus developed for the purpose of scoring dentitions disfigured by tooth surface and colour alterations of various aetiologies, based on society's perceptions of such anomalies.

It seems logical to assume that studies employing live subjects and actual clinical photographs as viewing stimuli are likely to have their outcomes biased by confounding factors such as morphology, gingival contour, and lip line. In the presence of multiple

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

Table 5. Dental judges' times for assessing dentitions against the IOTA (Part 2).

Judge number	Sex	Age	Qualifications	Type of practice	Time (in min) taken for scoring session 1	Time (in min) taken for scoring session 2	Correlation of scoring between the first and second scores
1	F	26	BDS	PDS & CDS	15	12	0.86
2	F	46	BDS	Clinical assistant–oral surgery	5	3	0.89
3	F	54	BDS	CDS	20	15	0.95
4	M	48	BDS	GDP	20	20	0.73
5	F	48	BA, BDS	GDP	2	4	0.87
6	F	52	BDS	DAC	20	15	0.90
7	M	49	BDS, PhD	Special care and paediatrics	20	15	0.79
8	F	43	BDS	DAC	11	9	0.77
9	F	28	BDS	GDP	7	6	0.92
10	M	49	BDS, MSc	CDS	30	30	0.97
Mean					15	12.9	0.87

Times taken and intrajudge correlation of scoring 35 clinical photographs of teeth of different appearance. Basic demographics are demonstrated to show sex, range of ages, type of practice and qualification of each judge. PDS: primary dental service; CDS: community dental service; GDP: general dental practice; DAC: dental access centre.

Table 6. Kendall's Tau-b coefficients for dental interjudge correlations for the first and second scoring sessions of IOTA (Part 2).

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10
J1 × 1	1	0.79	0.77	0.71	0.76	0.73	0.78	0.75	0.81	0.77
J1 × 2	1	0.85	0.82	0.69	0.88	0.83	0.78	0.68	0.80	0.76
J2 × 1		1	0.83	0.73	0.74	0.60	0.71	0.65	0.71	0.73
J2 × 2		1	0.83	0.69	0.78	0.71	0.75	0.64	0.72	0.70
J3 × 1			1	0.75	0.77	0.68	0.75	0.71	0.77	0.69
J3 × 2			1	0.76	0.80	0.79	0.74	0.74	0.82	0.71
J4 × 1				1	0.70	0.63	0.61	0.66	0.76	0.64
J4 × 2				1	0.64	0.67	0.58	0.66	0.71	0.59
J5 × 1					1	0.72	0.72	0.72	0.75	0.72
J5 × 2					1	0.79	0.75	0.66	0.75	0.78
J6 × 1						1	0.72	0.76	0.78	0.70
J6 × 2						1	0.73	0.78	0.84	0.73
J7 × 1							1	0.67	0.67	0.81
J7 × 2							1	0.66	0.71	0.79
J8 × 1								1	0.84	0.64
J8 × 2								1	0.69	0.65
J9 × 1									1	0.65
J9 x 2									1	0.67
J10 × 1										1
J10 × 2										1

Average correlation for the first scoring session: 0.72
Average correlation for the second scoring session: 0.74
J1–J10 = judge numbers, x 1 = scoring session 1, x 2 = scoring session 2. As this represents interjudge reproducibility, the correlation for each judge compared with their own score = 1.0.

confounders, it is not possible to eliminate bias and draw an individual's attention to the study variable alone. Therefore preparation of the study images by computer manipulation allowed for all features except the tooth appearance to be kept constant. Additionally, as study images depicted only teeth, the effect of background attractiveness and sex of the subject on the perception of attractiveness[22–25] was eliminated.

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

The IOTA images included only the six maxillary permanent anterior teeth. To be able to create realistic computer-generated representations of an entire dentition would have been difficult without extensive knowledge of computer graphics. In addition, as the labial surfaces of the maxillary incisors and canines are considered the most aesthetically important tooth surfaces in the mouth[26], it was felt that study images depicting these surfaces would suffice for determining the perceived attractiveness of a dentition. Other researchers have also used only the maxillary anterior teeth as the study stimulus[19,27–29].

Although individuals are likely to rate dental attractiveness significantly differently on different occasions, pooling the results obtained from a judging panel and increasing the size of the panel is likely to increase the validity of the results[23]. Keeping this in mind, the panel size in this study was increased from six as used by the developers of the SCAN[11], to 12. The judging panel for this study was constructed to represent dental and lay opinion for males and females over a wide range of demographic details. It is hoped that this panel acceptably reflected the views held by society at large as has been demonstrated by previous research using judging panels to rate dental attractiveness[22], However, the racial/ethnic profile of the study participants was not recorded in this study, so conclusions cannot be made with regard to this factor. For instance, it is possible that individuals living in an area where fluorosis is endemic might have judged those teeth affected with fluorosis differently from UK participants. Further work could assess the influence that this might have on the judgement of attractiveness.

Unlike the study that developed the SCAN[11], the present study included dentists in its panel. Some of the images used in this study were of severely discoloured dentitions. While it is likely that dentists may have come across such teeth at least during their training if not in their clinical practice, the lay judges may not have been exposed to such extremes of appearance. Hence dentists were included in the judging panel for two reasons; firstly, because they were more likely to be accustomed to the presented stimulus and were thus likely to increase stability of the results. Secondly, as the dentists were one of the intended future users of the index, it was considered appropriate to include dental opinion in its construction.

Although the lowest value for interjudge correlation was between a dentist and a lay person (0.53, between judges 6 and 11), the average correlation between dental and nondental judges for both ranking sessions was fair (0.81 and 0.79 for the first and second sessions, respectively). This suggests that on the whole the two groups show a fair degree of agreement even though some dentists and lay people may differ vastly in their aesthetic acceptability of tooth appearance.

In agreement with other work[30], some images in this study were ranked more consistently than others as reflected by a lower SD. The images anchoring the final rank order, that is the very attractive and very unattractive ones, were ranked with greater consistency than those in the middle of the rank order. This was probably because the images with minimal discolouration were easy to rank, as were the ones with very disfiguring colour and surface alterations. It is also possible that the judges, finding the markedly attractive and unattractive images easier to decide upon, ranked these first; after a certain point, fatigue is likely to have set in[27] which may have affected the judges ability to rank the remaining (the middle order) images effectively. In addition, the SD of images in the 'middle' of the rank order was always likely to be higher than that at either end of the scale owing to a greater 'choice' of rankings in the middle compared with the lowest and highest values. The judges' self reported rationale for ranking indicates that there are various ways that attractiveness is judged: some assess the difference in colour from the 'norm', others who consider pitting and other irregularities, and those who consider symmetry of discolouration. This may be one of the reasons for inconsistent rating of these images.

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

It is not possible for any illustrated index quantifying tooth appearance to include one representative image of each possible anomaly. This issue was addressed in the development of the IOTA by first establishing a continuum of perceived attractiveness of the study images and then selecting ten images separated by equal intervals as representative stops along the continuum.

When the IOTA was tested on a group of dentists it was found that they were able to use the IOTA with a high degree of agreement and in a short period of time. Training and calibration in the use of this index is likely to yield better agreement between examiners and more consistent scoring of clinical cases than that found in the current study. However the IOTA was tested on a small sample of judges. Further work is required to validate the IOTA index with a larger, more representative sample of participants. Validation of the index on patients in the clinical setting is required in order to increase its usefulness: it is possible that variation in lighting, the presence or absence of moisture, and clinical pressures (such as time) may in fact modify its usability or reliability. Additionally, further work to assess the relative importance of a malocclusion *versus* discolouration would be useful for paediatric dentists and orthodontists alike, when planning complex treatment.

Conclusions

An illustrated 10-point index for scoring all forms of tooth discolouration and surface defects was developed based on a representative panel's perception of the attractiveness of variously affected dentitions. It had good reproducibility and reliability and was acceptably easy and quick to use when tested. Further validation of this index requires tests on a larger sample and with members of the public. Possible applications of the index include research on the psychosocial impact of tooth discolouration, prioritization of treatment need, patient counselling, and evaluation of treatment success and effectiveness.

What this paper adds

- This paper reports the use of a novel index for measuring tooth attractiveness.
- This paper has demonstrated that teeth can be judged on their overall appearance irrespective of the aetiology, or histology of any discolouration.

Why this paper is important to paediatric dentists

- Paediatric dentists treat children presenting with a variety of tooth appearances, this paper could provide a scheme for prioritizing resource in their care.
- The index described in this paper could be used by paediatric dentists in research or for clinical assessment of treatment needs and success.

Contributions

Soma Modi carried out the bulk of the presented work as a Master of Clinical Dentistry (M Clin Dent) dissertation in the University of Bristol; Neil Davey taught and supported the electronic production of the images for all parts of the study; Lucy Williams edited the dissertation and expanded the literature supporting the issues raised in the paper; Rosemary Greenwood provided statistical support and analysis; Peter Crawford initiated and supervised the projects at all stages.

Acknowledgments

The authors would like to thank Sama Hameed, Alice Swarbrick and Rachel North who as undergraduates and with the generous help of Dr Angus Cameron of the Westmead Hospital, Sydney, carried out an earlier study which encouraged the development of the present work; and Dr Agnes Bloch-Zupan and her colleagues at the Université Louis Pasteur in Strasbourg, who with the help of a grant from COST (European Co-operation in the field of Scientific and Technical Research) helped with the initial pilot study. The authors would also like to thank Hazel Taylor at the University Hospitals Bristol NHS Foundation Trust for her assistance with the statistics presented here.

References

1 Rumsey N, Harcourt D. Visible Difference amongst children and adolescents: issues and interventions. *Dev Neurorehabil* 2007; **10**: 113–123.

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

2 Van der Geld P, Oosterveld P, Van Heck G, Kuijpers-Jagtman A. Smile attractiveness: self-perception and influence on personality. *Angle Orthodont* 2007; **77**: 759–765.
3 Kelleher MG, Roe FJ. The safety-in-use of 10% carbamide peroxide (Opalescence) for bleaching teeth under the supervision of a dentist. *Br Dent J* 1999; **187**: 190–194.
4 Kershaw S, Newton JT, Williams DM. The influence of tooth colour on the perceptions of personal characteristics among female dental patients: comparisons of unmodified, decayed and 'whitened' teeth. *Br Dent J* 2008; **204**: E9; discussion 256–257.
5 Alkhatib MN, Holt R, Bedi R. Prevalence of self-assessed tooth discolouration in the United Kingdom. *J Dent* 2004; **32**: 561–566.
6 Weintraub JA. Uses of oral health related quality of life measures in Public Health. *Community Dent Health* 1998; **15**: 8–12.
7 Shaw WC, Meek SC, Jones DS. Nicknames, teasing, harassment and the salience of dental features among school children. *Br J Orthod* 1980; **7**: 75–80.
8 Bos A, Hoogstraten J, Prahl-Andersen B. Dutch primary schoolchildren's attitudes toward their dental appearance. *Pediatr Dent* 2008; **30**: 439–442.
9 Do LG, Spencer A. Oral health-related quality of life of children by dental caries and fluorosis experience. *J Public Health Dent* 2007; **67**: 132–139.
10 Thylstrup A, Fejerskov O. Clinical appearance and surface distribution of dental fluorosis in permanent teeth in relation to histological changes. *Community Dent Oral Epidemiol* 1978; **6**: 315–328.
11 Evans R, Shaw W. Preliminary evaluation of an illustrated scale for rating dental attractiveness. *Eur J Orthodont* 1987; **9**: 314–318.
12 Goldstein RE. Esthetic dentistry – a health service? *J Dent Res* 1993; **72**: 641–642.
13 FDI Commisssion on oral health rae. An epidemiological index of developmental defects of dental enamel (DDE). *Int Dent J* 1982; **32**(2): 159–167.
14 Al-Alousi W, Jackson D, Crompton G, Jenkins OC. Enamel mottling in a fluoride and non-fluoride community, Parts I and II. *BrDent J* 1975; **138**: 9–15.
15 Murray JJ, Shaw L. Classification and prevalence of enamel opacities in the human deciduous and permanent dentitions. *Arch Oral Biol* 1979; **24**: 7–13.
16 Dean HT. Classification of mottled enamel diagnosis. *J Am Dent Assoc* 1934; **21**: 1421–1426.
17 Horowitz H, Driscoll WS, Meyers R, Heifetz S, Kingman A. A new method for assessing the prevalence of fluorosis- the tooth surface index of fluorosis. *J Am Dent Assoc* 1984; **109**: 37–41.
18 Pendrys DG. The fluorosis risk index: a method for investigating risk factors. *J Public Health Dent* 1990; **50**: 291–298.
19 McKnight CB, Levy SM, Cooper SE, Jakobsen JR, Warren JJ. A pilot study of dental students' esthetic perceptions of computer-generated mild dental fluorosis compared to other conditions. *J Public Health Dent* 1999; **59**: 18–23.
20 Levy SM, Warren JJ, Jakobsen JR. Follow-up study of dental students' esthetic perceptions of mild dental fluorosis. *Community Dent Oral Epidemiol* 2002; **30**: 24–28.
21 Ishihara S. *Tests for Colour Blindness*. 10th revised ed. Tokyo: Kanehara Shuppan, 1951.
22 Shaw WC. The influence of children's dentofacial appearance on their social attractiveness as judged by peers and lay adults. *Am J Orthod* 1981; **79**: 399–415.
23 Howells DJ, Shaw WC. The validity and reliability of ratings of dental and facial attractiveness for epidemiologic use. *Am J Orthod* 1985; **88**: 402–408.
24 Carlsson GE, Wagner IV, Odman P *et al.* An international comparative multicenter study of assessment of dental appearance using computer-aided image manipulation. *Int J Prosthodont* 1998; **11**: 246–254.
25 Eli I, Bar-Tal Y, Kostovetzki I. At first glance: social meanings of dental appearance. *J Public Health Dent* 2001; **61**: 150–154.
26 Horowitz HS, Driscoll WS, Meyers RJ, Heifetz SB, Kingman A. A new method for assessing the prevalence of dental fluorosis – the tooth surface index of fluorosis. *J Am Dent Assoc* 1984; **109**: 37–41.
27 Clark DC. Evaluation of aesthetics for the different classifications of the tooth surface index of fluorosis. *Community Dent Oral Epidemiol* 1995; **23**: 80–83.
28 Riordan PJ. Perceptions of dental fluorosis. *J Dent Res* 1993; **72**: 1268–1274.
29 van Palenstein Helderman WH, Mkasabuni E. Impact of dental fluorosis on the perception of well-being in an endemic fluorosis area in Tanzania. *Community Dent Oral Epidemiol* 1993; **21**: 243–244.
30 Roberts-Harry DP, Hathorn IS, Stephens CD. The ranking of facial attractiveness. *Eur J Orthod* 1992; **14**: 483–488.

© 2010 The Authors
Journal compilation © 2010 BSPD, IAPD and Blackwell Publishing Ltd

Copyright of International Journal of Paediatric Dentistry is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.