Clin Oral Invest (2005) 9: 7–24
DOI 10.1007/s00784-005-0018-z

CLASSICS IN CLINICAL DENTAL RESEARCH

John F. Cvar · Gunnar Ryge*

Reprint of Criteria for the clinical evaluation of dental restorative materials

Published online: 29 November 2005

Acknowledgements The criteria described in this report were developed by the former Materials and Technology Branch, Division of Dental Health, from August 1964 until February 1971. The Branch was responsible for an applied research program conducted to link basic physical, chemical, and biological properties of dental materials with clinical performance.

The work presented was begun in 1964 under the direction of the second author, now Assistant Dean for Research at the School of Dentistry, University of the Pacific, San Francisco, California. At that time, important conceptual contributions were made by Dr. Björn Hedegärd of the Odontologiska Kliniken, Stockholm, Sweden, in discussions with the second author and with Dr. R. J. McCune, now Director of Clinical Dental Research, Johnson and Johnson, New Brunswick, New Jersey and Dr. Richard Webber, of the National Institute of Dental Research, NIH. Miss Mildred Snyder developed many of the methods used to train and test examiners, and contributed greatly to the logical analysis of proposed criteria for evaluating dental restorations.

Others contributed to the formulation of the written criteria and helped to examine hundreds of restorations as the work progressed, including Drs. Bruce E. Johnson, Rudolph E. Micik, and Richard G. Weaver.

Special surveys were organized as part of residency projects conducted by Lt. Cols. Samuel C. Morgan and Warren A. Parker; their research reports are on file at the Dental Health Center. [1,2] Lt. Col. Parker deserves special recognition for undertaking numerous administrative duties during his tour of duty at the Dental Health Center.

The success of this work depended on the efforts of numerous other people, not all of whom can be succinctly listed. Progress was possible during all phases of the work through the efforts of Mrs. Irene Chavez, Administrative Assistant, and Miss Penelope Benton, Statistical Assistant.

First published in *U.S. Department of Health, Education, and Welfare, U.S. Public Health Service 790244,* San Francisco Printing Office 1971:1–42

See also introductory review: Bayne C, Schmalz G (2005) Reprinting the classic article on USPHS evaluation methods for measuring the clinical research performance of restorative materials. Clin Oral Invest 9, Issue 4

*Authors deceased

Contact: G. Schmalz (✉)
Poliklinik für Zahnerhaltung und Parodontologie,
Universität Regensburg,
Franz-Josef-Strauß-Allee 11,
93042 Regensburg, Germany
e-mail: Gottfried.schmalz@klinik.uni-regensburg.de
Tel.: +49-941-9446024
Fax: +49-941-9446025

Abstract Rating scales were developed for several factors that were considered relevant to the problem of clinically evaluating dental restorative materials. Examiners were trained to use the rating scales, and their performance was evaluated in field trials. Data analysis of examiner performance was used to revise the written criteria, and to train the examiners in making consistent judgments of dental restorations. Criteria were adopted when field testing indicated that examiners were able to duplicate their own judgments and judgments of other examiners at a predetermined level of acceptability.

Further experience with the rating scales in actual clinical studies led to the consolidation of anterior and posterior criteria, which had been developed separately, and to the deletion of certain rating scales which failed to

yield useful information. The rating scales which were finally adopted are for color match, cavo-survace marginal discoloration, anatomic form, marginal adaptation, and caries.

I. Introduction

Limited scientific data are available concerning the service life and clinical performance of dental restorations despite the fact that some restorative materials have been in constant use for many years. Silicate cement, for example, has been used for approximately seventy years, while dental amalgam has been used for about one hundred and thirty years. The major physical properties of these materials are well known, but the relation between physical properties and clinical performance remains a matter of considerable conjecture. Practicing dentists are therefore placed in the position of choosing among restorative materials with little clinical information to guide them. In recent years, their task has been complicated by the introduction of dozens of new restorative materials.

The lack of reliable information concerning the clinical performance of such materials is not due to lack of interest in performing the requisite research. Many researchers are acutely aware that clinical performance cannot be directly predicted from laboratory tests, but are discouraged from conducting appropriate research because of the lack of well-defined measures of clinical performance. Although current developments in engineering and technology offer some promise of permitting non-destructive tests of restorations to be conducted in clinical situations, the equipment needed may turn out to be very expensive, and the test procedures may be too time-consuming to be practical for many researchers. As an alternative, rating scales offer the possibility of producing meaningful clinical information, rapidly and inexpensively.

Rating scales are relatively rare in dental research, but have a long history of usage in other fields, notably psychology, where they have proven extremely valuable providing that they are carefully constructed and the raters are well-trained. The importance of training has not escaped the attention of dental researchers who have attempted to standardize examiners in making clinical judgments for the purpose of collecting survey information. Horowitz and Peterson[3] say, "Numerous surveys have been conducted to determine the prevalence of dental caries in given populations and to assess the efficacy of community water fluoridation. One of the problems in conducting such surveys and in evaluating their results when two or more examiners are used is the variability in diagnostic judgment that may exist between the examiners. Even when all examinations are made by a single examiner, it is difficult to compare results of one survey with others because of possible variations in the employed standards of diagnosis." Markén [4] remarks that, "The mere fact of being a dentist does not itself guarantee that the examiner will be an acceptable observer in a caries investigation." He adds that, "An observer must be well trained and the results of the training must be checked in some way."

These remarks are consistent with what is known by psychometricians concerning the use of rating scales. Guilford[5] says, "When a rater assigns ratings to a number of objects of a class, the frequency distribution is likely to show peculiarities attributable to the rater." He also remarks that "Within reasonable limits, nothing should be left undone to give the rater a clear, univocal conception of the continuum along which he is to evaluate objects and give all raters the same conception. The name of the trait is primarily useful as a label. Used without definition and without cues, it could be very misleading." Definitions, Guilford says, should be stated as much as possible in operational terms. Markén[4] also makes this point, saying, "Before starting a clinical trial it is necessary to define the criteria for the diagnosis of clinical caries and to use operational definitions."

The criteria for the evaluation of dental restorative materials presented in this report have been designed to measure clinically important features of restorations. Judgmental categories have been operationally defined, and have been arranged so that examiners arrive at ratings by making a series of bipolar decisions. For the convenience of the reader, the final rating scales are presented in the next section, followed by sections describing the conditions under which they are used, and on examiner training.

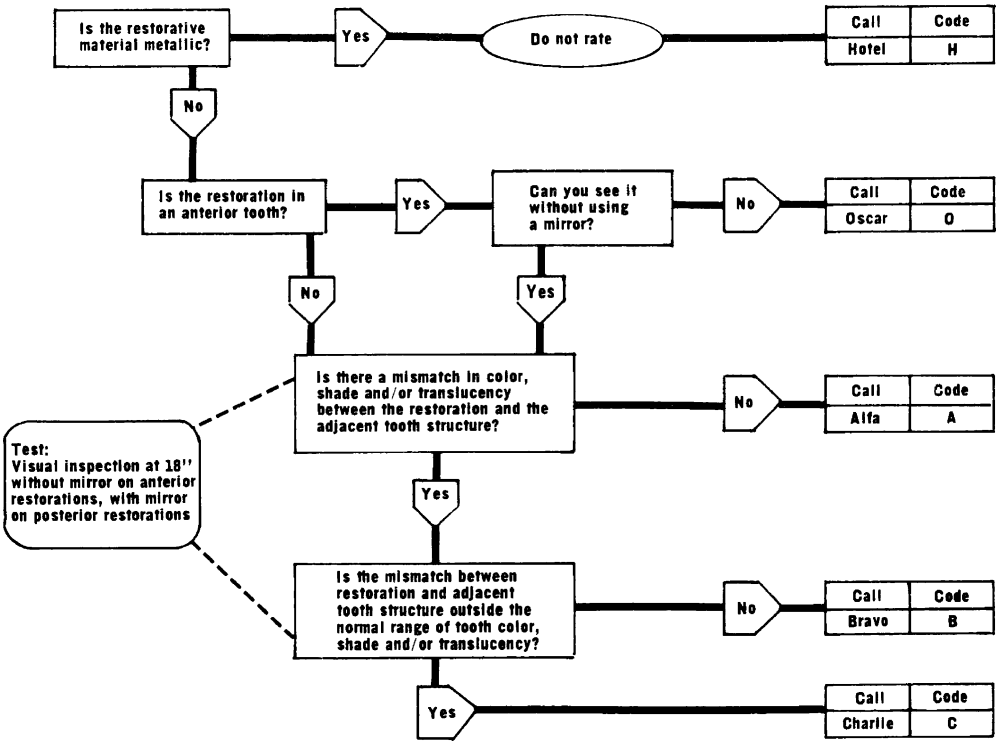
A section describing criteria development has been included as a part of Appendix I. Readers interested in methodology will find the appendices of value. Other may wish to limit their reading to the first four sections and the discussion in Section V.

II. The Criteria

The rating scales presented in this report are designed to reflect the aesthetic qualities and functional performance of restorations fabricated from a variety of dental restorative materials. The five characteristics represented are color match, cavo-surface marginal discoloration, anatomic form, marginal adaptation, and caries. Color match is judged on non-metallic restorations unless the restoration is located in an anterior tooth in a position where a mouth mirror must be used to see it. Poor color match is aesthetically displeasing, and may indicate chemical changes in restorative materials over a period of time. Judgments of color match are made at a distance of eighteen inches, equivalent to close conversational dis-

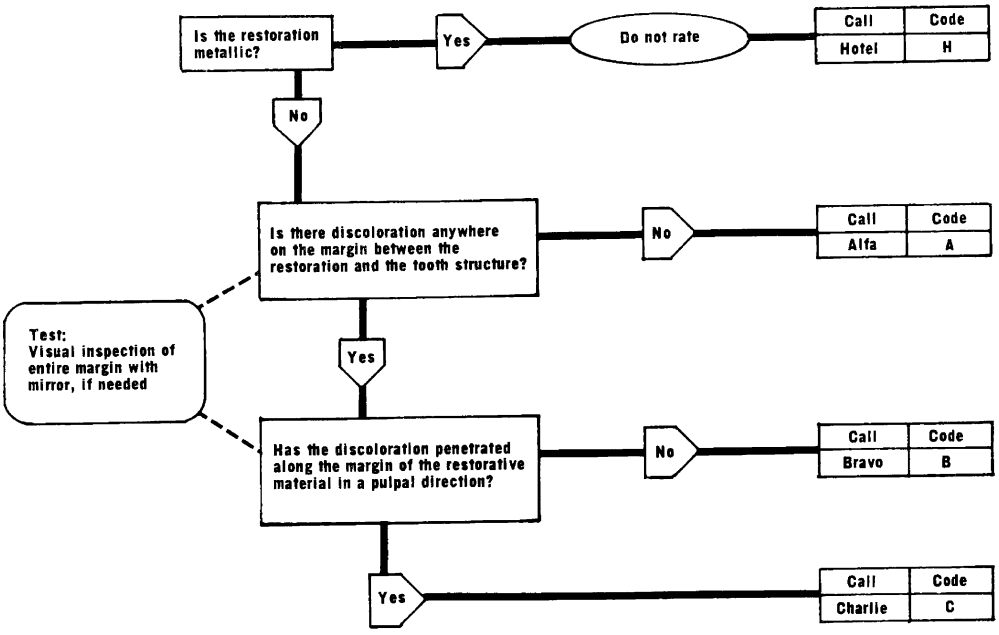
tance. The rating for color match, as for other characteristics, is reached by making a series of bipolar decisions, as indicated in the chart below:

COLOR MATCH



Cavo-surface marginal discoloration is discoloration at the interface of restoration and the tooth. If discoloration is severe, it is aesthetically displeasing; if it penetrates along the interface in a pulpal direction, it may indicate leakage, with a potential for caries processes to gain a foothold. Discoloration at the interface can also occur as a result of chemical reactions between restorative materials and liners.

CAVO SURFACE MARGINAL DISCOLORATION

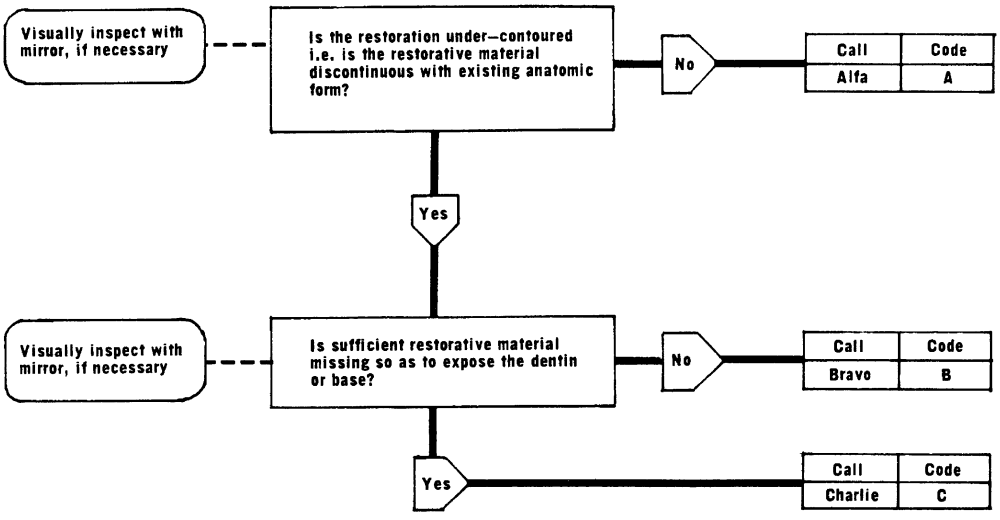


Anatomic form is a measure of loss of substance, and is useful in evaluating the clinical performance of restorative materials that are soluble or vulnerable to abrasion. Experience indicates that the most commonly used posterior restorative, dental amalgam, does not dissolve, while the performance of resins is largely unknown, and silicate cements and silicophosphate materials lose substance over a period of time. The clinical significance of loss of substance may vary; there is evidence that some materials dissolve, yet maintain a close adaptation with tooth tissue.

Exposure of dentin to oral fluids, bacteria, debris, and to thermal changes is considered to be damaging to tooth structure, and to offer a potential for decay to recur. The search for adhesive restorative materials has been largely motivated by a desire to gain the advantages that would accrue if restorations maintained close adaptation to tooth tissue indefinitely.

Caries at the margin of restorations is the final characteristic judged under these criteria. Some materials, notably silicate cements, are thought to inhibit decay at the

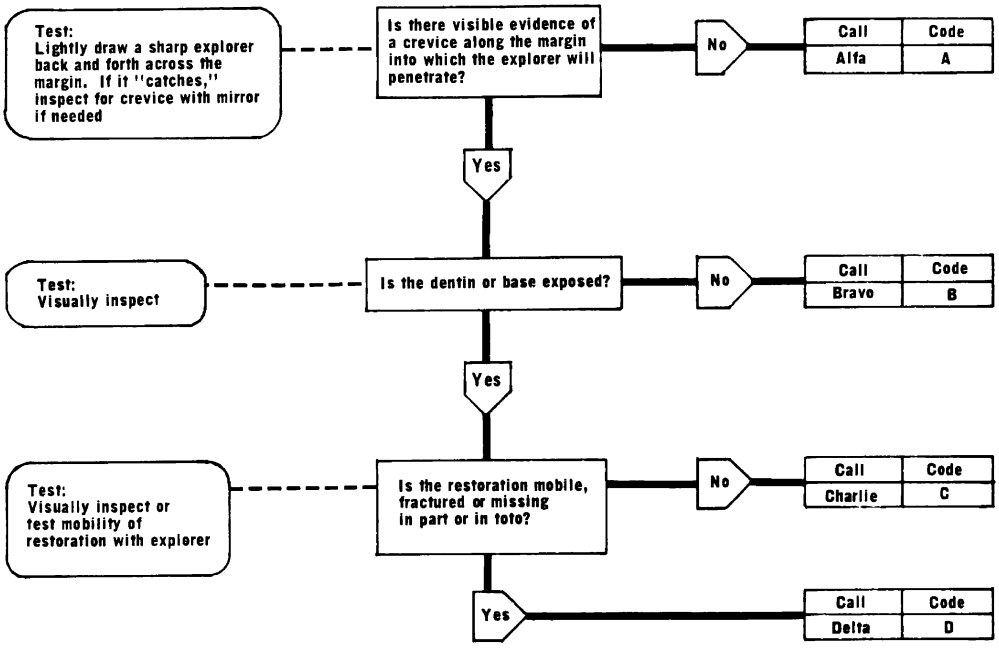
ANATOMIC FORM



Marginal adaptation is uniformly considered to be important in clinically evaluating restorations,[6,7,8,9,10,11,12] and is often investigated *in vivo* by laboratory methods.

interface of tooth and restoration because their solubility permits a continuous transport of fluoride ions from the restorative material to the tooth tissue. Some attempts have

MARGINAL ADAPTATION

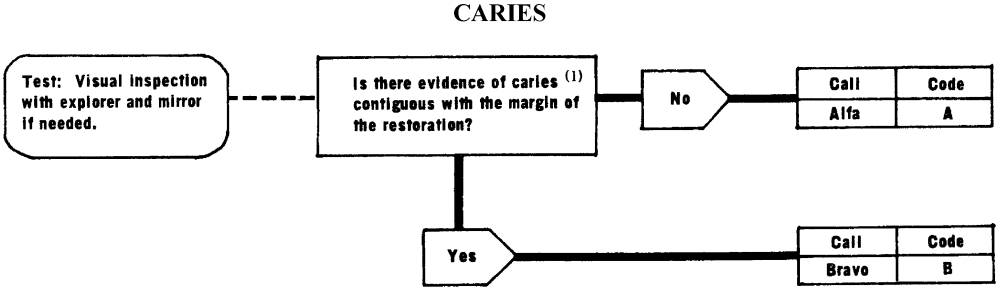


been made to incorporate fluoride in other dental materials, but in less soluble materials, the fluoride ion may not be available to the tooth tissue after initial placement. An otherwise good dental material that is non-soluble could be a second choice to the practitioner who is concerned about patients with poor oral hygiene, or patients considered to be highly susceptible to recurrent decay.

between paired restorations than is possible using rating scales, and it requires only the direct comparison of two objects, which is a relatively easy perceptual task.

Two trained examiners and a trained recorder are required to carry out the rating and ranking procedures.

The recorder is positioned so that he can hear the examiners easily, but they cannot readily see the recording



(1) An area at the restoration margin is carious if an explorer "catches" or resists removal after insertion with moderate to firm pressure, and is accompanied by one or more of the following:

a. Softness
b. Opacity at the margin, as evidence of undermining or demineralization
c. Etching or white spot as evidence of demineralization

An area at the margin is also considered carious if the explorer does not "catch", but conditions b or c are present.

III. Using the Criteria-Rating and Ranking

The written criteria presented in the last section constitute operationally defined rating scales for the judgment of five clinically important characteristics of dental restorations. Adequate training is essential if examiners are to use the rating scales reliably; this is discussed in Section IV. Although the rating scales could be used by trained examiners for many purposes, including assessing the work of dental students, they were specifically designed for comparing two different dental materials or two different dental procedures involving the same patient. For example, studies of conventional amalgam restorations vs. resin restorations are carried out by making two cavity preparations in similar teeth, and then assigning the two materials by reference to a table of random numbers. The patient thus serves as his own control.

Because two restorations can both receive "Bravo" ratings, yet one may be close to an "Alfa" while the other is nearly a "Charlie", a provision has been made for ranking two study restorations when they receive equal ratings. This procedure allows finer discriminations to be made

form. The duties of the recorder begin with naming the number of the tooth to be examined, the surfaces of the restoration, and the first characteristic, for example: " Number eight. Mesial. Color match." The examiner gives a rating in the phonetic code, for example, "Bravo." As the recorder writes "B" in the appropriate box of the evaluation record, he names the next characteristic. If the examiner takes too long to give a rating, the recorder prompts him by naming the characteristic again. Less experienced examiners usually take longer to decide upon a rating than their more experienced counterparts, but seem to perform best when they are forced to move along at a reasonably fast pace. Experienced examiners usually do not find themselves distracted by features of the restoration that are not included under the characteristics of the criteria, and often are able to complete the evaluation of restoration in ten seconds or less.

When the first examiner has evaluated both members of the pair, the second examiner takes his turn. If both examiners agree, the rating becomes final; if they do not agree, the recorder requests that a joint examination be conducted, and that the examiners agree on a final rating.

Most often, disagreements are resolved by adopting the less favorable rating given previously. Usually, one examiner has failed to notice a defect seen by the other examiner and accedes when it is demonstrated.

If the final ratings for the test restoration and the control restoration differ, a ranking of "one" is automatically assigned to the superior member of the pair, while the other is ranked "two". If the final ratings are the same, however, the examiners independently rank the restorations, using ranks of one, two and zero. Zero is applied to both restorations when an examiner cannot judge one to be superior to the other with respect to the characteristic under consideration. Ranks of "zero" are automatically assigned in the case of tied "Alfa" ratings for cavo-surface marginal discoloration, anatomic form and caries. For these characteristics, "Alfa" means, respectively, no marginal discoloration, no loss of substance, and no caries, so nothing would be gained by employing a ranking procedure.

The examination procedure given above requires that the participants accept certain roles, some of which are not commonly practiced by medical and dental personnel. Perhaps the most important difference is in the relation between the recorder (often a dental assistant) and the examiner (often a dentist). During examinations, the recorder is responsible for directing the rating procedures, which may include prodding examiners to rate restorations more quickly, or preventing them from becoming inquisitive about written ratings. The recorder, in short, occasionally issues directions to the examiners, who must learn to accept him as a director.

IV. Examiner Training

As stated in the introduction, well-trained examiners are essential if objects are to be rated with any degree of consistency. The objective of training is to attain consistency of three kinds: first, examiners should agree with each other; second, they should agree with their own judgments from one occasion to another; and finally, their judgments should be anchored in some way to prevent drift over a period of time. It is conceivable that examiners could replicate their own judgments quite well, and agree with each other, yet be subject to group drift. Photographs and models are useful in anchoring definitions which specify the characteristics being rated. A high level of inter-examiner agreement and intra-examiner agreement can to some degree be taken as evidence that drift is minimal, providing that there are several examiners being tested. It is less probable that group drift has occurred for eight to ten examiners who are in agreement with each other than for a group consisting of two or three examiners.

It is helpful to teach examiners that the goal of conducting examinations for research purposes is different from the goal of examining patients for treatment. In the latter case, all possible cues should be utilized in order to determine an optimal course of therapy; examinations conducted for survey or research purposes, however, usually have the goal of producing comparative data that reflect the status of two or more populations. Dentists have been trained to respond to subtle perceptual cues for the purpose of diagnosing individual cases, and find it difficult to ignore defects in restorations which are not covered by the written criteria. There may be some tendency, for example, for naive examiners to be dissatisfied with the rating scales when they are examining a restoration with an overhang. In training examiners, it is worth pointing out repeatedly that clinical conditions of this sort are as likely to appear in test groups as in control groups.

After the concept of rating scales has been explained, the next step for the trainee is to learn the written criteria. Printed training aids and tape-slide presentations are used to establish verbal knowledge of the scales, and to teach the order of the characteristics to be judged during examinations.

Following this, an instructor demonstrates the use of the rating scales, using photographs and models where possible. The next stage is to examine patients clinically, first with the instructor explaining how he arrives at several ratings, and then by obtaining independent examinations for the purpose of comparing trainee ratings with those given by the instructor.

Final testing consists of examining a large number of restorations (as many as 200) on two occasions, with several days separation between the examinations. Acceptable performance is defined as 85 % intra and inter-examiner agreement. Sequential analysis (see Appendix III) may be used to obtain a quick evaluation of whether the goal has been met, and if it has not, analysis of the nature of disagreements is performed, and further training is scheduled.

Once the desired performance levels have been attained, examiners can be utilized to evaluate paired restorations placed in clinical studies. Disagreements can be tabulated according to whether they are "Alfa-Bravo," "Alfa-Charlie," and so on. This will assist in detecting examiner drift, which may occur over a period of time. However, periodic review of the concepts underlying the use of rating scales, and periodic calibration of examiners in a test-retest situation is essential to monitor this measurement system.

V. Discussion

The rating scales for selected characteristics of dental restorations have been described, and examiner training has been discussed. In retrospect, the development and usage of such scales appears to be fairly straightforward. The unwary researcher should be warned, however, that a certain amount of confusion is likely to occur when investigators first attempt to construct such scales.

One source of confusion is the suspicion that rating scales are too subjective for use in measuring physical objects, and that physical testing of the objects is essential to obtain reliable and accurate information. In the presence of such doubts, the failure of the rating scales to measure certain attributes may cause grave misgivings on the part of trainees, which in turn can interfere with learning the

correct use of the scales. It is therefore worth examining the relation between physical testing and clinical evaluation of the sort described in this report.

Physical testing can be conveniently divided into two categories, depending on the intent of the investigator. One type of testing is intended to relate one physical property to another in order to contribute facts toward a body of theory. One of the end products, presumably, would be better materials, designed on theoretical grounds. The other type of test, as exemplified by American Dental Association specification tests, might be characterized as product testing. The intent of this sort of testing is to keep inferior materials out of the market place, and to encourage the development of superior products. Although this division is somewhat artificial in that there is a large amount of interplay between product testing and theoretical investigation, it does serve to point out that product testing does not presuppose a body of supporting theory. To know the extent to which a product performs its intended function requires that performance be measured; the facts that explain the performance are not essential, however desirable they may be for other reasons.

Perhaps the easiest way to relate clinical and physical performance tests is by way of an imaginary example: a dentist places some silicate cement restorations in the teeth of a patient, and using the same batch of material, makes some specimens. After conducting solubility tests, he somehow concludes that the restorations will not last for more than a year. Does he recall the patient in one year, and automatically replace the restorations? Clearly, he does not: if the margins of the restoration appear to be in good condition, there is no evidence of caries, and the patient has no complaints, the restorations will probably remain. One can easily imagine the reverse situation. The point is that as far as product testing is concerned, clinical results are primary; laboratory results are valuable to the extent that they are reliable predictors of clinical performance.

The rating scales have proven to yield useful information in the hands of highly-trained examiners. The routine for using the scales to rate restorations stabilizes the final ratings which constitute the raw data for clinically assessing dental materials. In other words, test-retest sessions, aimed at producing eighty-five percent inter and intra-examiner agreement, are a rather harsh test of the usefulness of the rating scales, because during the test sessions restorations are not judged in pairs. In actual clinical research, trained examiners have usually agreed more than ninety percent of the time, and report little difficulty in reaching final ratings. Despite this, frequent retraining and testing of examiners is essential if the rating scales presented in this report are to be reliable measures of the clinical performance of dental restorative materials.

Appendices

Appendix I describes Criteria Development and provides data from a study to develop rating scales for marginal adaptation. Appendix II provides data from studies to develop rating scales for evaluating anterior restorations [1] and posterior restorations [2]. Two characteristics developed during these studies have subsequently been eliminated. These are dark deep discoloration (anterior) and surface texture (posterior), both of which proved to be highly susceptible to examiner drift. Contour (anterior) and anatomic form (posterior) were nearly identical in wording, as were marginal integrity (anterior) and marginal adaptation (posterior), and were readily consolidated into a single system for examining both anterior and posterior restorations. The "Oskar" category for marginal adaptation was eliminated. Data on dark deep discoloration and surface texture are presented for the sake of historical accuracy, and to illustrate the point that reaching acceptable levels of performance in training sessions does not guarantee acceptable performance later on. There are no statistical data concerning examiner performance in judging caries at the margins of restorations because it is very difficult to locate a study population having a sufficient proportion of caries at the margin to warrant a special study. Examiner agreement on caries at margins has proved to be higher than for other characteristics judged during the course of clinical studies. However, this could be explained by assuming that examines can usually agree that caries is not present, particularly in studies conducted two to three years following placement.

Appendix III presents those statistical methods that are appropriate for data derived from rating scales used in clinical trials. Appendix IV contains references.

Appendix I

Criteria Development

The historical development of the rating scale for one characteristic (marginal adaptation) will be traced in this section to illustrate the methodology employed in developing rating scales for all the characteristics which comprise the criteria for evaluating dental restorative materials. Separate criteria were originally developed for evaluating anterior and posterior restorations, utilizing separate field trials. Data pertaining to these trials are provided in this Appendix.

The following outline summarizes the steps which were taken to select characteristics and develop associated rating scales:

1. Literature review and discussion to select relevant characteristics for clinical evaluation.
2. Development of written criteria to describe each characteristic selected for evaluation.
3. Clinical trial of the criteria using a small number of patients, followed by discussion among the examiners, and modification of the written criteria to remove ambiguities and to more closely specify the operational definitions of judgmental categories.
4. Consultation with a statistician to remove logical inconsistencies in the written criteria and to arrange

judgmental categories so none were superfluous and none captured an overwhelming majority of responses.
5. Training of examiners in using revised criteria.
6. Testing of examiners using criteria in a survey.
7. Criteria revision, examiner re-training, and further testing as needed.

The closeness of adaptation between restorative material and tooth structure is a characteristic that most investigators agree must be assessed in evaluating restorations [6,7,8,9,10,11]. Closeness of adaptation has been investigated by clinical assessment and by laboratory methods involving the measurement of dye penetration and the tracing of radioactive isotopes along the interface[12]. Although written criteria had previously been developed by the Materials and Technology Branch to assess the marginal integrity of anterior restorations,[1,13] the development of the first written criteria for marginal adaptation was carried out as if no previous model existed.

Phonetic code words were used to reduce misunderstandings when ratings were given orally by examiners. Alphabetic ratings also emphasize that the rating scales are considered to be ordinal, not interval. The first criteria for marginal adaptation were written as follows.

Code Word	
Alfa	The explorer does not "catch" when drawn across the restoration-tooth margin either from tooth to restoration or from restoration to tooth. If a "catch" exists, there is no visible crevice along the periphery of the restoration. The edge of the restoration appears to adapt closely to the tooth structure along the entire periphery of the restoration.
Bravo	The explorer does "catch" and there is visible evidence of a crevice into which the explorer will penetrate, indicating that the edge of the restoration does not closely adapt to the tooth structure. The dentin or base is not exposed, and the restoration is not mobile, fractured, or missing in part or in toto.
Charlie	The explorer penetrates into crevice indicating that a space exists between the restoration and the tooth structure. The dentin or the base is exposed at the periphery, but the restoration is not mobile, fractured, or missing in part or in toto.
Delta	The restoration is mobile, fractured, or missing in part or in toto.
Oscar	Marginal adaptation cannot be assessed due to an excess of restorative material at the margin.

Five dentists were trained in the use of the criteria at the Dental Health Center. During the first session the rationale for the criteria, the rating system, the coding system, and the record forms were explained and discussed. Ten restorations were then rated by the instructor to clinically illustrate the rating system. After the instructor explained the reason for assigning each rating the trainees examined the same restorations. Each trainee was encouraged to explain his interpretation of each characteristic and thus his reasons for agreement or disagreement with ratings assigned by the instructor. Where disagreements occurred, the categories were again explained so that all examiners would invoke the same concepts when using the rating scales.

A statistical consultant prepared notes on the training session, which were distributed to each examiner prior to the second session. The notes contained a resume of the discussion during the first training session. The following comments were noted:

1. All visible margins are to be examined.
2. Code Delta is used when the restoration is grossly fractured, that is, a fracture at the isthmus or when there is a fracture more than 1/2 mm from the margin or where the restoration is missing more than 1/2 mm from the margin. Fractures or loss of material less than 1/2 mm from the margin should be classified as Bravo or Charlie. Overt secondary caries is also included in this category.
3. Code Bravo should be used only when there is a visible crevice where the explorer catches.
4. Code Charlie should be used when there is evidence of secondary caries at the margin.

The examination procedures employed during the second training session simulated those to be used in the field surveys. The session was conducted in two phases. On the first day the instructor and each of the trainees independently examined 29 posterior restorations. The ratings were recorded and discussion was not allowed. Two days later the examiners rated the same 29 restorations for a second time. Two additional patients were examined to keep the examiners occupied, and to reduce discussion about the criteria between examinations. When the second examination was completed for all examiners, a discussion period was held with the patients present. Any restoration which presented a rating problem for any examiner was reviewed and discussed. Notes were prepared from this session also, and distributed to the examiners at a staff meeting prior to the first field survey. The comments from this session were as follows:

1. The objective is to assess the adaptation of the restorative material to the tooth structure at the margin.
2. Code Charlie should be used when there is evidence of secondary caries in dentin at the margin, or any exposure of the dentin or base at the margin.

The ratings assigned by the examiners during the training session were tallied on a sequential analysis graph to determine if an acceptable performance level had been attained. Performance was considered acceptable during this phase of training when examiners agreed with their own judgments and with consensus judgments more than 75 percent of the time. (Consensus was defined as a majority opinion.) Acceptable performance levels were attained, so the first field survey was scheduled at a nearby Coast Guard Station. The examiners were recruits, mainly eighteen to twenty years of age. Each restoration was

examined twice by the five examiners, the first and the second examination being separated by two days. For both examinations combined, there were a total of 1,280 judgments of marginal adaptation and 128 restorations. This distribution of ratings is given below in Table 1:

Table 2 provides the consensus ratings for the 128 restorations that were examined in Survey Number One, and Table 3 provides the ratings that were given by each examiner. Examiners A, B, C had previously been calibrated in using rating scales developed for evaluating anterior restorations, but examiners A and C had considerably more experience in rating restorations placed for comparative studies of dental materials. Examiners Y and Z were attempting to use rating scales for the first time. The relatively small number of "Charlie" and "Delta" ratings seemed reasonable in view of the age of recruits, and the probability that their restorations were not very old.

Table 4 provides the duplicate ratings given by each examiner from one examination to the next. The most experienced examiners were best able to duplicate their judgments, while the least experienced examiners did not perform as well. Only one examiner achieved the goal of eighty-five percent self-agreement, so further training was considered necessary.

Table 1 All ratings: marginal adaptation (survey number one)

	Total	Oscar	Alfa	Bravo	Charlie	Delta
Number	1,280	–	561	651	47	21
Percent	100.0	–	43.9	50.8	3.7	1.6

Table 2 Consensus ratings: marginal adaptation (number and percent; survey number one)

	Total	Oscar	Alfa	Bravo	Charlie	Delta
Number	128	–	59	66	2	1

Table 3 Marginal adaptation: ratings by examiner (survey number one)

Examiner	Examination	Total	Rating (%)				
			0	A	B	C	D
A	1	100.0	–	46.1	48.4	4.7	0.8
	2	100.0	–	47.6	49.2	3.1	–
B	1	100.0	–	54.0	43.7	2.3	–
	2	100.0	–	64.8	33.6	1.6	–
C	1	100.0	–	43.0	53.1	1.6	2.3
	2	100.0	–	43.8	52.3	0.8	3.1
Y	1	100.0	–	50.8	35.9	9.4	3.9
	2	100.0	–	38.2	47.7	9.4	4.7
Z	1	100.0	–	37.6	60.0	1.6	0.8
	2	100.0	–	13.3	83.6	2.3	0.8

Table 4 Duplicate ratings by examiner, marginal adaptation (survey number one)

Examiner	Total	Duplicate ratings	
		Number	Percent
A	128	110	85.9
B	128	101	78.9
C	128	105	82.0
Y	128	89	69.5
Z	128	90	70.3

Tabulations providing the nature if disagreements (for examiners compared with consensus and examiners compared with self) were obtained for Survey Number One, but owing to the small number of "Charlie" and "Delta" ratings, they were not particularly useful in deciding how to modify the written criteria for marginal adaptation. Generally, most disagreements were between "Alfa" and "Bravo" ratings.

Survey Number Two

Survey Number Two was based on the course of action a practicing dentist would be likely to follow in a clinical situation; that is, to re-examine a restoration in six months, to replace it for preventive reasons, or to replace it immediately because of damage to the tooth structure. The examiners were asked to separately list any aesthetic comments they might have concerning a restoration. The instructions provided to the examiners are duplicated below:

> Instructions: Assign *one* of the following ratings to each restoration, and give your reason for the rating. More than one reason may be given. Do not base the rating on aesthetic qualities. If you have a comment on the aesthetic qualities of the restoration, enter it under "b" in the comment box.

Alfa	Bravo	Charlie	Delta
Replacement unnecessary	Replacement Questionable	Replace for preventive reasons	Replace immediately

Survey Number Two utilized 22 patients from a Sixth United States Army unit stationed at Fort Baker near San Francisco. One hundred twenty-two restorations were examined by the same examiners that had participated in the first survey.

Analysis of results consisted of arranging the reasons given for ratings in a matrix which revealed how often factors were named as the sole reason for ratings and how often they were named in conjunction with other factors. For 582 non-Alfa ratings, factors were mentioned 205 times without other factors being named. Of these 205, "Margin" or "Open Margin" accounted for nearly seventy percent of the total, "Caries" for about ten percent,

"Inadequate Extension" for about eight percent and "Fractured Margin" for about five percent. The ratings associated with these factors are presented in Tables 5, 5a and 5b).

It can be seen that the most prevalent reason given for non-Alfa ratings was poor margins, and that most of these were assigned a "Bravo" rating, meaning that the examiner would like to see the restoration again in six months. Half of the ratings associated with inadequate extension were "Charlie," meaning that the restoration should be replaced for preventive reasons. When caries was mentioned as the sole reason for giving a rating, the rating was always "Delta."

Since most of the participants in Survey Number Two had some experience in using rating scales, survey results were checked by having several dentists not employed at the Dental Health Center and not familiar with the written criteria perform the same task. The results indicated that both trained and untrained examiners give the same reasons for assigning ratings to restorations, in the same order of importance.

The major conclusions regarding the criteria as used in the first survey, based on the results of surveys one and two, were that the written criteria described factors that are relevant clinical indicators of the status of restorations, but that some changes in oral instructions for using the criteria needed to be made in order to have the criteria reflect single rather than composite factors, and to restrict the examiner's judgment to what could be seen rather than what might be inferred. A new recording form was designed for Survey Number Three, with boxes for examiners to indicate the presence of caries and discoloration. Oral instructions were modified to eliminate the presence of caries or discoloration as a reason for assigning a rating for marginal adaptation. In other words, judgments were confined to the operational definitions presented in the written criteria, which were unchanged from those used in the first survey.

Table 5 Factors associated with non-Alfa ratings (survey number two)

Factor	Rating				
	Total	A	B	C	D
Margin or open margin	140	–	126	13	1
Caries	20	–	–	–	20
Fractured margin	11	–	7	2	2
Inadequate extension	17	–	9	8	–

Table 5b Number of times factor mentioned, posterior criteria study (survey number two)

Factor	Total	Without other factors	
		Number	Percent
Margin or open margin	284	140	49.3
Fractured margin	67	11	16.4
Inadequate extension	48	17	35.4
Gross or marginal ridge fracture	34	2	*
Caries	34	20	58.8
Probable recurrent decay	26	1	*
Restoration discolored	25	2	*
Poor contour	23		*
Restoration tarnished	13	4	*
Surface roughness	10	6	*
Overhang	5		*
Restoration mobile	4		*
Restoration pitted	4	1	*
Restoration corroded	3	1	*
Restoration missing	1		*
Poor contact	1		*

Percent not computed for factors mentioned for a total of less than 25 or for factors mentioned without other factors less than five times

Table 5a Factor matrix: posterior criteria study (survey number two)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Margin or open margin	1	140	47	20	10	8	14	20	16	5		1		2			1
Fractured margin	2		11	1	4				2	1		1					
Inadequate extension	3			17	3	4	1		1					1			
Gross or marginal ridge fracture	4				2	1	7	1	1				4			1	
Caries	5					20		1									
Probable recurrent decay	6						1		1						2		
Restoration discolored	7							2									
Poor contour	8											2					
Restoration tarnished	9									4	3						
Surface roughness	10										6						
Overhang	11												1				
Restoration mobile	12																
Restoration pitted	13													1			
Restoration corroded	14														1		
Restoration missing	15																
Poor contact	16																

The notes which had been prepared after the first and second training sessions were rescinded; they had served the purpose of resolving controversy among the examiners during training, but logical analysis revealed that they contributed little to specific definition of the factors to be rated. Additional experience made it easier for the examiners to accept the criteria as written, without serious disagreements over minor diagnostic points.

It was felt that the opportunity to note caries and discoloration would help the examiners to let their judgments be guided by the written criteria.

Results of survey number three

Prior to conducting the final field survey, a demonstration was conducted by the instructor, using extracted teeth with amalgam restorations, and silver plated models of posterior teeth containing restorations. A full range of conditions were represented for each characteristic. Following the demonstration with the extracted teeth and models, one patient with approximately ten posterior restorations was examined by the instructor and each examiner. The ratings were not recorded but the reasons for assigning ratings were reviewed by the examiners.

Survey Number Three was conducted at the Recruit Training Center, United States Coast Guard Base, Alameda, California. The examinees were recruits who had recently reported for duty. The five dentists who had participated in the previous field tests served as examiners, and recorders were obtained from the survey group. Randomly selected posterior quadrants were examined for each patient, yielding a total of 185 restorations which were rated twice by each examiner during trials held two days apart. Over 97 percent of the restorations were dental amalgams.

The distribution of 1850 ratings for marginal adaptation, obtained in the two examination sessions, is shown in Table 6. Consensus ratings are given in Table 7.

Marginal Adaptation ratings were distributed in every rating category, but the distribution, shown in Table 8, indicated that not all examiners were rating restorations identically. Examiner Z (one of the least experienced) appears to have been the most critical in assessing margins, especially in the "Alfa-Bravo" zone.

Although the examiners differed among themselves in rating marginal adaptation, they were able to duplicate their own ratings from the first to the second examination fairly well, as shown in Table 9. Because of inter-examiner disagreements, however, further training in using the rating scales for this characteristic was planned.

Table 10 indicates that most of the disagreements with consensus were of an "Alfa-Bravo" nature; this was not surprising, since most of the margins in this patient group had been rated as "Alfa" or "Bravo." As shown in Table 11, intra-examiner disagreements followed the same pattern as the disagreements with consensus.

The percent inter-examiner agreement and intra-examiner agreement in judging marginal adaptation on Survey Number Three is given in Table 12 Inter-examiner agreement ranged from 57.2 to 85.4 percent, and intra-examiner agreement ranged from 79.4 to 88.6 percent.

Graph 1 summarizes self-agreement for all examiners in the third survey. For example, examiner A agreed with his own judgments 82.7 percent of the time. Since this percentage can be expected to vary randomly from trial to trial, it is useful to determine what the range of variation is likely to be. The graph indicates that at the 95 percent

Table 6 All ratings, marginal adaptation (number and percent, survey number three)

	Total	Oscar	Alfa	Bravo	Charlie	Delta
Number	1,850	–	632	1,181	18	19
Percent	100.0	–	34.2	63.8	1.0	1.0

Table 7 Consensus ratings, marginal adaptation (number and percent; survey number three)

	Total	Oscar	Alfa	Bravo	Charlie	Delta
Number	185	–	61	121	1	2
Percent	100.0	–	33.0	65.4	0.5	1.1

Table 8 Marginal adaptation ratings by examiner (survey number three)

Examiner	Examination	Total	Rating (Percent)				
			0	A	B	C	D
A	1	100.0	–	44.3	53.5	2.2	–
	2	100.0	–	41.1	55.6	2.2	1.1
B	1	100.0	–	58.9	40.0	–	1.1
	2	100.0	–	48.6	49.3	0.5	1.6
C	1	100.0	–	28.1	70.3	0.5	1.1
	2	100.0	–	25.9	73.1	0.5	0.5
Y	1	100.0	–	26.5	70.2	2.2	1.1
	2	100.0	–	35.1	62.2	1.1	1.6
Z	1	100.0	–	18.4	80.0	0.5	1.1
	2	100.0	–	14.6	84.3	–	1.1

Table 9 Duplicate ratings by examiner, marginal adaptation (survey number three)

Examiner	Total	Duplicate ratings	
		Number	Percent
A	185	153	82.7
B	185	147	79.4
C	185	164	88.6
Y	185	152	82.1
Z	185	159	85.9

Table 10 Agreement and disagreement with consensus ratings by examiner, marginal adaptation (survey number three)

Examiner	Examination	Grand Total	Total Agreement	Total Disagreement	A–B	A–C	A–D	B–C	B–D	C–D
A	1	185	152	33	29	–	–	2	1	1
	2	185	150	35	31	–	–	2	1	1
B	1	185	132	53	52	–	–	1	–	–
	2	185	142	43	40	1	–	1	1	–
C	1	185	164	24	17	1	1	2	2	1
	2	185	166	19	16	–	–	2	1	–
Y	1	185	150	34	32	–	–	–	2	–
	2	185	161	24	22	–	–	1	1	–
Z	1	185	154	31	29	–	–	–	2	–
	2	185	150	35	34	–	–	1	–	–
Total	1	925	752	173	159	1	1	5	7	2
	2	925	769	156	143	1	–	7	4	1

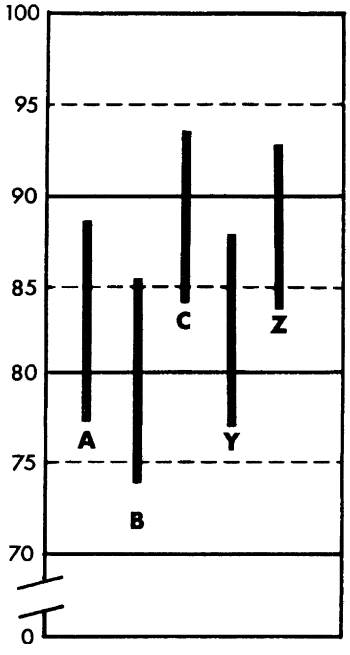
Table 11 Inter-examiner agreement and disagreement by examiner, marginal adaptation (survey number three)

Examiner	Grand total	Total agreement	Total disagreement	Nature of disagreement					
				A–B	A–C	A–D	B–C	B–D	C–D
A	185	153	32	28	–	–	2	2	–
B	185	147	38	36	1	–	–	1	–
C	185	164	21	18	–	–	2	1	–
Y	185	152	33	30	–	–	2	1	–
Z	185	163	22	19	–	–	1	2	–
Total	925	779	146	131	1	–	7	7	–

Table 12 Percent inter- and intra-examiner agreement, marginal adaptation (survey number three)

	Examiner	Second examination				
		A	B	C	Y	Z
First Exam	A	(82.7)	72.4	76.2	75.1	69.7
	B	70.8	(79.4)	70.2	75.6	63.7
	C	78.9	65.9	(88.6)	85.4	75.6
	Y	70.8	63.2	78.3	(82.1)	83.7
	Z	70.8	57.2	81.6	76.2	(88.1)

level of confidence self-agreement under Marginal Adaptation for examiner A was between 77.5 percent and 87.5 percent. It was encouraging to note that in no case did the lower limit for any examiner fall below 74 percent.



Graph 1 95 percent confidence intervals for self-agreement by examiner (survey number three)

Appendix II

Survey Results

Table 1 All ratings (survey number three), anterior criteria study, all characteristics

Characteristic	Ratings					
	Total	0	A	B	C	D
Color match	1,448	368	385	672	23	*
Cavosurface marginal discoloration	1,448	*	717	642	89	*
Dark deep discoloration	1,448	*	1139	309	*	*
Contour	1,448	*	1177	239	32	*

Table 2 Consensus ratings (survey number three), anterior criteria study, all characteristics

Characteristic	Ratings					
	Total	0	A	B	C	D
Color match	181	47	50	82	2	*
Cavo-surface marginal discoloration	181	*	95	77	9	*
Dark deep discoloration	181	*	148	33	*	*
Contour	181	*	149	28	4	*
Marginal integrity	181	*	87	86	6	2

*Not applicable

Table 3a Color match ratings by examiner, anterior criteria study (survey number three)

Examiner	Examination	Ratings				
		Total	0	A	B	C
D	1	181	45	38	95	3
	2	181	46	44	90	1
E	1	181	40	44	95	2
	2	181	38	46	95	2
A	1	181	46	59	73	3
	2	181	47	42	90	2
B	1	181	57	54	64	6
	2	181	49	58	70	4

Table 3b Cavosurface marginal discoloration ratings by examiner, anterior criteria study (survey number three)

Examiner	Examination	Ratings			
		Total	A	B	C
D	1	181	118	49	14
	2	181	133	41	7
E	1	181	73	96	12
	2	181	69	100	12
A	1	181	81	80	20
	2	181	59	114	8
B	1	181	96	74	11
	2	181	88	88	5

Table 3c Dark deep discoloration ratings by examiner, anterior criteria study (survey number three)

Examiner	Examination	Ratings		
		Total	A	B
D	1	181	153	28
	2	181	153	28
E	1	181	149	32
	2	181	137	44
A	1	181	141	40
	2	181	139	42
B	1	181	131	50
	2	181	136	45

Table 3d Contour ratings by examiner, anterior criteria study (survey number three)

Examiner	Examination	Ratings			
		Total	A	B	C
D	1	181	156	22	3
	2	181	144	32	5
E	1	181	142	31	8
	2	181	143	29	9
A	1	181	145	34	2
	2	181	133	44	4
B	1	181	153	27	1
	2	181	161	20	–

Table 3e Marginal integrity ratings by examiner, anterior criteria study (survey number three)

Examiner	Examination	Ratings				
		Total	A	B	C	D
D	1	181	87	84	7	3
	2	181	79	93	6	3
E	1	181	103	63	13	2
	2	181	87	81	10	3
A	1	181	89	85	5	2
	2	181	64	110	6	1
B	1	181	92	85	3	1
	2	181	70	107	–	4

Table 4a Percent inter- and intra-examiner agreement, anterior criteria study, color match (survey number three)

	Examiner	Second examination			
		D	E	A	B
First Examination	D	(85.1)	76.8	78.5	74.6
	E	78.5	(78.5)	74.0	68.0
	A	76.8	75.1	(76.8)	72.4
	B	72.9	69.6	69.6	(72.9)

Table 4b Percent inter- and intra-examiner agreement, anterior criteria study, cavo-surface marginal discoloration (survey number three)

	Examiner	Second examination			
		D	E	A	B
First examination	D	(83.4)	58.6	53.6	68.0
	E	64.1	(72.4)	69.1	72.4
	A	65.7	67.4	(65.2)	71.3
	B	70.2	68.0	68.5	(82.3)

Table 4c Percent inter- and intra-examiner agreement, anterior criteria study, dark deep discoloration (survey number three)

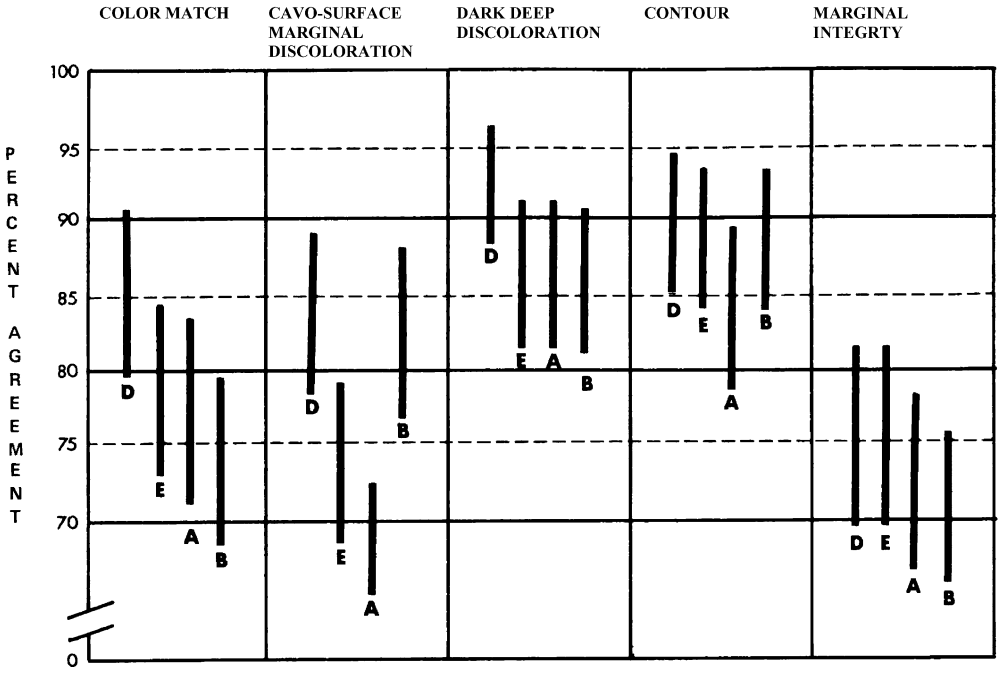
	Examiner	Second examination			
		D	E	A	B
First examination	D	(92.3)	88.9	86.7	87.3
	E	88.9	(86.7)	83.4	89.5
	A	88.9	84.5	86.7	81.8
	B	80.1	79.0	81.2	(86.2)

Table 4d Percent inter- and intra-examiner agreement, anterior criteria study, contour (survey number three)

	Examiner	Second examination			
		D	E	A	B
First examination	D	(90.1)	86.7	80.1	84.5
	E	84.5	(88.4)	81.8	82.3
	A	85.6	83.4	(84.0)	78.5
	B	87.8	86.2	85.6	(88.4)

Table 4e Percent inter- and intra-examiner agreement, anterior criteria study, marginal integrity (survey number three)

	Examiner	Second examination			
		D	E	A	B
First examination	D	(75.7)	74.6	75.1	66.8
	E	70.7	(75.7)	66.3	63.0
	A	70.1	66.3	(71.8)	68.0
	B	70.7	66.8	66.8	(70.2)



Graph 1 95 percent confidence intervals for self-agreement by examiner, anterior criteria study (survey number three)

Table 6 All ratings (survey number three), posterior criteria study, all characteristics

Characteristic	Ratings					
	Total	0	A	B	C	D
Surface texture	1,850	*	380	1,404	66	*
Anatomic form	1,850	*	1,780	68	2	*
Marginal adaptation	1,850	–	632	1,181	18	19

*Not applicable

Table 7 Consensus ratings (survey number three), posterior criteria study, all characteristics

Characteristic	Ratings					
	Total	0	A	B	C	D
Surface texture	185	*	53	129	3	*
Anatomic form	185	*	184	1	–	*
Marginal adaptation	185	–	61	121	1	2

*Not applicable

Table 8a Surface texture ratings by examiner, posterior criteria study (survey number three)

Examiner	Examination	Ratings			
		Total	A	B	C
A	1	185	47	132	6
	2	185	17	166	2
B	1	185	47	132	6
	2	185	27	153	5
C	1	185	62	122	1
	2	185	45	136	4
Y	1	185	37	134	14
	2	185	30	143	12
Z	1	185	44	131	10
	2	185	24	155	6

Table 8b Anatomic form ratings by examiner, posterior criteria study (survey number three)

Examiner	Examination	Ratings			
		Total	A	B	C
A	1	185	183	2	–
	2	185	183	2	–
B	1	185	183	1	1
	2	185	184	–	1
C	1	185	172	13	–
	2	185	178	7	–
Y	1	185	182	3	–
	2	185	185	–	–
Z	1	185	155	30	–
	2	185	175	10	–

Table 8c Marginal adaptation ratings by examiner, posterior criteria study (survey number three)

Examiner	Examination	Ratings					
		Total	0	A	B	C	D
A	1	185	–	82	99	4	–
	2	185	–	76	103	4	2
B	1	185	–	109	74	–	2
	2	185	–	90	91	1	3
C	1	185	–	52	130	1	2
	2	185	–	48	135	1	1
Y	1	185	–	49	130	4	2
	2	185	–	65	115	2	3
Z	1	185	–	34	148	1	2
	2	185	–	27	156	–	2

Table 9a Percent inter- and intra-examiner agreement, posterior criteria study, surface texture (survey number three)

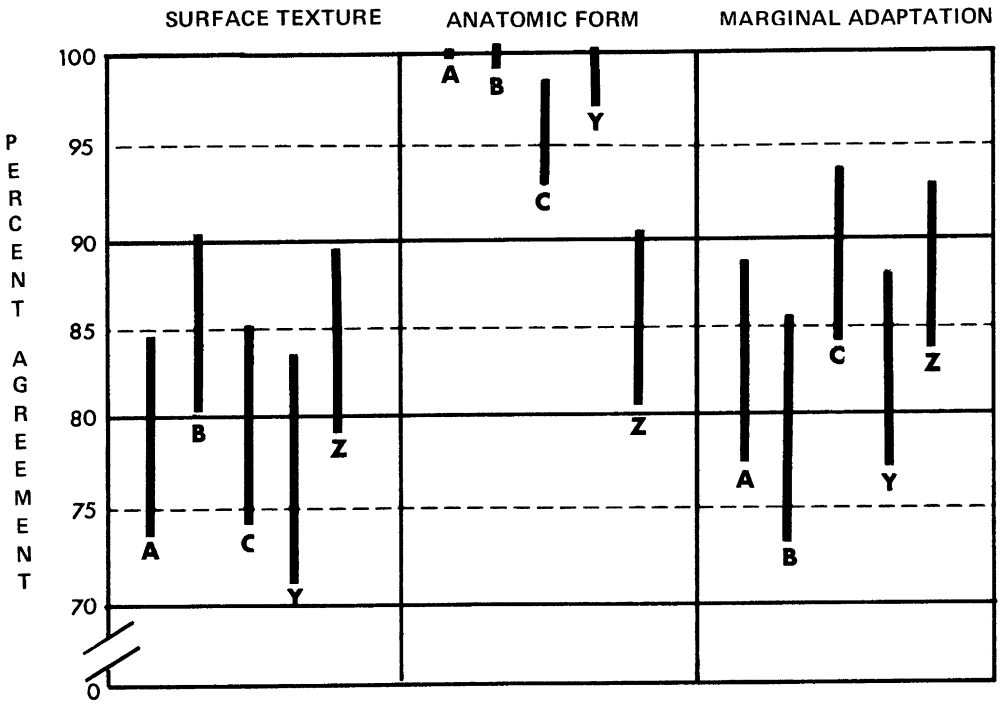
	Examiner	Second examination				
		A	B	C	Y	Z
First examination	A	(78.9)	82.7	80.0	82.7	85.4
	B	75.1	(85.4)	82.7	84.3	87.0
	C	74.5	80.0	(79.4)	80.5	83.2
	Y	70.2	78.3	76.7	(77.2)	76.7
	Z	75.1	76.7	75.1	75.1	(84.3)

Table 9b Percent inter- and intra-examiner agreement, posterior criteria study, anatomic form (survey number three)

	Examiner	Second examination				
		A	B	C	Y	Z
First examination	A	(100.0)	98.3	95.1	98.9	94.5
	B	97.8	(99.4)	96.2	99.4	94.0
	C	92.9	92.4	(95.6)	92.9	94.5
	Y	98.3	97.2	82.1	(98.3)	96.2
	Z	84.8	83.7	83.2	92.4	(85.9)

Table 9c Percent inter- and intrae-xaminer agreement, posterior criteria study, marginal adaptation (survey number three)

	Examiner	Second examination				
		A	B	C	Y	Z
First examination	A	(82.7)	72.4	76.2	75.1	69.7
	B	70.8	79.4	70.2	75.6	63.7
	C	78.9	65.9	(88.6)	85.4	75.6
	Y	70.8	63.2	78.3	(82.1)	83.7
	Z	70.8	57.2	81.6	76.2	(88.1)



Graph 2 95 percent confidence intervals for self-agreement by examiner (survey number three)

Appendix III

Statistical Methods

1. Sequential Analysis

Sequential analysis is useful for testing examiner performance, since each comparison can be dichotomized as either "agree" or "disagree." The advantage of sequential trials is that there is no fixed number of cases – the experiment can be ended when a decision is reached, thus eliminating unnecessary work. In addition, Type I error, α, and Type II error, β, are both specified in advance, in contrast to the usual arrangement where Type I error is fixed, and Type II error must be calculated. An excellent discussion of sequential analysis can be found in Chilton [14]. For the studies reported in this volume, α and β were both set at. 0.10, and the region of no decision was between 0.75 and 0.85. The sequential worksheet used in these studies appears at the end of the appendix.

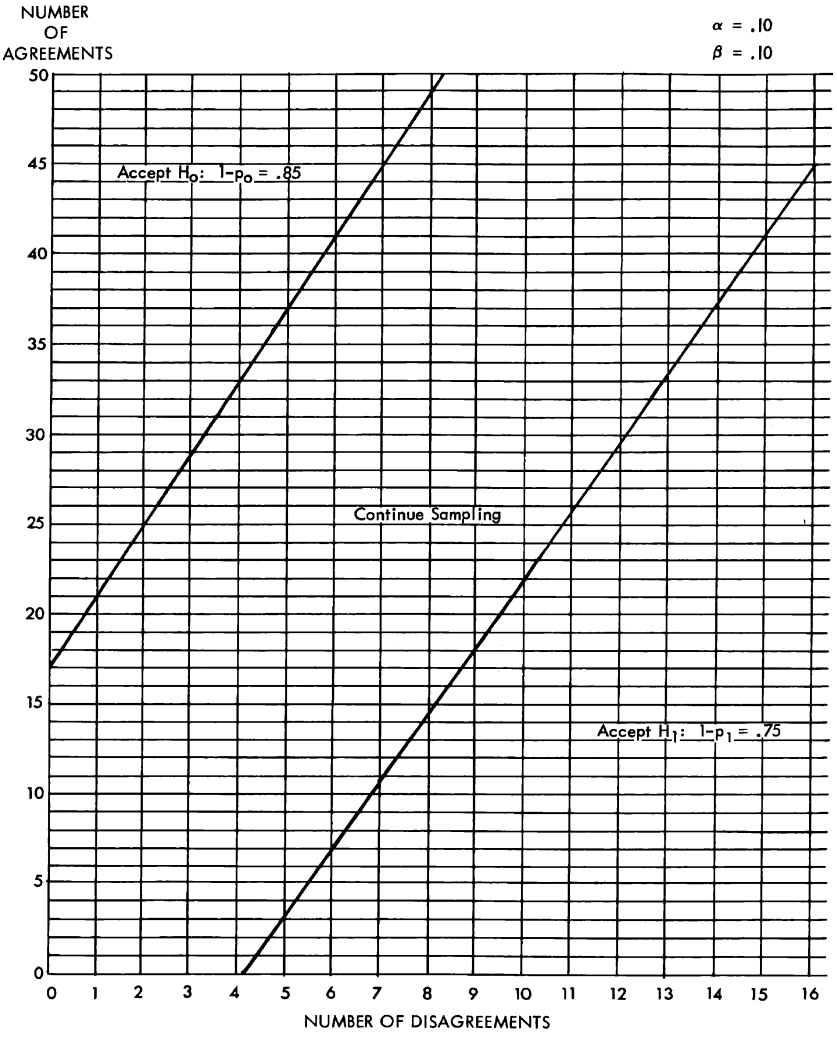
2. Data Analysis for Clinical Studies

In a completely randomized experiment where test and control groups of teeth constitute independent samples, and the outcome consists of graded results, a modified Wilcoxon 2-sample Test [15] is an excellent way of testing for statistical differences. The advantage of this test is that it makes use of data having an ordinal arrangement, while chi-square does not. However, when test and control teeth appear in the same mouth, with one randomly designated as test while the other is control, a test of significance for related samples must be used. The sign test is appropriate, providing that there are not too many tied pairs. An alternative is to assign numeric values to the letter ratings, and to apply a test to the distribution of differences between pairs. The choice of assigned values makes little difference in the outcome of this statistic.

An experiment using matched pairs has the advantage of controlling for environmental and biological factors, while independent samples depend upon randomization for control. Kish [16] presents a good discussion of the relation of statistical tests to variables that are controlled, uncontrolled, or randomized.

3. Methodological Note

Since the teeth of any given patient are not independent of each other, the patient is the unit of analysis. This presents no problem when each patient in a study has either one test or one control restoration (the case of independent samples), or when each patient has one pair of test-control restorations (the case of related samples). However, when these numbers are exceeded, some method must be devised to represent each patient by a single score before applying a test of significance.



Graph 3 Sequential analysis worksheet

Appendix IV

References

1. MORGAN, S. C.: Criteria for the Clinical Evaluation of Dental Anterior Restorative Materials. Residency Report, Dental Health Center, San Francisco, California, 1966.
2. PARKER, W. A.: A Study of Criteria for Clinical Evaluation of Posterior Restorative Materials. Residency Report, Dental Health Center, San Francisco, California, 1967.
3. HOROWITZ, H. S. and PETERSON, J. K.: Evaluation of Examiner Variability and the Use of Radiographs in Determining the Efficacy of Community Fluoridation, *Archs Oral Biol*., II: 867–875,1966.
4. MARKEN, K.-E.: The Training of Observers, pp., 81–87. In *Advances in Fluorine Research and Dental Caries Prevention*, Vol. 4, Proceeding of the 12th Congress of the European Organization for Research on Fluorine and Dental Caries Prevention, Utrecht, The Netherlands, 8–11 June, 1965, Oxford, Pergamon Press, 1966.
5. GUILFORD, J. P.: *Psychometric Methods*, New York, McGraw-Hill Book Company, Inc., p. 292, 1954.
6. PHILLIPS, R. W., BOYD, D. A., HEALEY, N. J., and CRAWFORD, W. H.: Clinical Observations on Amalgams with Known Physical Properties, *J. Dent. Res*., 22: 167–172, 1943.
7. WILSON, C. J. and RYGE, G.: Clinical Study of Dental Amalgam, *J. Amer. Dent*. Ass., 66: 673–771, 1963.
8. PAFFENBARGER, G. C., SCHOONOVER, I. C., and SOUDER, W.: Dental Silicate Cements: Physical and Chemical Properties and a Specification, *J. Amer. Dent. Ass*., Vol. 25, January 1938.
9. CIVJAN, S. and BRAUER, G. M.: Physical Properties of Cements Based on Zinc Oxide, Hydrogenated Rosin, o-Ethoxybenzoic Acid, and Eugenol, *J. Dent. Res*., 43: 281–299, 1964.
10. CIVJAN, S. and BRAUER, G. M.: Clinical Behavior of o-Ethoxybenzoid Acid - Eugenol - Zinc Oxide Cements, *J. Dent. Res*. 44: 80–83, 1965.
11. MACRAE, P. D., ZACHERL, W., and CASTALDI, C. R.: A Study of Defects in Class II Dental Amalgam Restorations in Deciduous Molars, *J. Canad. Dent. Ass*., 28: 491-502, 1962.
12. GOING, R. E., MASSLER, J., and DUTE, H. L.: Marginal Penetration of Dental Restorations by Different Radioactive Isotopes, *J. Dent. Res*., 39: 273–284, 1960.
13. RYGE, G.: Dental Materials, pp., 253-267. In Goldman, H. M., Forrest, S. P., Byrd, D. L., and McDonald, R. E. (eds.): *Current Therapy in Dentistry*, Vol. 3. Saint Louis, The C. V. Mosby Company, 1968.
14. CHILTON, N. W.: *Design and Analysis in Dental and Oral Research*. J. B. Lippincott, 1967.
15. ARMITAGE, P.: Tests for Linear Trends in Proportions and Frequencies, Biometrics, 11: 375–386, Sept. 1955.
16. KISH, L.: Some Statistical Problems in Research Design, Amer. Soc. Review, 24: 328–388, June, 1959.