



Walter A<sup>1</sup>, Möltner A<sup>2</sup>, Böckers A<sup>3</sup>,  
Rüttermann S<sup>4</sup>, Gerhardt Szép S<sup>4\*</sup>

<sup>1</sup>Department of Operative Dentistry, Centre for Dentistry and Oral Medicine (Carolinum), Goethe-University Frankfurt, Frankfurt, Germany

<sup>2</sup>Competence Centre for Examinations in Medicine/Baden-Württemberg, Medical Faculty, University of Heidelberg, Heidelberg, Germany

<sup>3</sup>Medical Faculty, Institute of Anatomy and Cell Biology, University in Ulm, Albert-Einstein Allee 11, 89081 Ulm, Germany

<sup>4</sup>Department of Operative Dentistry, Centre for Dentistry and Oral Medicine (Carolinum), Goethe-University Frankfurt, Frankfurt, Germany

**Received:** 29 September, 2018

**Accepted:** 16 October, 2018

**Published:** 18 October, 2018

**\*Corresponding author:** Susanne Gerhardt-Szep, Department of Operative Dentistry, Centre for Dentistry and Oral Medicine (Carolinum), Goethe-University Frankfurt, Frankfurt, Germany, Tel. No: +49-69-6301-7505; Email: S.Szep@em.uni-frankfurt.de

<https://www.peertechz.com>



## Research Article

# Video-based assessment of practical operative skills for Undergraduate dental students

## Abstract

**Introduction:** The aim of this study is to evaluate, within the scope of an experimental design, to what extent the assessment of two different settings of prepared cavities, based on video sequences, containing digital analysis tools of the prepCheck software, as well as to what extent they deviate from one another and are reliable.

**Materials and Methods:** For this prospective, single-centred, experimental study, 60 examination cavities related to a ceramic inlay preparation were assessed by four trainers in two different settings (A: video film versus B: video film plus an analogue model assessment) by using a standard checklist. The examined parameters contained: the 1. preparation / outer edges, 2. surface & smoothness / inner edges, 3. width & depth, 4. slide-in direction, 5. outer contact positioning and 6. overall grade on a Likert scale of 1 = 'excellent', 2 = 'very good', 3 = 'good', 4 = 'satisfactory' to 5 = 'unsatisfactory'. An evaluation questionnaire with 33 items was additionally addressed to the concept of application of a digital-analytic software. The statistical analysis, using SAS 9.2 (SAS Institute Inc., Cary, USA, PROC MIXED) and R (Version 2.15, Package lme4) concerned the reliability, inter-rater correlation and significant factors at a p of 0.05.

**Results:** The assessment of the individual criteria and overall grade of the control group (A) were, on average, lower (i.e. better) than in the study group (B), yet with the exception of the 'outer contact positioning', without conclusive statistical significance. The reliability lay at an average of  $\alpha=0.83$  (A) and  $\alpha=0.79$  (B). The maximum reliability of the criteria 'preparation edge', 'surface', 'width & depth' as well as 'overall grade' were reasonable in the assessment mode, with  $\alpha > 0.7$ . The inter Video-based Assessment 3 rater correlation was at an average of  $0.43 < r < 0.74$  higher in assessment mode A than B that comprised  $0.35 < r < 0.60$ .

**Conclusion:** The current examination shows an average reliability in the assessment mode A that exceeds the requirements for practical examination ( $\alpha \geq 0.6$ ) and also fulfils the general requirements for 'high-stake' examinations of  $\alpha \geq 0.8$ .

## Abbreviations

Video-Based Assessment; Dental; Objective Evaluation; Practical Skills; Checklist; Performance; Operative Dentistry; OSPE; Undergraduate degree

## Introduction

The practical development of skills, i.e. the process of gaining expertise in procedures and techniques required for operative dentistry, incorporates a fundamental part of any study in dentistry. In this way, the aim of the sixth semester of the dental education is also, in scope of the phantom course of operative dentistry, to optimally prepare students for the treatment of patients. Above all in regards to cavity

preparation, which is one of the basic competencies required by any dentist later in their career, the students were unsure of themselves at first. These, above all include parameters such as cavity depth, i.e. width, surface smoothness and the cavity edge form [1]. Deviating assessments provided by different trainers may lead to frustration and confusion among students [2]. Assistance should be provided by means of modern media, such as computer-generated digital analysis tools [3-5], of which the implementation in the dental curriculum was supported in many different ways [5-8]. To achieve this, the services provided by students are monitored and assessed by trainers at various times, both formatively and summatively. Ideally, therefore, all of the assessments should incorporate all characteristics of reliability, validity, responsibility, flexibility, comprehension, implementation ability and relevance [9, 10].

The assessment of practical skills in dental and medical schools requires considerable time and effort on part of a supervising faculty [1, 11, 12]. The live assessment of these skills poses a significant problem of resources to dentistry schools and results in complications arising in the execution and scheduling of their daily activities [11]. Dental literature, on the other hand, acknowledges the need for the objective assessment of skills in operative training [1, 10]. Structured grading systems, such as the Objective Structured Practical Examinations (OSPE) or OSCEs (Objective Structured Clinical Examinations) were specifically designed to reduce subjectivity [1,10,13,14]. In order to fulfil the general requirements of 'high-stake' examinations, a specific number of examiners as well as checklists, should be implemented in the assessment of cavities in an OSPE-design 1.

A major disadvantage of requiring live assessment is the substantial demand on time and resources involved in getting several staff members to observe and assess students' performance 1. As an alternative, supervisors could use videos that reduce some of the logistical overhead [11]. Video-based assessment allow raters to be blind to certain aspects of the performance, such as the identity of trainees, that may otherwise engender bias in rating [11, 12, 15-23]. It further facilitates a more detailed review of a learner's performance and provides additional time for the rater to fully focus on the performance of a trainee [12]. In addition, videos can be reviewed several times, by several different raters. Finally, trainees can review video recordings themselves and thus be given the opportunity of enhancing their learning through debriefing methods [12].

Such video-based assessments have already been implemented and evaluated in various surgical subject areas [14, 15, 17, 21, 22]. Most of them conclude that videobased assessment provides an efficient and reliable educational environment with satisfactory rater consistency and evidence for validity [12, 14, 17-20, 23].

The type of video assessment for cavity preparations in dental medicine most suitable for 'high-stake' examinations, has not been clarified in any work of literature, up to now. It is also not clear, whether the video assessment of dental cavities (in a simulation model) alone, i.e. their additionally-regarded manual-analogue component, may possibly imply a difference in grading. The additional consideration of models also leads to a higher demand in time and personnel, as this must be carried out individually and assessed by the examiner him or herself. This setting, namely that each model must be individually assessed by examiners and consequently evaluated within a unanimous procedure, where a final grade is allocated, describes the current stand in the type of situation for grading. Optimally, we assume that three to four examiners are required here [1].

To address this gap, this study aims to compare two different settings for the videobased practical operative skills evaluation, including an analysis tool. In an experimental design, it should be evaluated to what extent different assessments of prepared cavities based on sequences of videos

containing digital analysis tools, deviate from one another and the reliability that each possesses. In the study control group (Part A), examiners assessed examination activities, which they observed in a video that illustrated various parameters of a digital analysis tool. Finally, the examiners received the opportunity in Part B (study group), to additionally regard the real examination activity model themselves and modify their previously-provided video assessment. Two main research questions should be answered for the examination:

1. Do the various modes of assessment used in both examined settings (control and study group) affect their reliability?
2. What influence do the different modes of assessing the examined settings (control and study group) have on the overall assessment of study participants (trainers)?

In addition, we were interested in the evaluation of study participants, with regards to the application concept of the digitally-analysed software and study procedure.

## Materials and Methods

This is a prospective, single-centered experimental study conducted at the Goethe University, in Frankfurt. The application for ethical approval was received on 9th August 2015 and approved by the Ethics Committee of the Department of Medicine at the Goethe Universität Frankfurt-on-Main on 9th September 2015, holding the reference number 302/2015. The examination period was provided as being between 7th October 2015 and 13th October 2016.

## Participants and Assessed Parameters

The criteria for selecting suitable study participators (examiners) was determined in the run-up. Their inclusion criteria included belonging to the department of operative dentistry and the fact that they proved to have little or no experience of the PrepCheck software (had only worked with it up to ten times). In preparation of the study, the trainers were first prepared for the assessment scenario through two trainthe-teacher events and their evaluation skills calibrated. The exact time frame is represented in (Table 1). For the experimental study, 60 cavity preparations were assessed by 4 trainers in two different assessment modes (Part A and B). The exact time frame is represented in (Table 2).

The cavity preparations were scanned using the CEREC-Omniscam (Dentsply Sirona, York, USA) that in the scope of an obligatory and summative OSPE examination of students in their 6th study semester, was submitted as examination material (in the winter semester 2013/2014 and summer semester 2014) and graded by the trainers.

This dealt with distal occlusion preparations for ceramic inlays in premolar teeth. These examinations lay three semesters behind at the time of the study, so that the study participants (examiners) had no memory of either the grades provided, nor of the students whose examination papers these represented.

**Table 1:** Schedule and procedure of both train -the-teacher events preceding the experimental study. Five model videos were assessed for each round(min=minute).

Train-the-teacher no:1			Train-the-teacher no:2
Time	Duration	Task	Time
15:15-15:17	2 min	introduction, clarification of procedure	13:37-13:38
15:17-15:23	6 min	video of demo model for practice and explanation, clarification of questions	13:38-14:43
15:23-15:39	16 min	video assessment(Round 1)	13:43-14:15
15:39-15:44	5 min	Consensus on provision of grades	14:15-14:23
-	-	break	14:23-14:36
-	-	video assessment(Round 2)	14:36-14:55
-	-	Consensus on provision of grades	14:55-15:12
15:44-15:49	5 min	Evaluation	15:12-15:17
	34 min	Total duration	

**Table 2:** Schedule and procedure of the experimental study including part A and part B(min=minute).

Experimental study cavities 1-30			Experimental study cavities 31-60	
Time	Duration	Task	Time	Duration
13:20-13:21	1 min	Introduction	13:20-13:21	1 min
13:21-13:25	4 min	Demo of a model as a reminder,clarificaton of questions	13:21-13:26	5 min
13:25-14:10	45 min	10 videos(part A)+10 models(part B)	13:26-14:11	38 min
14:10-14:15	5 min	Break	14:11-14:23	12 min
14:15-15:00	45 min	10 videos(part A)+10 models(part B)	14:23-15:00	37 min
15:00-15:05	5 min	Break	15:00-15:14	14 min
15:05-15:50	45 min	10 videos(part A)+10 models(part B)	15:14-15:42	28 min
15:50-15:55	5 min	Evaluation	15:42-15:47	5 min
	2 hr 35 min	Total duration		2 hr 27 min

### Assessment Tool

The assessment of the cavities occurred by means of checklists comprised by Schmitt et al. 2016 1, in support of the study. These incorporated five items (1. preparation edge / outer edges, 2. surface & smoothness / inner edges, 3. width & depth, 4. slide-in direction, 5. outer contact positioning and 6. overall grade). The individual assessments (Table 1) were indicated on a Likert Scale of 1 = excellent, 2 = very good, 3 = good, 4. = satisfactory to 5 = unsatisfactory (Table 3, Figures 1-7). After completion of the assessments, the examiners were questioned on general matters (n = 3), by means of an evaluation questionnaire containing 33 items, such as age, gender, teaching experience, the application concept of the digital-analytical software (n = 17), individual assessment preferences (n = 3), study procedure (n = 10) (Tables 6 and 7). Freely-composed commentaries rounded off the evaluation questionnaire.

### Procedure

By means of the Wilcoxon-Matched-Pairs-Test using the Bonferroni version, a case number of n=60 was determined from the results of a preceding train-the-teacher event at  $\alpha=0.0125$  and a probability of  $P(X+X'>0)=0.25$ , in order to guarantee a power of 80% for four trainers.

The cavities were randomly allocated to both groups (Parts A and B) of the experiment. The randomisation took place by

**Table 3:** Sequence and duration of the parameter settings in the prepCheck video and associated items of the checklist, for the assessment of grades(min=minimum,s=seconds).

Step	setting in prepcheck	Grade assessment criteria	approx.duration in video(seconds)
1	(zoom in to cavity)	(first view of model)	3
2	preparation edge	1.preparation edge(outer edges)	17
3	surface consistency, undercut	2.surface & smoothness(inner edges)	31
4	side section images d-m	3.width & depth	60
5	undercut	4.slide-in direction	6
6	undercut	5.Outer contact positioning	5
		total duration	2 min 02 sec

entering coded models into an online randomizer (<https://www.random.org>).

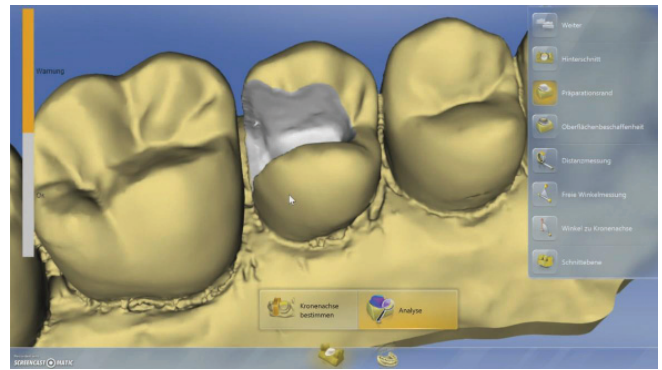
### Video-based Assessment

The composition of the video of the digitalised teeth was created in the so-called analysis mode of the prepCheck software (Dentsply Sirona, York, USA) and followed by the free-of-charge programme 'ScreenCast-O-Matic' (Softonic International, Barcelona, Spain. Version 2.0). The duration of



the individual videos encompassed 122 seconds on average, while they portrayed six different settings that were selected in the run-up of the prepCheck software (Dentsply Sirona, York, USA) (Table 3). A beamer and a screen, as well as a connection to a laptop were required for the videos. The environment of the room for both scenarios (Part A and B) is represented in (Figures 8,9). In Part B, the participants (examiners) agreed on unanimous assessment conditions, in regards to the enlargement aids used (2.7 x with light).

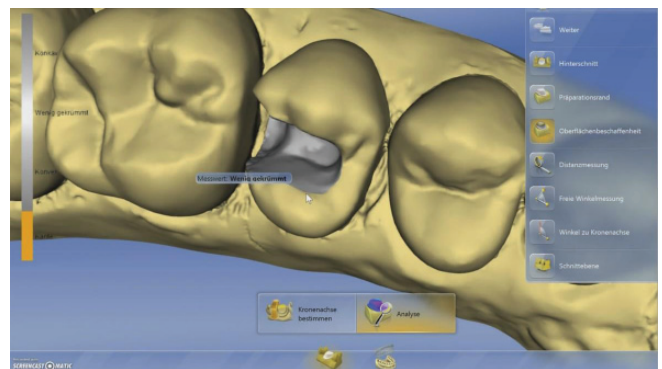
For Part A (control group with a prepCheck video), the participants could enter their assessment questionnaires, while the video was played (Image 1). For this, they had maximum 120 seconds time. The plastic tooth 15 to be assessed that was built-in to a simulation model, indicated an occlusal width of approx. 0.7 x 0.9 cm. The size of the tooth on the screen comprised an average of approx. 50 x 70 cm, which encompassed an approx. 75 x enlargement. For Part B (study group with prepCheck video + consequent models), the preparations to be assessed



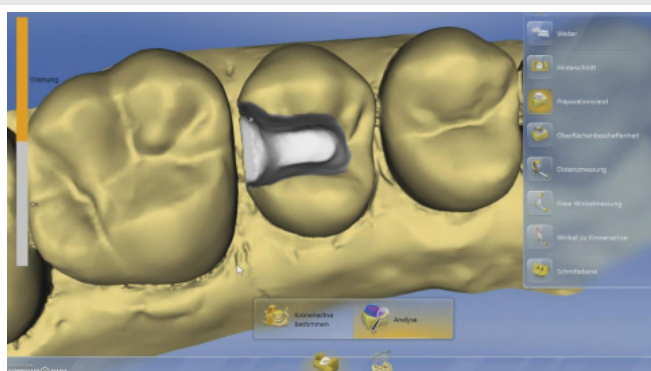
**Figure 3:** In the second step, the outer edges of the preparation are represented from all sides for approx. 17 seconds, while the prepCheck is set to 'preparation edge', and the setting is automatically slanted in all directions (buccal, oral, mesial and distal), so that one always has a direct view for the assessment. This setting is entered into the assessment questionnaire as the first parameter (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnitt Ebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, Warnung = warning).



**Figure 1:** In the first step (1st Part), a first view is taken of the model to be assessed as an overview when the prepCheck is set to 'zoom to cavity' for approx. 1 second (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnitt Ebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, Warnung = warning).



**Figure 4:** In the third step, the surface, smoothness and inner edges of the preparation are represented from all sides for approx. 31 seconds in the prepCheck setting 'surface consistency', by automatically slanting the setting in all directions (buccal, oral, mesial and distal), so that one always has a direct view for making the assessment. The programme provides support through indications on the concavness i.e. convexness of the preparation. Recognisable edges are coloured in orange. At the cursor, which is represented by an arrow here, a direct response to the consistency is provided. This setting is entered into the assessment questionnaire as the second parameter (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnitt Ebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, konkav = concave, wenig gekrümmt = slightly bent, konvex = convex, Kante = edge, Meßwert = measured value).



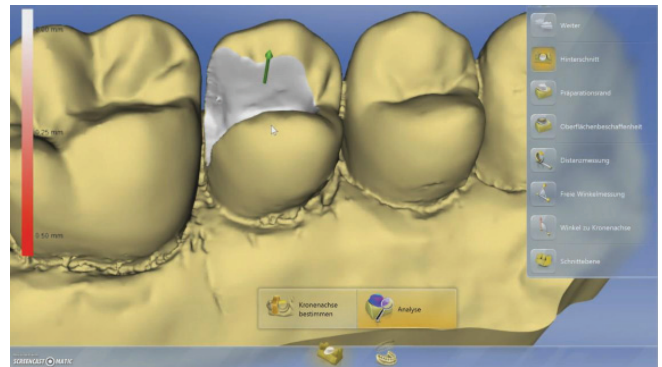
**Figure 2:** In the first step (2nd Part), an enlarged view is taken at the cavity to be assessed, while the prepCheck is set to 'zoom to cavity' for approx. 2 seconds (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnitt Ebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, Warnung = warning).

were maintained in models (tooth model, KaVo Dental GmbH, Biberach, Germany) on a table (Figure 2). At every seat, basic dental utensils (a mirror, probe) were provided that included a lead pencil and cotton wool buds. The examiners used the model with the corresponding reference number and the filled-in checklist with the corresponding individual assessment from Setting A. They examined the already available individual grade and modified these, where necessary. For the assessment, the teeth could be taken out of the models and the preparation edges marked with a lead pencil, where necessary. This was meant to assist in more easily recognising undesired bevels

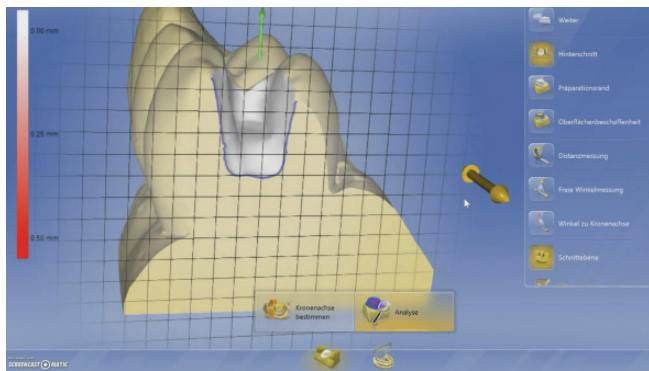
during the preparation. Before completion of the assessment time and prior to their being passed onto the next examiner, the cavities had to be cleaned with a moist cotton wool bud. Maximum 120 seconds was foreseen for the assessment of each model. In the background, a count-down timer ran above the beamer that could be viewed by all participants (Figure 2).

## Statistical Analysis

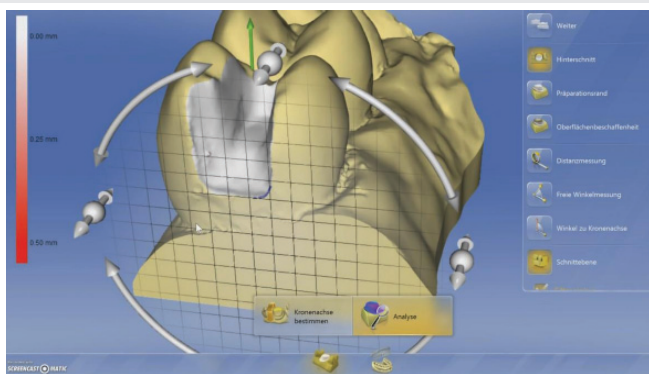
The case number calculation took place in co-operation



**Figure 7:** In the final step, the outer contact positioning of the preparation is represented for approx. 5 seconds in the prepCheck setting 'undercut'. Here, the cavity can also be pointed in all directions. The programme provides further support through giving indications on the measurements of the preparation and through a green arrow pointing at the preparation axis. The film ends after approx. 120 seconds. This setting is entered into the assessment questionnaire as the final parameter (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnittebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, konkav = concave, wenig gekrümmt = slightly bent, konvex = convex, Kante = edge, Meßwert = measured value).



**Figure 5:** In the fourth step, the width and depth of the preparation are represented from all sides for approx. 60 seconds in the prepCheck setting 'sectional plane'. A screening of 1mm is projected onto the preparation, so that one even receives a metrical analysis. The programme provides support through indicating measurements of the preparation and already at this stage points a green arrow at the preparation axis. This setting is entered into the assessment questionnaire as the third parameter (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnittebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, konkav = concave, wenig gekrümmt = slightly bent, konvex = convex, Kante = edge, Meßwert = measured value).



**Figure 6:** In the fifth step, the slide-in direction of the preparation is portrayed more precisely in the prepCheck setting 'undercut' for approx. 6 seconds. A wheel is projected on the preparation, by which the rotation can slant the cavity in more directions and in that way also examine the indications. The programme furthermore provides support of indication measurements of the preparations and a green arrow pointing at the preparation axis. This setting is entered into the assessment questionnaire as the fourth parameter (Translation for: weiter = further, Hinterschnitt = undercut, Präparationsrand = preparation edge, Oberflächenbeschaffenheit = surface consistency, Distanzmessung = measurement of distance, freie Winkelmessung = free angle measurement, Winkel zur Kronenachse = angle of the crown axis, Schnittebene = cut level, Kronenachse bestimmen = determine crown axis, Analyse = analysis, konkav = concave, wenig gekrümmt = slightly bent, konvex = convex, Kante = edge, Meßwert = measured value).

with the Institute of Biostatics and Mathematical Modelling, in Frankfurt-on-Main. The assessment of the results occurred by means of the statistic programmes SAS 9.2 (SAS Institute Inc., Cary, USA, PROC MIXED) and R (Version 2.15, Package lme4). Basic data was retrieved and an analysis of the similarity of the mean values carried out between the observers (ANOVA for dependent observations, as the same models were used).

Finally, the inter-correlations of the assessments of the four raters were calculated among each other. For the comparison of the assessments in Part A and B, each of the four observer ratings were determined and both parts (A and B) tested using a ttest for paired samples. In order to determine the overall reliability of both scenarios, the six single-assessment parameters were complemented by a further 'mean' variable. In addition, a test was carried out to determine the differences between both alpha values for Parts A and B, followed by the reliability test for the 'mean' of the grades of both scenarios.

The statistical assessment was carried out in co-operation with the Competence Centre for Examinations in Medicine, Baden-Württemberg of the Medical Faculty, Heidelberg.

## Results

### Collected preparation parameters and inter-rater correlations

The descriptive, statistical assessment of the individual assessment providing the mean value, standard deviation, median, minimum and maximum, as well as the calculation of the reliability, took place simultaneously for all criteria ('mean'), i.e. separately from one another in regards to the 'preparation edge/outer edges', 'surface & smoothness/inner edges', 'slide-in direction', 'outer contact positioning', 'width & depth' and 'overall grade' (Table 4). In conclusion, the following results can be summarised in the following way: the



assessments of the individual criteria and overall grade were in the control group on average lower (i.e. better) than in the study group (prepCheck video + consequent model), however with one exception that showed no statistical significance. For the assessment of the parameter 'outer contact positioning', the alpha significantly rose from 0.56 (Part A) to 0.74 for Part B. The results of the inter-rater correlations are outlined in (Table 4).

### Assessment questionnaire

All distributed assessment and evaluation questionnaires

were returned after being filled-in. The exclusion rate lay at 0%. The indications on the included study populations are to be taken from (Table 5). The results of the evaluation can be viewed in (Tables 6,7). An excerpt from the freely-composed commentaries is to be taken from (Table 8).

### Discussion

This study establishes evidence to support the reliability of video-based assessments of operative competency in performing cavity preparations in dentistry. To the best of our knowledge, this is the first study to prospectively compare

**Table 4:** Result of the individually-examined criteria of the control and study groups(ICC='inter-rated correlations',MV='mean value',SD='standard deviation', med='median', min='minimum',max='maximum',\*='significant difference').

Group	parameter	MV	SD	med	min	max	cronbach's alpha	ICC(min-max)
Controls(Teil A)	All	2.98	±0.55	3	2	5	0.833	0.43-0.74
	Preparation/outer edges	2.98	±0.86	3	1	5	0.778	0.37-0.60
	Surface & smoothness/inner surfaces	2.83	±0.68	2.75	2	5	0.737	0.24-0.64
	slide-in direction	2.8	±0.69	2.75	1	5	0.498	0.10-0.30
	Outer contact positining	2.74	±0.73	2.87	1	5	0.564	0.06-0.37
	width & depth	3.23	±0.88	3.12	1	5	0.789	0.36-0.64
	Overall grade	3.29	±0.84	3.25	1	5	0.768	0.26-0.72
Study(part B)	All	3.28	±0.54	3.25	3	5	0.793	0.35-0.60
	Preparation/outer edges	3.2	±0.93	3.5	1	5	0.705	0.17-0.57
	Surface & smoothness/inner surfaces	3.14	±0.72	3	2	5	0.727	0.34-0.45
	slide-in direction	2.97	±0.71	2.75	2	5	0.584	0.19-0.39
	Outer contact positining	3.14	±0.89	3.37	1	5	0.741	0.22-0.62
	width & depth	3.55	±0.90	3.62	2	5	0.736	0.35-0.53
	Overall grade	3.68	±0.89	3.75	2	5	0.732	0.29-0.54

**Table 5:** Data on subject group(consisting of four examiners: one man and three women; MV='mean value', SD='standard deviation', med='median',min='minimum',max='maximum')

Group population	MV	SD	Med	Min	Max
...(age) in years	38.75	±10.44	37.25	29	52
...years of teaching experience	12.25	±10.65	10.75	2.5	25
Have used prepCheck... times	4.75	±1.260	4.5	3	7

**Table 6:** Result of question 6, assessed on a scale of 1 to 5, where 1='fully agree';2='partially agree';3='uncertain';4='do not agree';Max='maximum').

Question 6:During the assessment of the cavity by means of prepCheck, I found it useful that i(opposed to without this help)...	MV	SD	Med	Min	Max
could assess the depth better.	3	±0.96	2.5	2	4
could assess the occlusal width better.	3	±1.11	3.25	1	4
could assesss the width of the gingival step.	3	±0.98	3	2	4
could assess the approximate outer contact positioning better.	4.12	±0.54	4.25	3	5
could assess the form of the edges better.	2.65	±0.98	2.25	2	4
could assess the edges of the cavities better.	2.87	±1.12	3	1	4
could assess the sharp edges on the inner cavity angles better	3.17	±0.77	3	2	4
can see points under the slilde-in direction better.	2.65	±0.98	2.25	2	4
could assess the angle of the conicity better.	2.75	±0.91	2.5	2	4
could assess whether the cavity base was straight	2.75	±0.96	2.5	2	4



**Table 7:** Result of questions 5 and 7 up to 24, assessed on a scale of 1 to 5, where 1='fully agree',2='partially agree';3='uncertain';4='do not agree';5='do not agree at all'(MV='mean value'; SD='standard deviation';Med='median';Min='minimum';Max='maximum').

Evaluation	MV	SD	Med	Min	Max
Question 5: I feel the introductory event with the prepCheck was a good preparation.	1	±0.00	1	1	1
Question 7: In my opinion, the dental examiners cannot be replaced by prepCheck for the assessment of cavities	1	±0.00	1	1	1
Question 8: I recommend prepCheck to colleagues.	1.87	±0.95	1.75	1	3
Question 9: I feel working with prepCheck is useful for calibrating trainers.	1.87	±0.95	1.75	1	3
Question 10: I feel that one should assess cavity preparations composed in the scope of examinations by means of prepCheck.	3	±1.41	2.5	2	5
Question 11: Please provide an assessment of prepCheck in form of a school grades: 1= excellent to 6=unsatisfactory	2.87	±0.89	2.75	2	4
Question 12: When assessing preparations, I like to look at images.	2.5	±1.11	2.75	1	4
Question 13: When assessing preparations, I like to look at videos.	2.5	±1.11	2.75	1	4
Question 14: For the assessment of preparartions, I feel it is important to get advice from colleagues.	2	±0.93	2	1	4
Question 15: I was able to concentrate well enough during the provision of grades.	2	±0.50	1	1	3
Question 16: I was able to easily recognise incorrect representations (such as scan errors, incorrectly-placed preparation in a row.	2.5	±1.36	2	1	5
Question 17: The number of preparations in a row was just right.	1.12	±0.25	1	1	2
Question 18: I would have preferred to have assessed less preparations in a row.	3.87	±0.54	3.75	3	5
Question 19: The representation on the screen was easy to recognise.	1	±0.00	1	1	1
Question 20: The lighting was sufficient.	1	±0.00	1	1	1
Question 21: I would have preferred more time for the assessment of videos.	4.5	±0.58	4.5	4	5
Question 22: I would have preferred more time for the assessment of models.	4.5	±0.58	4.5	4	5
Question 23: The intervals were long enough.	2.5	±1.82	2.25	1	5
Question 24: On the whole, the procedure was very manageable.	1	±0.00	1	1	1

**Table 8:** Excerpt from the freely-composed commentaries of the examiners at the end of the evaluation questionnaire.

" I felt the experience agained...through the calibration helpful."
"The colour variation in the representation of the prepared and non-prepared part appears useful and helpful."
" I found the assessments partially very tiring."
"...The scenario is great for calibration puposes! Would be great for the train-the-teacher events!"
"... the assessment could have been more effective by using another sequence of the assessment criteria"
"...Well prepared(setting), comfortable atmosphere."
"I found the representation of the preparation edges approximate to using the PC particularly helpful!
"I was able to easily reconise frequent sloping i.e., errors in the secondary preparation!"

purely video-based assessment appears far more suitable. For an OSPE, this would mean that one would save on further dental personnel during the examination and could reach a consensus on the grades through four examiners and by means of videos, in considerably less time after the OSPE. This would, however, require the cavities to be scanned-in prior to the assessment. The extent to which the situation of the inlay preparation could also be applied to other examination material, such as, for example, the provision of fillings, would have to be discussed in further studies. In dissent situations surrounding the assessment, digital models (scans) could be useful, particularly, where the assessed working steps would have to be 'hidden' throughout the course of an examination. This occurs, for example, when during the application of a

two different settings of video-based assessments of cavity preparation performance using predefined checklists.

The reliability of this study lay at an average of  $\alpha=0.79$  (Part B: study group) and at  $\alpha=0.83$  (Part A: control group). In other literature, one can find reliability values in the form of Cronbach's  $\alpha$  of around 0.5 for examinations using CAD systems [24-26]. The reliability for OSPE without CAD systems, on the other hand, is depicted between the range of  $\alpha=0.68$  and  $\alpha=0.87$  [1, 10, 27]. The current experimental study, thereby, is more closely aligned with the results of these latter results. Due to the reliability values determined, the setting of Part A could be applied to 'high-stake' examinations. Part B only lies slightly below the value of  $\alpha = 0.8$  and requires an additional assessment step beyond the models. Thus, the



**Figure 8:** Four examiners (viewed from behind) watch the video film in Part A of the study on a screen on which the beamer projects the presentation in a darkened room. After 120 seconds, there is a brief interval (1 to 2 secs) until the next video film is played.



**Figure 9:** Four examiners (viewed from above) make an assessment in Part B of the study. They wear enlargement glasses and evaluate independently to one another. A countdown monitor runs in the background of the screen, so that they always have the maximum provided duration of 120 seconds in view. At the end of this time, an acoustic signal sounds and the next model is displayed for assessment.

restoration, the preceding preparation is 'hidden' before the end of an under-filling or at the end of a filling.

As a result, participants, on average, assessed Part A ('control group' with prepCheck video) with lower values (i.e. better grades) than Part B ('study group' with prepCheck video + consequent model). The overall grade of the control group ended up being 0.39 lower than in the study group. Here, it was interesting to see that the results of the study group merely deviated by 0.07 grade points from the overall grade in the real, live-performed OSPE. Thus, this setting appeared to depict the examination situation most clearly. This can be explained by the fact that the assessment of the live OSPE is equally performed with the aid of models, which resulted in procedural uniformity being detected here.

In studies on video-based examinations, some reliability data is provided in the form of ICC (interclass correlation coefficients) values. LAEEQ, CHEN and SCAFFIDI report an ICC of 0.62 [12, 14, 15]. KATEEB ICC values of  $0.47 \leq r \leq 0.78$  [28], provided in publications on CAD systems; inter-rater correlations of  $0.17 \leq r \leq 0.56$  are mentioned by ESSER [29]. URBANKOVA determined an ICC value of  $0.69 \leq r \leq 0.90$  [30]. This experimental study is therefore closest aligned to ESSER [29], KATEEB [28], LAEEQ and CHEN [14, 15]. The statement by SAMPAIOFERNANDES that there is a lot of deviation between individually implemented examiners [31], is in any case applicable, which also occurs in this study, regardless of whether this problem was tried to be counteracted through the train-the-teacher events. The effects of the training were less than optimal, however, so that a greater need for more information and practice would have been required above all concerning the parameters 'slide-in direction' and 'outer contact positioning'. The fact that the outer contacting positioning correlated to low ICC values within the control group, i.e. that were therefore exclusively assessed on grounds of the prepCheck videos, is not surprising. For, in inlay preparations, the outer contact positioning is conceivable, due to the given extension surfaces

and therefore the relation to more difficult conditions for scanning the cavities. These areas would certainly be easier to demonstrate in full-crown preparations. Were additional models provided for the assessment, the ICC values doubled, as the scanning no longer played a role here and one could assess the outer contact positioning i.e. correct the assessment, better. Here, the software would have to be improved on part of the manufacturer. In addition, a significant increase of Cronbach's alpha occurred in Setting B, when the 'outer contact positioning' was evaluated. This is also not surprising, as one was in the position of assessing these areas more carefully on the model. Appropriately, the study participants assessed the possibility of being able to assess the approximate outer contact positioning via prepCheck at a mean average of  $4.12 \pm 0.54$ . The assessment of the form of the cavity edge and slide-in direction, on the other hand, however, appear to represent clear indications of the analysis software. The study participants identified a further advantage of the examined analysis tool through the process of the calibration of their colleagues and perceived each of their applications with a mean of  $1.87 \pm 0.95$ . It is generally regarded as fundamentally important, however, to primarily perform the assessment by use of the analysis tool for examinations ( $3.00 \pm 1.41$ ). It is not surprising, that it is generally agreed that "dental assistants cannot be replaced by prepCheck when assessing cavities" ( $1.00 \pm 0.00$ ). For, the sole use of digital analysis tools in the current valid version alone, may require critical parameters in the grading, such as, for example, to insufficiently depict an image of the outer contact positioning. The overall assessment of the prepCheck analysis tool, ended up being a rather modest at  $2.87 \pm 0.89$  (on a Likert scale of 1 = excellent to 6 = unsatisfactory) and points to the above-mentioned problematic areas that can certainly be optimised on part of the software.

In order to reduce the limitations of the study, various points were considered. First of all, the order of the displayed videos and models was randomised by means of an online randomiser. As the variable of the experimental parts, i.e. examination teeth, was independent to the participants, a 'selection effect' did not take place. Secondly, the study took place with the same four study participants in both parts, at the same time (13:20) and in the same time frame (approx. two hours and 27 minutes), using the same procedure in the same rooms. Thirdly, the lighting of both of these settings were equally also the same, as well as also the duration of the videos (2 mins 0-10 secs) and the sequence of the settings portrayed in the individual films. Furthermore, it was taken into consideration that the participants were selected from the trainers of the department of operative dentistry, who were already actively taking part in practical preparation exercises (phantoms course for the study of conservative dentistry) while in their sixth semester of study and proved to have assessment experience. In order to reduce the problematics of the lack of realistic representation, it was attempted to perform the study in such a way that it reflected the circumstances of examination as closely as possible. In this way, the duration of the live-assessment of a cavity preparation was determined in preliminary studies and the assessment questionnaires compared to the checklists familiar from the examinations



[1, 27]. In order to eliminate the problem of generalisation occurring through the differing teaching experience, it was attempted to calibrate the assessment of the cavities in the preceding train-the-teacher events. Despite this, the following limitations should be taken into consideration: it is conceivable that when assessing a model (Part B), the evaluation was generally more rigid, as the preliminary grades from after the first part were already known. It is also possible that in scope of the whole experimental part, a practice-effect took place that became evident to each individual assessor to a different degree. This could explain why, despite the preceding train-the-teacher events, the inter-rater reliability differed. The influence of gender, age and teaching experience of the subject group was not a main part of this examination, although it could well be addressed in future studies.

## Conclusion

1. This examination illustrates an average reliability of  $\alpha = 0.833$  in the assessment mode control group (Part A) that supersedes the demands for practical examinations ( $\alpha \geq 0.6$ ) and also encompasses the general requirements of 'highstake' examinations of  $\alpha \geq 0.8$ . In Part B, a reliability of  $\alpha = 0.797$  was determined, without this being of specific significance to the control group.
2. The overall assessment did not significantly differ between both examination groups (Parts A and B). In the 'outer contact positioning' parameter, however, significant differences could be determined between A and B.
3. The ICC values with a mean average of  $0.43 < r < 0.74$  for the control group assessment mode (Part A) are higher than in the study group assessment mode (Part B) with  $0.35 < r < 0.60$ . The ICC values of the 'slide-in direction' and 'outer contact positioning' criteria of the assessment mode of the control group (Part A) are minimal. The maximum reliability of the criteria of 'preparation edge', 'surface', 'width & depth' and 'outer contact positioning' in the assessment mode of the control group (prepCheck video) is acceptable at  $\alpha > 0.7$ .
4. The assessment of the study participants in regards to the application concept of the digital-analytic software and study procedure generally proved to demonstrate a positive tendency.

## References

1. Schmitt L, Möltner A, Rüttermann S, Gerhardt-Szep S (2016) Study on the Interrater Reliability of an OSPE (Objective Structured Practical Examination) – Subject to the Evaluation Mode in the Phantom Course of Operative Dentistry 33: 61. [Link: https://tinyurl.com/y7wtolnj](https://tinyurl.com/y7wtolnj)
2. Natkin E, Guild RE (1967) Evaluation of preclinical laboratory performance: a systematic study 31: 152-161.
3. Jasinevicius TR, Landers M, Nelson S, Urbankova A (2004) An evaluation of two dental simulation systems: virtual reality versus contemporary non-computer-assisted 68: 1151-1162. [Link: https://tinyurl.com/ybl4p2yr](https://tinyurl.com/ybl4p2yr)

4. Kournetas N, Jaeger B, Axmann D, Groten M, Lachmann S, et al. (2004) Assessing the reliability of a digital preparation assistant system used in dental education 68: 12281-234. [Link: https://tinyurl.com/y9run9ts](https://tinyurl.com/y9run9ts)
5. Cardoso JA, Barbosa C, Fernandes S, Silva CL, Pinho A (2006) Reducing subjectivity in the evaluation of pre-clinical dental preparations for fixed prosthodontics using the Kavo PrepAssistant 10: 149-156. [Link: https://tinyurl.com/y7t2w4xh](https://tinyurl.com/y7t2w4xh)
6. Kenneth AE (2004) E-learning—new technologies but slow progress 7: 115-117.
7. Kikuchi H, Ikeda M, Araki K (2013) Evaluation of a virtual reality simulation system for porcelain fused to metal crown preparation at Tokyo Medical and Dental University. 77: 782-792. [Link: https://tinyurl.com/y834vzrx](https://tinyurl.com/y834vzrx)
8. Taylor CL, Grey NJ, Satterthwaite JD (2013) A comparison of grades awarded by peer assessment, faculty and a digital scanning device in a pre-clinical operative skills course 17: 16-21. [Link: https://tinyurl.com/y9wwcco2](https://tinyurl.com/y9wwcco2)
9. Turnbull J, Gray J, MacFadyen (1998) Improving in-training evaluation programs. J Gen Intern Med 3: 317-323. [Link: https://tinyurl.com/ya3tfd5g](https://tinyurl.com/ya3tfd5g)
10. Gerhardt-Szep S, Güntsch A, Pospiech P, Söhnel A, Scheutzel P, et al. (2016) Assessment formats in dental medicine: An overview. GMS J Med Educ 33: Doc65. [Link: https://tinyurl.com/yd5lb8v2](https://tinyurl.com/yd5lb8v2)
11. Zia A, Sharma Y, Bettadapura V (2016) Automated video-based assessment of surgical skills for training and evaluation in medical schools. Int J CARS 11: 1623-1636. [Link: https://tinyurl.com/yddz5rfx](https://tinyurl.com/yddz5rfx)
12. Scaffidi MA, Grover SC, Carnahan H, Yu JJ, Yong E, et al. (2017) A prospective comparison of live and video-based assessments of colonoscopy performance. [Link: https://tinyurl.com/yc5gjs0j](https://tinyurl.com/yc5gjs0j)
13. MacLuskey M, Durham J, Balmer C, Bell A, Cowpe J, et al. (2011) Dental student suturing skills: a multicentre trial of a checklist-based assessment. Eur J Dent Educ 15: 224-229. [Link: https://tinyurl.com/y8ta8dg7](https://tinyurl.com/y8ta8dg7)
14. Chen AC, Lee MS, Chen WJ, Lee ST (2013) Assessment in orthopedic training—an analysis of rating consistency by using an objective structured examination video. J Surg Educ 70: 189-192. [Link: https://tinyurl.com/ybykh3rc](https://tinyurl.com/ybykh3rc)
15. Laeeq K, Infusino S, Lin SY, Reh DD, Ishii M, et al. (2010) Video-based assessment of operative competency in endoscopic sinus surgery. Am J Rhinol Allergy. May-Jun; 24(3):234-7. [Link: https://tinyurl.com/yauyald6](https://tinyurl.com/yauyald6)
16. Podsakoff NP, Podsakoff PM, Mackenzie SB, Klinger RL (2013) Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures. J Appl Psychol. [Link: https://tinyurl.com/ybk8xv7e](https://tinyurl.com/ybk8xv7e)
17. Sarkiss CA, Philemond S, Lee J, Sobotka S, Holloway TD, et al. (2016) Neurosurgical Skills Assessment: Measuring Technical Proficiency in Neurosurgery Residents Through Intraoperative Video Evaluations. World Neurosurg 89: 1-8. [Link: https://tinyurl.com/y8znzezo](https://tinyurl.com/y8znzezo)
18. Perron NJ, Louis-Simonet M, Cerutti B, Pfarwallner E, Sommer J, et al. (2016) Feedback in formative OSCEs: comparison between direct observation and videobased formats. Med Educ Online 21: 32160. [Link: https://tinyurl.com/yb7d9a45](https://tinyurl.com/yb7d9a45)
19. Takazawa S, Ishimaru T, Harada K (2015) Video-Based Skill Assessment of Endoscopic Suturing in a Pediatric Chest Model and a Box Trainer. J Laparoendosc Adv Surg Tech A 25: 445-453. [Link: https://tinyurl.com/y6uwuhuy7](https://tinyurl.com/y6uwuhuy7)
20. Massey D, Byrne J, Higgins N, Weeks B, Shuker MA, et al. (2017) Enhancing OSCE preparedness with video exemplars in undergraduate nursing students. A mixed method study. Nurse Educ Today: 54: 56-61. [Link: https://tinyurl.com/y7v9jcgj](https://tinyurl.com/y7v9jcgj)

21. Deie K, Ishimaru T, Takazawa S, Harada K, Sugita N, et al. (2017) Preliminary Study of Video-Based Pediatric Endoscopic Surgical Skill Assessment Using a Neonatal Esophageal Atresia/Tracheoesophageal Fistula Model. *J Laparoendosc Adv Surg Tech A* 27: 76-81. [Link: https://tinyurl.com/y8skg2b6](https://tinyurl.com/y8skg2b6)
22. Simpson D, Helm R, Drewniak T, Ziebert MM, Brown D, et al. (2006) Objective Structured Video Examinations (OSVEs) for Geriatrics Education. *Gerontol Geriatr Educ* 26: 7-24. [Link: https://tinyurl.com/ycul69kp](https://tinyurl.com/ycul69kp)
23. Pérez-Escamirosa F, Chousleb-Kalach A, del Carmen Hernández-Baro M, Sánchez-Margallo JA, Lorias-Espinoza D, et al. (2016) Construct validity of a video-tracking system based on orthogonal cameras approach for objective assessment of laparoscopic skills. *Int J CARS* 11: 2283–2293. [Link: https://tinyurl.com/yb3wz2vq](https://tinyurl.com/yb3wz2vq)
24. Weigl P, Felber R, Brandt J, König E, Lauer HC (2015) Fully automated and objective quality inspection of a clinical tooth preparation. *Proceedings 40th Annual Meeting of the Association for Dental Education in Europe (ADEE)*. *Eur J Dent Educ* 19: e8–e35.
25. Stumpf A, Weigl P, Gerhardt T, Felber R, Heidemann D, Gerhardt-Szép (2015) Computer-aided 3D analysis of cavities via prepCheck - a pilot study. *Proceedings 40th Annual Meeting of the Association for Dental Education in Europe (ADEE)*. *Eur J Dent Educ* 19: e8–e35. [Link: https://tinyurl.com/yabr9pub](https://tinyurl.com/yabr9pub)
26. Roopa VA (2017) The Calibration of a Software Programme to Assess Ceramic Crown Preparations in a Pre-clinical Setting. [Link: https://tinyurl.com/y9t7n2un](https://tinyurl.com/y9t7n2un)
27. Petkov P, Knuth-Herzig K, Hoefer S, Stehle S, Scherer S, et al. (2017) The reliability and predictive validity of a sixth-semester OSPE in conservative dentistry regarding performance on the state examination. *GMS J Med Educ* 34. [Link: https://tinyurl.com/y9gu4jqj](https://tinyurl.com/y9gu4jqj)
28. Kateeb ET, Kamal MS, Kadamani AM, Abu Hantash RO, Arqoub MM (2016) Utilising an innovative digital software to grade pre-clinical crown preparation exercise. *Eur J Dent Educ*. [Link: https://tinyurl.com/ydz8oc3d](https://tinyurl.com/ydz8oc3d)
29. Esser C, Kerschbaum T, Winkelmann V, Krage T, Faber FJ (2006) A comparison of the visual and technical assessment of preparations made by dental students. *Eur J Dent Educ* 10: 157-161. [Link: https://tinyurl.com/y77ecmm7](https://tinyurl.com/y77ecmm7)
30. Urbankova A (2010) Impact of computerized dental simulation training on preclinical operative dentistry examination scores. *J Dent Educ* 74: 402-409. [Link: https://tinyurl.com/y7gqazv](https://tinyurl.com/y7gqazv)
31. Sampaio-Fernandes MA, Sampaio-Fernandes MM, Fonseca PA, Almeida PR, Reis-Campos JC, et al. (2015) Evaluation of occlusal rest seats with 3D technology in dental education. *J Dent Educ* 79: 166-176. [Link: https://tinyurl.com/ycvykecy](https://tinyurl.com/ycvykecy)