Validity of the M-L scale and Vivadent scale in clinical wear study

Shu-cheng Chang, DDS, MS,[a] Shiro Suzuki, DDS, PhD,[b] and Yu-Hwa Pan, DDS, MS[c]

[a]Mackay Memorial Hospital, Taipei, Taiwan, [b]Clinical Research and Department of Prosthodontics, University of Alabama at Birmingham, Birmingham, USA, and [c]Dental Department, Chang-Gung Memorial Hospital, Keelung, Taiwan

Purpose: The purpose of this study was to evaluate the validity of two commonly used visual inspection methods, the M-L scale and the Vivadent scale, in a clinical wear study.
Materials and Methods: Sixty study casts were evaluated using both the Micro-Vu metrology system and visual inspection methods. Values obtained by using the visual inspection methods were correlated to those measured by the Micro-Vu metrology system. The same principles were applied to prepare specimens to test the precision of the proposed technique.
Results: This new impression technique in conjunction the Micro-Vu system was proved to be precise. The mean of the differences between repeat measurements was 1.26 μm with a standard deviation of 0.8 μm. When comparing the validity of the two scales, the M-L scale was found to be better than the Vivadent scale. Values obtained by using visual inspection methods were larger than values obtained by using the Micro-Vu metrology system. Overall values obtained by using the M-L scale were smaller than those obtained by the Vivadent scale. The effect of evaluators and the interactions between influencing factors were prominent as to significantly reverse the effect of some evaluators.
Conclusion: Based upon the results of this study, it was concluded that the validity of the currently used visual inspection methods was considerably influenced by many factors, the desirability of continuing these types of measurements for definitive wear assessment should be reconsidered. **(Int Chin J Dent 2005; 5: 22-32.)**

Key Words: Baldwin illusion, mental average procedure, M-L scale, validity, Vivadent scale.

Introduction

Occlusal wear has been a major clinical concern for all restorative materials that are used for restorations of posterior teeth. Assessment of the amount of occlusal loss is an important issue for any new resin composite restorative material. According to the original guidelines for the acceptance of a posterior composite by the American Dental Association in 1984,[1] the wear loss should be less than 250 μm after 5 years of clinical service. These guidelines were revised in 1989, and the posterior composites were classified into two categories.[2] The Category-A composites are for unrestricted use in posterior teeth, while Category-B composites are restricted for small, conservative cavity preparations in which the occlusal stress is limited. To be qualified as a Category-A composite, after an initial 6-month period, no more than 50 μm wear should occur during the following 2 years, and less than 100 μm wear during a 4-year period. For a Category-B composite, the corresponding amounts of wear should be less than 125 μm and 175 μm, respectively. Wear requirements were based on the average wear of individual restorations. The revised guidelines published in 1991 and 1996 excluded the possibility of a composite being accepted for restricted use.[3,4]

Measurement of clinical wear can be divided into two methodologies, direct and indirect measurements. A direct evaluation method applies to wear evaluation techniques that employ direct visual observation by evaluators. The indirect methods allow evaluators to evaluate information recorded in the patient's mouth. Indirect methods can be further divided into two categories, non-instrumental and instrumental methods. From the literature, many wear evaluation methods have been proposed and are summarized in Table 1.[5-34] These methods differ widely in their ability to generate quantitative values, their availability, the cost of equipment, and the time and effort required for evaluations. As the instrumental methods are often costly, non-instrumental

methods are more feasible in large clinical studies.[33] ADA guidelines in 1989 restricted the evaluation of wear to indirect methods only and required that the precision and accuracy of the method used should be stated.[2] Until the present time, it remains unclear which indirect non-instrumental method provides the best precision and validity. Validity is the extent to which the study measures what it is intended to measure; precision is the reproducibility of the study results.[35] Most indirect non-instrumental methods for clinical wear compare the distance between the worn restorative surface and its original margin with a set of standards. In order to determine the validity of non-instrumental methods, reliable instrumental methods must be explored that measure the same distance. Some studies had been performed to compare the precision of different scales.[15,36,37] However, due to the difficulties of measuring this distance as well as the multi-factorial character of the indirect non-instrumental methods themselves, only a few of them focused on validity.[17] None of these studies had been designed to compare the validity of different scales.

Table 1. Direct and indirect methods in clinical wear measurements.

Category/ Methods	Technique	Reference
Direct		
USPHS	Visual comparison	[5]
Clinical ranking	Visual comparison	[6]
Modified USPHS	Visual comparison	[7]-[9]
Indirect/ non-instrumental		
Photographic method	Photograph comparison	[6], [10]
Total ranking model	Model comparison	[11], [12]
Categorial scoring	Model comparison	[11]
Leinfelder's method (I)	Model compared to scale	[12]
SEM photograph rating	SEM photo comparison	[13]
M-L scale	Model compared to scale	[14]
Vivaden scale	Model compared to scale	[15]
Step wedge meothod	Model compared to scale	[16]
Sabri/Boghosian scale	Model compared to scale	[17]
Indirect/ instrumental		
Handelman's method	Coping impression for volume	[18]
Dennison's method	Coping impression for volume	[19]
Urquiola's method	Coping impression for volume	[20]
Vrijhoef's method	Coping impression for volume	[21]
Santucci's method	Topographic micrometer for depth	[22]
Leinfelder's method (II)	Traveling microscope	[12]
Profilometry	Profiling for depth	[23]-[25]
Roulet's method	Computer 3D occlusal mapping	[26]
Eick's method	Stereo-camera with reference system	[27]
Williams' method	Laser occlusal mapping	[28]
Lambrechts' method	3D measuring microscope	[29]
DeLong's method	Computer graphics and servohydraulics	[30]
Michigan system	Computer-graphic measuring system	[31]
Reflex microscope	Optical measuring system	[32]
Impression sectioning	SEM photograph for depth	[25]
SEM photogrammetry	SEM photography with computer	[25]
Chadwick's method	Stereomicroscope with camera	[8]
Laser profilometry	Laser profiling for depth	[33]
Mehl's method	Optical 3D device	[34]

Reviewing the literature involving the use of instrumental evaluation for wear of composite resins revealed that most of the studies focused on the volume of material loss as the main factor to judge wear.[18-21,26-32,34] However, Chadwick et al.[8] considered that it may be misleading to use this approach because a large restoration with a minimal depth of material wear loss may have an apparent volume loss that is greater than a small restoration that shows a considerable depth of wear loss. They suggested that calculating the depth of material loss might be a fairer way of comparing the clinical performance of materials. Few techniques have been advocated to measure the depth of wear loss.[8,12,22-25,33] Among them, only Chadwick et al.,[8] Leinfelder et al.,[12] and Bayne et al.[33] measured the depth of wear in the clinical situation.

Factors that might influence the result of the measurement when an indirect method is applied include: different scales,[15,36,37] different evaluators,[15,38] use of a mental average procedure,[15,16,33] the Baldwin illusion,[17,39] and degree of experience gained with repeat measurements.[38]

The objectives of this study were: a) to introduce a new instrumental method to measure the amount of wear loss in a clinical study, and b) to investigate the validity of two commonly used indirect wear evaluation methods, the M-L scale and the Vivadent scale. It was expected that this study would provide a better understanding of the relationship between these two methods, and the underlining factors that might influence their validity.

Materials and Methods

Sixty study casts, including upper or lower molars, which were representative of varying degrees of occlusal wear, were chosen from several long-term wear studies previously conducted in the Research Clinic of the University of Alabama, School of Dentistry. To reduce the variables and preserve the original casts for future studies, all casts were duplicated with a silicone- based impression material (Express, 3M Co., St. Paul, MN, USA) and poured with dental stone (Silky Rock, Whip-Mix Corp., Louisville, KY, USA). The reason for choosing this dental stone is its low expansion value (0.09%). The buccal cusps of all sixty experimental casts were sorted by their height. The lowest 20 casts were coded as “L”, low cusp height; the medium 20 casts were coded as “M”, medium cusp height; and the highest 20 casts were coded as “H”, high cusp.

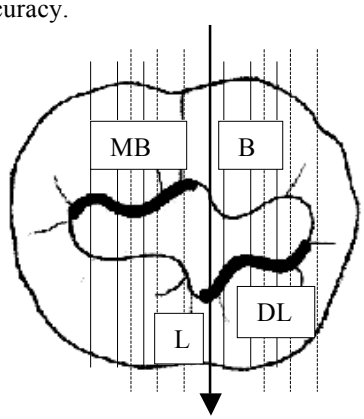
The definition of wear in this study is the mean distance from the cavosurface margin to the occlusal surface of the composite resin, as described by Leinfelder et al.[12] This kind of wear of the specimens is a contact-free area wear[24] and caused almost entirely by three-body abrasion by abrasive particles in the masticatory slurry.[16] To fulfill the objectives of this study, the selected experimental casts were evaluated using two methodologies, instrumental and non-instrumental evaluations.

For the instrumental evaluation, the Micro-Vu optical metrology system (Model Qubix 8x6, Micro-Vu Corporation, CA, USA) was used to measure wear on the negative replicas of sixty casts (Fig. 1). The Micro-Vu metrology system is a three dimensional non-contacting measuring tool. It consists of three main parts: the measuring stage, the InSpec Vision System and the InSpec Metrology Software. The motorized stage is an electro-mechanical assembly that sends accurate x, y, and z coordinates information to the measuring software. The InSpec Vision system has the capability of quantitative measurements of the desired features on a specimen. By taking the positional information from the measuring stage, the InSpec metrology software is able to determine the size and location of the features of interest and execute appropriate functions as needed. For the x and y axes, the resolution is 0.5 μm and repeatability is 1.0 μm.

For each experimental cast, three impressions were taken with a silicone impression material (Express).

Sectioning of the impression was accomplished with a sharp #12 surgical blade. All sections were in a bucco-lingual direction. After one arbitrary cut was made on the first replica, the sectioned replica was used to mark the designated points in the buccal and lingual margins on the stone model prior to further sectioning. For the other two negative replicas, three equally spaced sections were made on both the mesio-buccal area and the disto-lingual area of the margins (Fig. 2). This sectioning of multiple impressions is advantageous as it produces thick sections and minimizes the risk of distortion and loss of accuracy.

1



2

Fig. 1. Micro-Vu optical metrology system.
Fig. 2. Scheme for evaluation. Two points (designated as points of the buccal margin, B, and the lingual margin, L) are at the intersections of the arrow and margins. Two areas (mesio-buccal area, MB, and disto-lingual area, DL) are bolded. Arrow denotes the position of arbitrary section on the first replica. Solid lines are sections of the second negative replica and dash lines are sections of the third one.

All sections of negative replicas for a tooth were then placed on a glass slide and measured by using the Micro-Vu system. The profiles of wear facets were magnified 35 to 365 times as necessary to create an image of a sufficient size on the monitor in order to identify the location of the cavosurface margin and the worn restorative surface. The distance between these two locations was calculated by the computer software. The data obtained by examining the sections through designated buccal and lingual points for each cast were coded as "point" data. Four measurements for each area were averaged then coded as "area" data.

As a precision test for the measurement technique, a similar procedure with exactly the same materials for clinical casts was performed on 18 cylinders with known depths of prepared depressions ranging from 0 to 1,000 μm. Impressions were taken and negative replicas were sectioned as described for wear specimens. Three parallel sections with one at the diameter of circular depressions were made for each negative replica. Sections facing the diameter of the replicas were placed on glass slides and measured by the Micro-Vu metrology system. For each section, two depths of depression were measured. Seventy-two measurements were obtained. Repeat measurements of all impressions were performed at a different time period.

For the non-instrumental evaluation two methods, the M-L scale and the Vivadent scale, were investigated in this study. Moffa and Lugassy first developed the M-L scale in 1986.[14] The M-L standards are 18 cylinders with flat-bottomed circular depressions of known depth centered on the flat end surfaces, representing incremental defects at intervals ranging from 0 to 1,000 μm. The difference between each scale in the series is 25 μm when the wear is less than 100 μm, 50 μm when the wear is between 100 μm to 500 μm, and 100 μm when the wear is more than 500 μm. The Vivadent scale is an 18-category standard scale that consists of tooth-shaped, tooth-sized dies with restoration-like incremental defects ranging from 25 μm to 1,000 μm. The extent of the artificial material loss at the margins is constant. The difference between each standard in the series is 25

μm when the wear is less than 200 μm, 50 μm when the wear is between 200 μm to 400 μm, and 100 μm when the wear is more than 400 μm. Both the incremental depth pattern of the M-L scale and that of the Vivadent scale reflect the capacity of evaluators to visually detect smaller differences when the actual amount of existing loss is less.[15]

Five dentists denoted C, L, N, P, and S were trained to use both scales following a training program. Among them, only evaluator S had previous experience using the M-L scale for evaluating wear on clinical casts. The standards for both scales were duplicated and labeled randomly. For the M-L scale, evaluators had to evaluate thirty-six stone replicas and achieve 90% accuracy prior to proceeding. Evaluators then evaluated ten additional clinical casts with known amounts of wear and were required to get 80% accuracy before they proceeded to evaluate the experimental casts. For the Vivadent scale, a similar training program was followed with the only difference being that replicas of the Vivadent scale were used.

Each evaluator was instructed to conduct evaluations in a standardized manner and received a protocol clearly describing the procedures and necessary information. A high intensity incandescent lamp was positioned in front of the evaluator and the experimental casts and arranged to emphasize shadows cast by exposed walls of the cavity preparation and to aid in evaluation of the degree of wear. No amplifier was used.

For each experimental cast, two points on the buccal and lingual margins and two areas (mesio-buccal and disto-lingual area) were evaluated (Fig. 2). A red mark was placed near the point being evaluated, and comparisons could be made with standard scales and coded as such. For those two areas, a "mental average procedure" was utilized.

After taking the training program for the M-L scale, evaluators examined all experimental casts with that scale. At least a 24-hour lapse was required before evaluating experimental casts with the other scale to prevent the possibility that evaluators might remember the values from the previous evaluation. The same rules applied for evaluating casts using the Vivadent scale. Repeat measurements were done after a 1-week lapse. Before a repeat measurement was made, a new training program was performed. After finishing their evaluation, evaluators were asked three subjective assessment questions regarding which scale was easier to learn, faster to use, and their confidence that they had used it correctly.

The experience factor was determined by the improvement of validity of repeat measurements with the same scale. The effect of the mental average procedure was tested by the difference between the evaluations of points and areas. The mental average procedure was employed only when evaluating areas. Therefore, the difference between the results obtained from the area and point evaluations indicated the effect of the mental average procedure. The effect of "Baldwin illusion" was judged by the influence of cusp height.

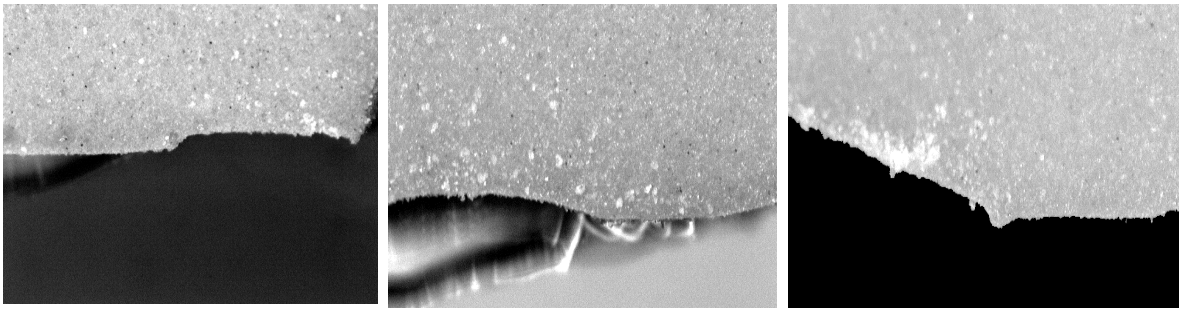
Correlation and regression analyses were used for the precision test of the impression technique with the Micro-Vu system. A combination of analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) were used to analyze the validity. The confidence level for hypothesis tests was set at 0.05. Scheffe's test was applied when the F-test was significant. The dependent variable tested was the difference between square root of values obtained by Micro-Vu system and the square root of values obtained by evaluators using two different scales. This variable was tested to show approximately a normal distribution. "Validity" is represented by the mean of this variable. The closer to zero this mean is, the better is the validity. A negative value of mean indicates that the values obtained by using non-instrumental scales were larger than that obtained by the Micro-Vu system.

Results

Instrumental evaluation

The correlation between the results obtained by the Micro-Vu system and the depths of cylinders was very high. R^2 for the first and repeat measurements were 0.9966 and 0.9970, respectively. The mean of the difference between the first measurement of replicas and second measurement was 1.26 μm with a standard deviation of 0.8 μm.

Three different types of profiles of impression replicas were observed with the Micro-Vu metrology system (Fig. 3): flat floors of worn restorations, curved floors of worn restorations and marginal degradation. When a curved floor of composite was observed, higher magnification was used to define the locations of cavosurface margin and worn surface. The depth of marginal degradation was excluded as a part of the wear facet.



a b c

Fig. 3. Buccal profiles of wear facet seen on Micro-Vu system. a) wear on a flat worn composite Surface; b) wear on a curved worn surface; and c) marginal degradation.

Results of the instrumental evaluations of 60 experimental casts are shown in Table 2. The distribution of wear on 120 points and the averaged wear on 120 areas are shown in Table 3. Mean of point values was larger than that of area values. The range of measurements was 1,000 μm. Sixty percent of all measurements were less than 300 μm. The standard deviations of measurements were large due to their wide range and distribution.

Table 2. Measurements of wear with the Micro-Vu system.

	Number	Mean (μm)	S.D. (μm)	Std error (μm)	Range (μm)
Point	120	172.43	237.51	21.68	1,000
Area	120	149.25	202.33	18.47	1,000
Total	240	160.84	220.47	14.23	1,000

Table 3. Distribution of the wear on points and averaged wear on areas.

Wear (μm)	0-99	100-199	200-299	300-399	400-499	500-599	600-699	700-799	800-899	900-999	1,000	Total
Number	124	7	16	9	4	1	3	1	1	2	9	240

Non-instrumental evaluation

Table 4 is the complete model for this study. The effect of sixty casts, five evaluators, two scales, and the mental average procedure (MAP) were all statistically significant. The overall effect of repeat measurement was not statistically significant (p=0.2252). The Baldwin illusion was not included in this model because only buccal cusps were ranked. Before further discussion of the results, the importance of the interaction between factors is noteworthy. No interaction higher than 3-way was significant.

Table 4. Full model of ANOVA used in this study.

Source	df	F	Pr>F
Main effect			
Method	1	9.19	0.0024*
Evaluator	4	75.04	0.0001*
MAP	1	13.15	0.0003*
Repeat	2	1.49	0.2252
Cast	59	38.71	0.0001*
Two-way interaction			
Method x Evaluator	4	32.73	0.0001*
Method x MAP	1	1.68	0.1947
Method x Repeat	2	7.18	0.0008*
Method x Cast	59	2.43	0.0001*
Evaluator x MAP	4	9.83	0.0001*
Evaluator x Repeat	8	3.86	0.0002*
Evaluator x Cast	236	2.78	0.0001*
MAP x Repeat	2	0.72	0.4884
MAP x Cast	59	12.71	0.0001*
Repeat x Cast	118	0.98	0.5537
Three-way interaction			
Method x Evaluator x MAP	4	0.15	0.9639
Method x Evaluator x Repeat	8	14.24	0.0001*
Method x Evaluator x Cast	236	1.64	0.0001*
Method x MAP x Repeat	2	0.02	0.9819
Method x MAP x Cast	59	0.45	0.9999
Method x Repeat x Cast	118	0.64	0.9991
Evaluator x MAP x Repeat	8	0.47	0.8774
Evaluator x MAP x Cast	236	0.74	0.9987
Evaluator x Repeat x Cast	472	0.71	1
MAP x Repeat x Cast	118	0.2	1
Four-way interaction			
Method x Evaluator x MAP x Repeat	8	0.18	0.9941
Method x Evaluator x MAP x Cast	236	0.41	1
Method x Evaluator x Repeat x Cast	472	0.7	1
Evaluator x MAP x Repeat x Cast	472	0.23	1
Method x MAP x Repeat x Cast	118	0.17	1
Five way interaction			
Method x Evaluator x MAP x Repeat x Cast	472	0.2	1

*p<0.05, statistically significant difference.

Table 5. Comparison of difference on influencing factors.

Factor	Variable	df	F	Pr>F	Mean	S.D.
Method		1, 7,198	14.33	0.0097*		
	Vivadent				-0.32998	0.060948
	M-L				-0.09902	0.065158
Evaluator		4, 7,195	56.34	<0.001*		
	C				-0.57643[b]	0.09348
	L				-0.07122[c]	0.08765
	N				-0.88753[c]	0.12203
	P				-0.53516[c]	0.09059
	S				0.99785[a]	0.09374
MAP		1, 7,198	9.59	0.0020*		
	MAP				-0.07641	0.062857
	NMAP				-0.35259	0.063294
Repeat measurement		2, 7,197	1.09	0.337		
	1				-0.23088	0.078920
	2				-0.28561	0.078661
	3				-0.12700	0.074219
Cusp height		2, 1,797	32.57	<0.001*		
	H				0.84400[a]	0.15365
	M				-0.00440[b]	0.13131
	L				-0.79752[c]	0.14565

*p<0.05, statistically significant difference.
** Means with the same letter are not significantly different.

The results of overall difference between methods, evaluators, with mental average procedure (MAP) or without mental average procedure (NMAP), repeat measurements, and cusp height are presented in Table 5. The only factor that was not significant is the effect of repeat measurements (p=0.337).

The results of most important interactions between influencing factors are shown in Table 6. Interactions among method, evaluator, and the mental average procedure (MAP) were all statistically significant.

Table 6. Comparison of most important interactions between influencing factors.

Factor	Variable	df	F	Pr>F	Mean	S. D.
Method by evaluator		9, 7,190	37.25	<0.001*		
	Vivadent x C				-1.06788	0.13329
	Vivadent x L				0.33673	0.11886
	Vivadent x N				-0.73395	0.16845
	Vivadent x P				-0.42581	0.12306
	Vivadent x S				0.24101	0.12446
	M-Lx C				-0.08498	0.12862
	M-Lx L				-0.47918	0.12712
	M-Lx N				-1.04111	0.17654
	M-Lx P				-0.64450	0.13293
	M-Lx S				1.75469	0.13449
MAP by evaluator		9, 7,190	29.56	<0.01*		
	MAP x C				-0.56281	0.12653
	MAP x L				-0.18185	0.12915
	MAP x N				-0.32234	0.17962
	MAP x P				-0.53333	0.12132
	MAP x S				1.21831	0.12713
	NMAP x C				-0.59005	0.13773
	NMAP x L				0.03941	0.11847
	NMAP x N				-1.45271	0.16265
	NMAP x P				0.53698	0.13465
	NMAP x S				0.77740	0.13739
Method by MAP		3, 7,196	5.84	0.006*		
	Vivadent x MAP				-0.24129	0.085006
	Vivadent x NMAP				0.08848	0.092478
	M-L x MAP				-0.41867	0.08734
	M-L x NMAP				-0.28651	0.091628

*p<0.05, statistically significant difference.

The correlation of the answers to the three subjective assessment questions and the scale with the better validity used by individual evaluator are listed in Table 7. Except for Evaluator S, the more confidence that the evaluator had toward the scale, the more valid were the data obtained by using that scale.

Table 7. Answers for three subjective assessment questions.

Evaluator	C	L	N	P	S
Easier to learn	Vivadent	M-L	M-L	M-L	Vivadent
Faster evaluation	M-L	M-L	Vivadent	Vivadent	M-L
More confidence on scale	M-L	Vivadent	Vivadent	Vivadent	M-L
More valid on scale	M-L	Vivadent	Vivadent	Vivadent	Vivadent

Discussion

Instrumental evaluation

The impression technique with the Micro-Vu optical metrology system was shown to be ideal for evaluating clinical wear due to the high resolution (0.5 μm) and repeatability (1.0 μm) of the system and the small standard deviation of this technique (0.8 μm). In addition, the ability of viewing the profile of restorations along with their wear facets at a proper magnification facilitates measuring wear on different shaped restorations with different marginal degradation patterns (Fig. 3). When the worn composite surface is flat, it is easy to measure

the occlusal wear (Fig. 3a). However, it is difficult when it is curved since the junction between the floor of composite and enamel becomes more obscured (Fig. 3b). By increasing the power of magnification, the locations of floor and margin can be identified easily. Marginal degradation is characterized by a breakdown at the resin-enamel wall interface. Reduced tensile strength due to decreased particle size of composite and weak bonding strength of bonding agent are believed to play main roles in marginal degradation.[40] Because of the difference in the mechanisms involved, marginal degradation (Fig. 3c) should not be confused with wear. With the aid of the Micro-Vu metrology system, this mistake is prevented.

The technique proposed in this study is similar to the technique used by Leinfelder et al.[12] In this present study, more sections were made and a different viewing and measuring system was used. It is quite clear that the computer-aided system used in this study may achieve a more favorable performance than an optical traveling microscope.

Roberts and Soderholm's study showed that among three techniques compared, cross-sectioned impression technique, SEM photometric techniques and profilometer evaluation, the standard error for the impression technique was the smallest.[32] Our study also indicated that this technique was very precise.

Non-instrumental evaluation

For comparison of the overall validity of the non-instrumental evaluation methods, three major findings were: a) the M-L scale was more valid than the Vivadent scale, b) the overall values obtained by using the M-L scale were smaller than values by the Vivadent scale, and c) the values obtained by both the M-L scale and the Vivadent scale were larger than the values measured by the Micro-Vu system (Table 5). To our knowledge, no previous study has been published that compared the validity of different scales. Our second major finding was the opposite of Bryant's study[15] and was also different from the study by Taylor et al.,[38] whose results showed that the values obtained by the M-L scale and the Vivadent scale were equivalent. The third finding disagreed with Winkler et al.[17] who found that the visual estimates of wear using the clinical casts were significantly lower than those using mechanical measurement.

Data obtained by different evaluators were significantly different ($p<0.001$). The data obtained by Evaluator L were most valid while those by Evaluator S were least valid. The effect of the evaluator factor was so influential as to shift the direction of the results when its interactions with other factors were considered. For example, the overall validity of the M-L scale was better than that of the Vivadent scale. When the evaluator effect was further analyzed, this statement was true only for Evaluator C. For the other four evaluators, the validity of the M-L scale was worse than that of the Vivadent scale (Table 6). Our results support the findings by Taylor et al.[36] that the controlling limitation for non-instrumental methods is the evaluators' ability to discriminate wear.

For the mental average procedure, Taylor et al.[36] and Bryant[15] stated that because the amount of occlusal loss was usually uneven along the margins, the evaluator presumably performed this procedure for both the casts of the experimental restorations and standards while making the comparison. Surprisingly, using the mental average procedure increased the validity of the results ($p=0.0020$). However, when the effect of this procedure was analyzed for its interaction with the evaluator effect (Table 6), the results were different. In fact, the mental average procedure only increased the validity for Evaluator N. For the other four evaluators, the mental average procedure either had no influence or decreased their validity. Again, the interaction of evaluator effect was so powerful as to reverse the overall effect. For the two methods investigated in this study, their interaction with

the mental average procedure was also significant (p=0.006). Regardless of whether the mental average procedure was employed, the Vivadent scale was more valid than the M-L scale. Regardless of which scale was used, evaluations with the mental average procedure were less valid than without it. The former result seems to be contradictory to the result of overall validity, for which the M-L scale was more valid. The reason is that evaluators tended to give smaller values, even smaller than values obtained by the Micro-Vu system, when using the Vivadent scale while the mental average procedure was not indicated (mean=0.088). Therefore, our conclusion for the mental average procedure is that this procedure might influence different evaluators to a certain extent and the validity of different methods as well.

The effect of experience was tested by repeat measurements. Evaluators gained experience by repeat training programs and measurements. There was no significant difference among repeat measurements (p=0.337). Although some authors suggested that experienced evaluators achieved higher intra-evaluator and inter-evaluator agreement,[15] our study showed that experience did not increase the validity of the results. In other words, more practice or experience might benefit the precision of the measurements, but not the validity for a particular scale.

Baldwin first studied the effect of size contrast in 1895. In general, size contrast means that any stimulus surrounded by larger stimuli seems smaller than it really is.[39] Winkler et al.[17] speculated that the Baldwin illusion had an influence for underestimation of the depth of wear at the margin. The results showed that cusp height played an important role in this study (p<0.001). The validity for the evaluations of medium height cusps was best among the three different cusp height categories. The effect of the Baldwin illusion was not significant.

As far as subjective assessment is concerned, both scales were equally favored (Table 7). This is different from the evaluators in Bryant's study who were all more confident with the Vivadent scale.[15] The relationship between the confidence level and validity showed that the validity increased with subjective confidence except for Evaluator S, who had previous experience with the M-L scale.

Conclusion

The impression technique in conjunction with the Micro-Vu optical metrology system for measuring wear of composite resins utilized in this study was proven to be highly precise. The ability to view the profile of a wear facet on a thin section as well as to magnify it for detailed appreciation was advantageous for assessing wear. This new technique is recommended for use in future studies involving wear loss evaluation.

For non-instrumental methods, the three major findings in this study were: a) different scales, evaluators, and the mental average procedure influenced the validity of the results, b) the validity was not improved by practice or experience, and c) interactions between factors, especially with evaluators, were high, which in turn definitely affected the result obtained by using non-instrumental evaluation methods. Therefore, the desirability of continuing these types of evaluation techniques to assess wear should be reconsidered.

Acknowledgment

The authors would like to thank Dr. Jack Lemons and Mr. Preston Beck for their advise and assistance when using the Micro-Vu system, and Drs. Adam Clark, Samir Nanjapa, Winnie Lee for their efforts with evaluations.

References

1. American Dental Association Council on Dental Materials, Instruments and Equipment. Acceptance program guidelines for composite resin materials for occlusal Class I and Class II restorations. 1984.
2. American Dental Association Council on Scientific Affair. American dental association acceptance program guidelines for composite resins for posterior restorations. 1989.
3. American Dental Association Council on Dental Materials, Instruments and Equipment. American dental association acceptance program guidelines for tooth-colored restorative materials for posterior teeth. 1991.

4. American Dental Association Council on Scientific Affair. Acceptance program guidelines for restorative materials. 1996.
5. Cvar JF, Ryge G. Criteria for the clinical evaluation of dental restorative materials. US Public Health Service Publications, no. 790-244, USGPO, 1971.
6. Osborne JW, Phillips RW, Gale EN, Binon PP. Three-year clinical comparison of three amalgam alloy types emphasizing an appraisal of the evaluation methods used. J Am Dent Assoc 1976; 93: 784-9.
7. Walls AWG, Murray JJ, McCabe JF. The management of occlusal caries in permanent molars. A clinical trial comparing a minimal composite restoration with an occlusal amalgam restoration. Br Dent J 1988; 164: 288-92.
8. Chadwick RG, McCabe JF, Walls AWG, Mitchell HL, Storer R. Comparison of a novel photogrammetric technique and modified USPHS criteria to monitor the wear of restoration. J Dent 1991; 19: 39-45.
9. Collins CJ, Bryant RW and Hodge KL. A clinical evaluation of posterior composite resin restorations: 8-year findings. J Dent 1998; 26: 311-7.
10. Mahler DB, Terkla LG, Van Eysden J, Reisbick MH. Marginal fracture vs. mechanical properties of amalgam. J Dent Res 1970; 49: 1452-7.
11. Goldberg AJ, Rydinge E, Santucci EA, Racz WB. Clinical evaluation methods for posterior composite restorations. J Dent Res 1984; 63: 1387-91.
12. Leinfelder KF, Taylor DF, Barkmeier WW, Goldberg AJ. Quantitative wear measurements of posterior composite resins. Dent Mater 1986; 2: 198-201.
13. de Rijk WG, Conner ML, Jennings KA. The in vivo wear resistance of dental composites with enhanced polymerization. J Dent Res 1984; 63: 337. Abstr 951.
14. Moffa JP, Lugassy AA. Calibration of evaluators utilizing the M-L occlusal loss scale. J Dent Res 1986; 65: 302. Abstr 1197.
15. Bryant RW. Comparison of three standards for qualifying occlusal loss of composite restorations. Dent Mater 1990; 6: 60-2.
16. Mair LH. The measurement and analysis of clinical abrasion- a modified approach. Dent Mater 1990; 6: 271-5.
17. Winkler MM, Lautenschlager EP, Boghosian A, Greener EH. Visual versus mechanical wear measurement of dental composite resin. J Oral Rehabil 1996; 23: 494-500.
18. Handelman SL, Jensen QE, Pameijer CH. Quantitative assessment of sealant wear in vivo. J Prosthet Dent 1978; 40: 531-3.
19. Dennison JB, Powers JM, Charbeneau GT. Measurement of in vivo wear on posterior composite restorations. J Dent Res 1980; 59: 318. Abstr 202.
20. Urquiola NJ, Charbeneau GT. Quantitative evaluation of clinical wear of posterior composite resin restorations. J Dent Res 1981; 60: 583. Abstr 1094.
21. Vrijhoef MMA, Letzel H, Hendriks FHJ. A method to determine the loss of substance of dental restorations. J Oral Rehabil 1985; 12: 9-16.
22. Santucci EA, Racz WB. One year evaluation of posterior restorations in a primate model. J Dent Res 1982; 61: 247. Abstr 618.
23. Mitchem JC, Gronas DG. In vivo evaluation of the wear of restorative resin. J Am Dent Assoc 1982; 104: 333-5.
24. Lutz F, Phillips RW, Roulet JF, Setcos JC. In vivo and in vitro wear of potential posterior composites. J Dent Res 1984; 63: 914-20.
25. Roberts MJ, Soderholm KJM. Comparison of three techniques for measuring wear of dental restorations. Acta Odontol Scand 1989; 47: 367-74.
26. Roulet JF, Reich T, Lutz F. High precision occlusal mapping: a new method for measuring of posterior composites. J Dent Res 1983; 62: 220. Abstr 457.
27. Eick JD, McGarrah HE, Lamb RD. Application of stereo-photogrammetry to measure wear of posterior composites. J Dent Res 1984; 63: 335. Abstr 1482.
28. Williams DF, Cunningham J, Lalor MJ, Groves D, Atkinson JT.Laser techniques for the evaluation of wear in Class II restorations. J Oral Rehabil 1983; 10: 407-14.
29. Lambrechts P, Vanherle G, Vuylsteke M, Davidson CL. Quantitative evaluation of the wear resistance of posterior dental restorations: a new three-dimensional measuring technique. J Dent 1984; 3: 252-67.
30. DeLong R, Pintado M, Douglas WH. Measurement of change in surface contour by computer graphics. Dent Mater 1985; 1: 27-30.
31. McDowell GC, Bloem TJ, Lang BR, Asgar K. In vivo wear. Part I: the Michigan computer-graphic measuring system. J Prosthet Dent 1988; 60: 112-20.
32. Adams LP, Wilding RJC. Tooth wear measurements using a reflex microscope. J Oral Rehabil 1988; 15: 605-13.
33. Bayne SC, Taylor DF, Rekow ED, Wilder AD, Heymann HO. Confirmation of Leinfelder clinical wear standards. Dent Mater 1994; 10: 11-8.
34. Mehl A, Gloger W, Kunzelmann KH, Hickel R. A new optical 3-D device for the detection of wear. J Dent Res 1997; 76: 1799-807.
35. Ahlbom A, Norell S. Introduction to modern epidemiology, 2nd ed. Chestnut hill, MA: Epidemiology Resources, 1990.
36. Taylor DF, Bayne SC, Sturdevant JR, Wilder AD. Correlation of M-L, Leinfelder, and USPHS clinical evaluation techniques for wear. Dent Mater 1990; 6: 151-3.
37. Taylor DF, Bayne SC, Sturdevant JR. Vivadent comparison to M-L, Leinfelder, and USPHS clinical scales. J Dent Res 1990; 69: 160. Abstr 416.
38. Soderholm KJM, Roberts MJ, Antonson DE, Anusavice KJ, Mauderli AP, Sarrett DC, Warren JW. Visual and profilometric wear measurements. Acta Odontol Scand 1992; 50: 121-7.
39. Coren S, Girgus JS. Seeing is deceiving: the psychology of visual illusion. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978.
40. Suzuki S, Leinfelder KF. Localized wear and marginal integrity of posterior resin composites. Am J Dent 1993; 6: 199-203.

Reprint request to:

Dr. Shiro Suzuki
Department of Prosthodontics, University of Alabama at Birmingham, School of Dentistry
1919 7th Avenue South, Birmingham, AL 35294-0007, USA
Fax: +1-205-975-6108 E-mail: shiro@mail.dental.uab.edu

Received January 19, 2005. Revised February 13, 2005. Accepted February 16, 2005.