ED 095 096   SP 008 273

AUTHOR   Hunter, Harold G.; And Others
TITLE   Development and Evaluation of Self-Applied Plaque
                Indices for Children.
INSTITUTION   Human Resources Research Organization, Alexandria,
                Va.
SPONS AGENCY   National Institutes of Health (DHEW), Bethesda, Md.
                Div. of Dental Health.
REPORT NO   HumRRO-PP-10-74
PUB DATE   Jun 74
NOTE   15p.

EDRS PRICE   MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS   *Dental Health; *Dentistry; *Health Education;
                *Hygiene; *Preventive Medicine

ABSTRACT
        This paper reports on one of the main goals of
preventive dentistry, that is, encouraging children to remove plaque
at least once a day. Two self-scoring systems were combined with two
disclosants for a total of four experimental systems administered to
128 children. In the count method, the child counts the number of
stained teeth; in the rating method, the child selects one of five
color photographs that looks most like his own mouth. While both
methods appeared to be satisfactory for scoring plaque, the count
method does not depend on additional materials and is superior in
reliability and teachability. The authors state that while
self-scoring systems may not be satisfactory for routine evaluation
in a preventive program, they are reasonable substitutes for
professional indexes in epidemiological surveys. (Author)

HumRRO

ED 095096

Development and Evaluation of Self-Applied Plaque Indices for Children

Harold G. Hunter and C. Dennis Fink
Human Resources Research Organization

Charles L. Broring, Robert M. Morgenstein, and
Louis L. Lesche
Georgetown University School of Dentistry

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street ● Alexandria, Virginia 22314

June 1974

The Human Resources Research Organization (HumRRO) is a nonprofit corporation established in 1969 to conduct research in the field of training and education. It is a continuation of The George Washington University Human Resources Research Office. HumRRO's general purpose is to improve human performance, particularly in organizational settings, through behavioral and social science research, development, and consultation.

The contents of this report do not necessarily represent the
official opinion or policy of the sponsor of the HumRRO research.

PREFATORY NOTE

DEVELOPMENT AND EVALUATION OF
SELF-APPLIED PLAQUE INDICES FOR CHILDREN

PROBLEM

One approach to preventive dentistry would be to convince children to remove their plaque at least once a day.

If children are to remove their own plaque, they need some means to judge their own performance. A review of published plaque scoring systems (1-8) failed to reveal any that were designed for self-application. Development of a self-scoring method was considered important for several purposes:

- Self-evaluation of plaque removal.
- Evaluation by others, such as parents and peers.
- Evaluation of plaque-removal teaching strategies.
- A basis for public standards of oral hygiene.

Minimum requirements for a self-applied plaque-scoring system were established as follows:

- It should be easily learned and applied (i.e., in less than 10 minutes).
- Materials should be inexpensive and easily available.
- Self-scores should correlate well with scores taken by professionals, both on the self-scoring system used by the child, and on a standard (published) plaque-scoring system.

DEVELOPMENT OF THE EXPERIMENTAL SYSTEMS

Two scoring systems were devised for experimental evaluation: the Count Method and the Rating Method. Both require that the child's teeth be stained.

Under the Count Method, the child is asked to count the number of teeth showing any stain. The count is based on facial surfaces of the 16 most anterior teeth, including all four first bicuspids. These are all the teeth most children can see clearly using only a hand mirror. For simplicity, substitutions were not made for missing teeth.

Under the Rating Method, the child is asked to pick one of five color photographs that looks most like his own mouth. Photos range from clean to dirty.

Two commercial disclosants were compared, TRACE™ and PLAK-LITE™.[1] TRACE™ was selected because it is in standard use at the dental clinic where data were

---

[1] Endorsement of these products is neither intended nor implied by HumRRO or the National Institutes of Health.

collected, and because another study (6) recommended solutions over tablets for research purposes. The PLAK-LITE™ was used because it is a relatively new product whose disclosing effectiveness merited comparison with the effectiveness of TRACE™.

Evans (7) had developed a five-point photographic scale using TRACE™, for earlier research.[1] Each photograph in this scale showed a close-up of a normal set of teeth with finger retraction used to expose 14 to 16 teeth at the gum lines. Psychometric methods were employed to select five photographs which collectively depicted teeth ranging from being completely free of plaque to being almost completely covered with plaque, especially at the gum lines. This scale was requested from, and generously supplied by, the University of Houston. It constituted the basis for the Rating Method under TRACE™ disclosing.

A comparable set of five photographs was developed, using similar psychometric methods, for the Rating Method under PLAK-LITE™ disclosing.[2]

The PLAK-LITE™ photographs were taken with the PLAK-LITE™ as the sole source of lighting. The light was held approximately six inches away from and slightly above the mouth. Plastic retractors were used to expose as many teeth and gum line areas as possible. The film used was high-speed daylight Ektachrome (ASA 160) pushed to 400 ASA. Satisfactory photographs also can be obtained using GAF 500. An SLR camera was used with autobellows and a 135-mm lens. This arrangement allows close-ups from a distance of 8 to 10 inches. Exposure time was 1/60 of a second; the f-stop was 2.8.

To obtain high-quality prints, the 35-mm negatives were first converted to 4 x 5 inch inter-negatives. During this process, the yellowness of the flourescein-stained plaque was slightly accentuated. Glossy prints then were made from the inter-negatives. Persons wishing to duplicate these procedures should be cautioned that: (a) the PLAK-LITE™ must be held above the line between camera lens and object of photo; and (b) because of the extreme shallowness of the depth of field, the subject must be asked to "be steady" just before taking each shot. Using the same f-stop, it is recommended that each subject be photographed at speeds of 1/120, 1/60 and 1/30 of a second.

The two scoring methods combined with the two disclosants yielded a total of four experimental systems:

- TRACE-Count, or counting after disclosing with TRACE™
- TRACE-Rate, or rating after disclosing with TRACE™
- LITE-Count, or counting after disclosing with PLAK-LITE™ dye
- LITE-Rate, or rating after disclosing with PLAK-LITE™ dye

---

[1]Trained raters compared intraoral color photos with the five standards, in order to measure the effects of persuasive communications.

[2]A set of photographs consisting of the TRACE™ scale and the PLAK-LITE™ scale can be obtained from the Human Resources Research Organization for a fee of $8.00 per set.

RESEARCH METHOD

Research Design

Thirty-two children were assigned to each of the four experimental systems. For each child, the first of two professional examiners took a Patient Hygiene Performance (PHP) score (6) and an experimental score (E-score) using the system to which the child was assigned. The second examiner then took an independent PHP and E-score, taught the system to the child, asked for his or her self-score, and collected questionnaire data.

Questionnaire data were always collected last, and self-scores next to last. However, other sequences were counterbalanced within each experimental system as follows:

- PHP, then E-score   (examiner A)
  E-score, then PHP   (examiner B)   —   8 subjects

- E-score, then PHP   (examiner A)
  PHP, then E-score   (examiner B)   —   8 subjects

- PHP, then E-score   (examiner B)
  E-score, then PHP   (examiner A)   —   8 subjects

- E-score, then PHP   (examiner B)
  PHP, then E-score   (examiner A)   —   8 subjects

Sequences and treatments (experimental systems) were assigned at random to the 128 children, within subject quotas defined above.

Subjects

Subjects were children normally appearing at the Georgetown University clinic for routine care. Analyses of age and sex data, performed after the fact, showed the mean age to be about 11-12, with girls slightly older than boys (11.9 to 11.2). The sex ratio was well balanced, 66 girls and 61 boys. The four experimental groups were comparable, with respect to age and sex ratios.

Many of the subjects had been exposed to preventive concepts through previous clinic experience. However, these subjects were not identified for special handling.

Examiners

Examiners were faculty members[1] in the Pedodontics Department of the Georgetown University School of Dentistry.

PHP Scores

To estimate the validity of the children's self-scores, the examiners collected data using the Patient Hygiene Performance, or PHP index. This system was selected because it

[1] The examiners were Dr. Charles L. Broring, Dr. Robert M. Morgenstein, and Dr. Louis L. Lesche.

enjoys widespread acceptance in the professional community, and because it typically yielded extremely reliable data (6). For example, a pilot study to familiarize the examiners with the PHP yielded a Pearson product-moment correlation coefficient in the low .90s, based upon 30 subjects.

Self-Scores

Self-scores were collected from the children using procedures described earlier. Under the Count Method, the examiner indicated the teeth the child should look at, and asked how many were stained. Only four children had missing teeth. Their self-scores (and E-scores) were extrapolated upward to reflect a base of 16, for computational purposes.

Only two children refused to score themselves, both older girls (14 and 16) in the LITE-Rate group.

Questionnaire

Each child was asked three questions:
   (1) How much fun was it to score yourself?
   (2) How hard was it?
   (3) How often would you do it at home?
Responses were recorded as brief phrases or words, such as "easy," "pretty hard," "once a day," and so on.

RESULTS

For all analyses, except questionnaire data, actual group sizes were as follows:

TRACE-Count =	31	LITE-Count =	32
TRACE-Rate =	32	LITE-Rate =	30

Data were first analyzed for order effects. Recall that the first four activities performed by the two examiners were counterbalanced within each of the four experimental groups. It was desirable to rule out order effects, in order to avoid complicated analyses with respect to the primary experimental variables.

To test that the counterbalancing was effective, one-way analyses of variance were performed within each of the four experimental groups. Thus, data were treated in terms of the order in which they were collected. No significant F ratios were obtained. Based upon these assurances, all remaining analyses were performed without regard for the order in which data were collected. With the exception of questionnaire data, all analyses consisted of Pearson product-moment correlation coefficients computed from data within each of the four groups.

4

Validity

The validity of a self-score was defined as its correspondence with a PHP score, taken by a professional. Since both examiners took PHP scores on each child, two indications of validity were available for each self-scoring system. Correlations are as follows:

Examiner	TRACE-Count	TRACE-Rate	LITE-Count	LITE-Rate
A	.75	.66	.71	.73
B	.65	.70	.56	.88

Reliability

Reliability was defined as the degree to which the two examiners agreed between themselves with regard to their E-scores, or scores taken on the children's self-scoring system. These correlations were:

TRACE-Count	TRACE-Rate	LITE-Count	LITE-Rate
.93	.87	.96	.84

Teachability

Teachability was defined as the correspondence between self-scores and E-scores. Both examiners took an E-score on each child, as follows:

Examiner	TRACE-Count	TRACE-Rate	LITE-Count	LITE-Rate
A	.88	.62	.84	.70
B	.84	.74	.80	.59

Questionnaire Data

Since children were allowed to respond freely to examiner questions, their responses were categorized after the fact, and the frequencies within each category were compared.

(1) Fun. The first question was, "How much fun was it to score yourself?"

Responses were categorized as positive, neutral, negative, or missing. Inspection of the distributions of responses across the four experimental systems reveals few differences; no statistical analyses were performed.

	Positive	Neutral	Negative	Missing
TRACE-Count	18	9	3	2
TRACE-Rate	16	10	5	1
LITE-Count	18	8	6	0
LITE-Rate	15	10	4	3

(2) <u>Difficulty</u>. The second question was, "How hard was it to score yourself?"

Responses were categorized as positive (easy), negative (hard), or missing, and treated as for the first question.

	Positive	Negative	Missing
TRACE-Count	27	3	2
TRACE-Rate	30	1	1
LITE-Count	20	2	0
LITE-Rate	26	3	3

(3) <u>At Home</u>. The last question was, "How often would you score yourself at home?"

Responses referencing a daily or more frequent rate (e.g., "every day," "twice a day," etc.) were called "daily," anything based on a week was called "weekly," and everything else ("monthly," "once in a while," "when I feel like it," and "no") was labeled "other." Again, no differences were apparent across groups.

	Daily	Weekly	Other	Missing
TRACE-Count	19	8	2	3
TRACE-Rate	19	11	1	0
LITE-Count	15	11	5	1
LITE-Rate	16	10	3	3

DISCUSSION

Count Versus Rating Methods

The correlations for reliability and teachability suggest that the Count Method is more reliable and more easily taught than is the Rate Method. This is a reasonable suggestion. The criterion in the Count Method—based on counting the number of teeth showing plaque—is easy to apply. Conversely, the Rating Method requires a more global judgment. One can receive the impression of "plaqueness" if all teeth have some plaque at the gumlines, or if a few teeth are completely covered with plaque. Therefore, to some extent, a photographic scale for plaque is a product of two dimensions—number of teeth showing plaque, and extensiveness of plaque on any given tooth.

To estimate differences in reliability using the two methods the reliability correlations for the two Count groups (.93 and .96) were combined using Fisher's z. The correlations obtained using the Rating Method were similarly combined. The average correlations for the Count and Rating Methods were .95 and .86 respectively. These correlations were significantly different ($p < .05$).

Counting was also compared with Rating for teachability, using similar procedures. As an example, when the four individual teachability correlations for the Count Method

(.88, .84, .84, and .80) are combined, the resultant average correlation is .84. Similarly, combining the four teachability correlations for the Rate Method yielded an average correlation of .67. These two correlations (.84 and .67) are significantly different ($p < .01$).

Validity

The PHP (the standard chosen to validate the experimental system) is based on six surfaces, only two of which are facial anterior surfaces. Nevertheless, validity coefficients were moderately high throughout (as shown earlier), ranging from .56 to .88.

The entire issue of validity is highly arguable with respect to plaque-scoring systems in general. For example, no previously published plaque-scoring system is known to even mention validity, much less present data to support validity claims.[1] Indeed, only rarely are inter-rater reliability data presented. Rather, validity is implied on the basis of analytic inferences.

More defensible claims for validity would consist in demonstrations that plaque scores correspond well with independent measures of periodontal disease. To the authors' knowledge, no such data exist.

The most damaging argument against the experimental systems as valid plaque-scoring systems is that they sample only anterior facial surfaces. Plaque distribution studies consistently demonstrate that these surfaces are already the cleanest in the mouth (especially maxillary surfaces).

Nevertheless, the same studies usually show a consistent relationship among different areas of the mouth. For example, if the facial anteriors are clean, it can be predicted that other surfaces will be only slightly less clean. If the facial anteriors are dirty, other areas can be expected to be more dirty. These sorts of predictions are supported in the present study by the coefficients between the children's self-scores (facial anteriors) and the examiners' PHP scores (two facial anterior surfaces and four posterior surfaces).

Thus, all four self-scoring systems are reasonable substitutes for epidemiological purposes. Self-scores predict PHP scores with moderate accuracy, and are easier and quicker to obtain in quantity.

They are, however, suspect within the context of dental prevention, since they invite children to concentrate on just those areas that need the least work, and ignore the areas that deserve the most attention.

Reliability

The inter-professional reliability coefficients, using the experimental systems, were remarkable. The two Count Methods, in particular, showed the highest reliability

---

[1] Evans (7) claims validity for the TRACE℠ photographic scale in the sense that trained raters found reliable differences as a function of persuasive communications, using that scale.

coefficients ever reported for any plaque scale, .93 for TRACE-Count and an incredible .96 for LITE-Count. The two rating systems fared only slightly worse, .84 for LITE-Rate and .87 for TRACE-Rate.

Confidence limits calculated for these coefficients (using Fisher's z' transform) suggest that, for comparable subject populations and administrative procedures, the Count Method reliability should remain in the .90s about 95% of the time, and the Rate Method coefficients are likely to remain in the .80s.

Professional inter-rater reliability with respect to PHP scores was also high. To estimate differences in reliability using the two disclosants, the PHP coefficients for the two TRACE™ groups (.94 rating and .88 counting) were combined using Fisher's z. The PHP coefficients under PLAK-LITE™ (.76 counting and .91 rating) were similarly combined. Inter-rater reliability was significantly higher ($p < .05$) using TRACE™. However, TRACE™ was more familiar to the examiners, from prior experiences in the clinic.

Teachability

Time required to teach and obtain self-scores on the four systems was not clocked, because it was so short—less than a minute. Children took less time to rate than to count, but since both were so short, there seemed little to choose from.

In terms of child-examiner agreement on the self-scores, however, the data favored the count systems. Visual inspection of data from the four methods shows the rank order to be TRACE-Count, LITE-Count, TRACE-Rate, and LITE-Rate.

The teachability (child-examiner agreement) data are critical from practical considerations, since they indicate the degree to which a professional can trust a child's self-evaluation. For the Count Methods, agreement was in the middle .80s, and for the rating, in the upper .60s. The coefficients for the Count and for the Rate Methods were separately combined using Fisher's z procedures. The resultant average coefficients were .84 and .67 respectively. The difference between these two coefficients was tested for significance using a test for uncorrelated coefficients. The difference was found to be significant ($p < .01$) in favor of the Count Method.

As might be expected, the examiners tended to count more teeth as stained than did the children. Both groups counted more teeth as stained under the PLAK-LITE™ than under TRACE™.

	TRACE™		PLAK-LITE™	
	Examiners	Children	Examiners	Children
Mean Count	8.5	7.7	9.8	8.2

Clinical Impressions

The examiners preferred the photographs (rating system) over counting. It seemed easier for the children, and appeared to the examiners to carry greater motivational potential. The latter impression was not, however, supported by the questionnaire data.

Peer Evaluation

At the conclusion of each session, patients were paired in the order in which they were processed (e.g., the first with the second), and asked to score each other, using the system they had used on themselves. The purpose was to estimate whether peers could substitute for professionals. Because this activity had not been built into the original design, the pairs were not controlled in terms of scoring systems or disclosants. For example, TRACE-Count patients occasionally counted peers who had rated themselves under the PLAK-LITE©™.

Of the 64 pairs of subjects, both members of 50 pairs were taught the same method. For 25 of these pairs, the Count Method was used by the self-rater and the peer, and both members of the other 25 pairs used the Rating Method. In the remaining 14 pairs, one child had been trained using one type of disclosant and then rated his partner who had been stained using another disclosant. Inspection of the data indicated that this did not affect the ratings. Thus, the results were analyzed by method only. The findings are shown below.

	Average Self-Rating	Average Peer-Rating	t Test Comparison
Count Method	9.40	6.72	.05
Rate Method	2.40	2.20	n.s.

For both methods, the peer ratings were lower than the self-ratings. However, this difference was statistically significant only when using the Count Method.

The agreement between self and peer ratings was determined by calculating the reliability coefficient between the two ratings for each of the methods. For the Count Method, this correlation was .42, a value significantly different from zero. For the Rate Method, a reliability coefficient of .25 was obtained. This value is not significantly different from a zero correlation. It must be concluded, therefore, that peer ratings, as obtained under the conditions that prevailed during this study, do not provide a reliable index of the amount of plaque on another person's teeth.

CONCLUSIONS

Both counting and rating appeared satisfactory for scoring plaque, whether used by professionals or for self-scoring by the children themselves. With respect to reliability and teachability, the Count Method was found to be superior to the Rate Method. Also

favoring the Count Method is the fact that counting does not depend upon the presence of additional materials (photographs). However, counting does take somewhat longer than the Rate Method, and was less favored by the professional examiners than was rating from photographs.

With respect to the type of disclosant used, the inter-rater reliability coefficients using PLAK-LITE™ and TRACE™ (.85 and .92, respectively) were very high. However, there was a statistically significant difference in favor of TRACE™. With respect to method validity, reliability, and teachability, the differences using TRACE™ and PLAK-LITE™ were not significant.

The above findings suggest that the Count Method, using TRACE™ as the disclosant, is the preferred way to identify the presence and extensiveness of plaque. Supporting this conclusion is the practical observation that TRACE™ is the cheaper and more widely available disclosant. However, the routine use of TRACE™ may be less motivationally attractive than the use of the PLAK-LITE™.

The data suggested that children scored themselves more severely and consistently than they did peer partners, and that peer ratings tend to be unreliable.

Self-scoring systems of the type tested are reasonable substitutes for professional indices in epidemiological surveys, but perhaps not for routine evaluation in a preventive program, since they sample only facial anterior surfaces.

For children who are old enough to manipulate a mouth mirror and are on a routine preventive program, a whole-mouth self-scoring system might consist of yes-no (stain or no stain) discriminations per sextant, both facial and lingual. This would yield a 13-point (1-12) scale. Such a system should be tried out and evaluated.

REFERENCES

1. Green, J.C. and Vermillion, J.R. "Oral Hygiene Index: A Method for Classifying Oral Hygiene Status," *Journal of the American Dental Association*, vol. 61, no. 172, August 1960.

2. Green, J.C. and Vermillion, J.R. "The Simplified Oral Hygiene Index," *Journal of the American Dental Association*, vol. 68, January 1964, pp. 7-13.

3. Quigley, Gertrude A. and Hein, J.W. "Comparative Cleansing Efficiency of Manual and Power Brushing," *Journal of the American Dental Association*, vol. 65, 1962, pp. 26-29.

4. Turesky, S., Gilmore, N.D. and Glickman, I. "Reduced Plaque Formation by the Chloromethyl Analogue of Vitamin C," *Journal of Periodontology*, vol. 41, 1970, pp. 41-43.

5. Volpe, A.R., *et al.* "Antimicrobial Control of Bacterial Plaque and Calculus and the Effects of These Agents on Oral Flora," *Journal of Dental Research*, vol. 48, 1969, pp. 832-841.

6. Podshadley, A.G. and Haley, J.V. "A Method for Evaluating Oral Hygiene Performance," *Public Health Report*, vol. 83, March 1968, pp. 259-264.

7. Evans, R.I., *et al.* "New Measure of Effects of Persuasive Communications: A Chemical Indicator of Toothbrushing Behavior," *Psychological Reports*, vol. 23, December 1968, pp. 731-736.

8. Collier, D.R. and Williams, J.E. "The Evaluation of an Educational Program in Preventive Periodontics," *Journal of the Tennessee Dental Association*, vol. 48, April 1968, pp. 92-103.

16. Abstracts
A main goal of preventive dentistry is to encourage children to remove plaque at least once a day. Two self-scoring systems were combined with two disclosants for a total of four experimental systems administered to 128 children. In the first system, the Count Method, the child counts the number of stained teeth; the second system, the Rating Method, calls for a selection of one of five color photographs that looks most like the child's own mouth. While both methods appeared to be satisfactory for scoring plaque, the Count Method was superior in reliability and teachability. Also, the Count Method does not depend on additional materials (photographs). Since self-scoring systems sample only facial anterior surfaces, they may not be satisfactory for routine evaluation in a preventive program. They are, however, reasonable substitutes for professional indices in epidemiological surveys.

17. Key Words and Document Analysis. 17a. Descriptors

Dental care
Dentistry
Health
Medical services
Preventive medicine
Teeth

17b. Identifiers/Open-Ended Terms

Children
Plaque indices
Preventive dentistry
Self-application

17c. COSATI Field/Group 0605/0614/0616

18. Availability Statement
Distribution of this document is unlimited.

19. Security Class (This Report) UNCLASSIFIED

20. Security Class (This Page) UNCLASSIFIED

21. No. of Pages 12

22. Price