

**Mehrwert eines digitalen Analysesystems für die Bewertung von
Zahnpräparationen für eine vollverblendete Krone in der vorklinischen
Ausbildung**

Von der Medizinischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule
Aachen zur Erlangung des akademischen Grades einer Doktorin der Zahnmedizin
genehmigte Dissertation

vorgelegt von

Katharina Svenja Heinen, geborene Merkel

aus Würselen

Berichter: Universitätsprofessor Dr. med. dent. Stefan Wolfart
Universitätsprofessor Dr. med. dent. Michael Wolf

Tag der mündlichen Prüfung: 01.12.2025

***Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online
verfügbar.***

Widmung

Diese Arbeit widme ich Max.

Inhaltsverzeichnis

Abkürzungsverzeichnis	
Einleitung	1
Grundlagen zur Leistungsbeurteilung	2
Qualitative Beurteilung von Leistungsüberprüfungen	2
Beurteilung praktischer Fertigkeiten von vorklinischen Zahnmedizinierenden	4
Bewertungsmethoden	4
Evaluierung der gängigen Bewertungsmethoden	4
Einflussfaktor Notenskala	6
Einflussfaktor Rater beziehungsweise Ratertraining	6
Computer-aided Learning im Zahnmedizinistudium	8
Einbindung von Computer-aided Learning (CAL) in den Kursalltag	9
Zahnpräparation	11
Präparationsrichtlinien	11
Richtlinien für festsitzende metallkeramische Restaurationen	14
Kontrolle der Präparation	14
Computergestützte Zahnmedizin	15
Zielsetzung	17
Material und Methoden	19
Studiendesign	19
Ethische Aspekte	19
Probanden	19
Eingeschlossene Präparationen	19
Bewertungsbogen	20
Analysesoftware	22
Analyseparameter	23
Ablauf	28
Erste Sitzung (Praetest)	28
PrepCheck-Schulung	29
Vorbereitung für zweite Sitzung	29
Zweite Sitzung (Posttest)	30
Statistische Auswertung	33
Ergebnisse	35
Untersuchungen zum Praetest	35
Untersuchungen zum Posttest	39
Unterschiede zwischen erstem und zweitem Durchgang	41
Diskussion	50
Diskussion der Methode	50
Analysesoftware prepCheck	52
Umgang mit der CEREC Omnicam	54
Diskussion der Ergebnisse	56
Interrater-Reliabilität	56
Intrarater-Reliabilität	58

Einfluss der Analysesoftware prepCheck auf die Beurteilung	61
Implementierbarkeit in den Kursalltag	62
Stärken der Studie	63
Limitationen der Studie	63
Schlussfolgerungen	64
Zusammenfassung	65
Summary	66
Abbildungsverzeichnis	67
Tabellenverzeichnis	68
Literatur	69
Danksagung	74
Erklärung § 5 Abs. 1 zur Datenaufbewahrung	75
Eidesstattliche Erklärung gemäß § 5 Abs. (1) und § 11 Abs. (3) 12. der Promotionsordnung	76

Abkürzungsverzeichnis

ZAppro	Zahnärztliche Approbationsordnung
CAL	Computer-aided Learning
OSCE	Objective Structured Clinical Examination
prepCheck	Name einer Analysesoftware
CEREC	Chairside Economical Restoration of Esthetic Ceramics (computergestütztes System von Sirona)
CAD/CAM	Computer-Aided Design/Computer- Aided Manufacturing
SPSS	Statistical Package for the Social Sciences (Software zur statistischen Analyse)

Einleitung

Das Zahnmedizinstudium soll die Studierenden zu kompetenten Zahnärzten auszubilden, die nicht nur über das theoretische Fachwissen, sondern auch über die nötigen manuellen Fertigkeiten verfügen, um Patienten adäquat beraten und behandeln zu können. Studierende der Zahnmedizin führen im klinischen Studienabschnitt bereits eigenständig Behandlungen am realen Patienten durch. Die entsprechenden praktischen Fertigkeiten, die für diese eigenständigen Behandlungen beherrscht werden müssen, werden in dem vorklinischen Studienabschnitt an Phantompatienten vermittelt. Die Studierenden erlernen hierfür im Studienabschnitt I und II des Zahnmedizinstudiums (ZAppro 2019) in einem simulierten klinischen Behandlungsumfeld unter anderem die Präparation von Pfeilerzähnen für festsitzende prothetische Restaurationen.

Um gewährleisten zu können, dass die Studierenden über die Kompetenz zur Patientenbehandlung verfügen, muss der Lernfortschritt bei der Ausübung der simulierten Patientenbehandlung fortwährend überprüft werden. Die Beurteilung erfolgt in der Regel durch die Mitarbeiter der jeweiligen Klinik. Bei dieser Form der Leistungsbeurteilung kann es durch einen Mangel an Objektivität, Reproduzierbarkeit und Transparenz zu Verwirrung und Frustration unter den Studierenden kommen (Sharaf, AbdelAziz et al. 2007, Renne, McGill et al. 2013).

Diese Problematik ist schon lange bekannt: Bereits seit den 1970er-Jahren findet die Leistungsbeurteilung in der vorklinischen Ausbildung von Zahnmedizinierenden Aufmerksamkeit in der Literatur (Natkin and Guild 1967, Fuller 1972, Dhuru, Rypel et al. 1978).

So muss das wissenschaftliche Ziel sein, die Art der Benotung in den praktischen vorklinischen Kursen des Zahnmedizinstudiums zu verbessern und eine gesteigerte Reliabilität und Objektivität zu erreichen.

Grundlagen zur Leistungsbeurteilung

Im Rahmen einer Leistungsbeurteilung wird der Lernfortschritt, also Menge, Niveau und Qualität der Lernergebnisse, bewertet. Die Art der Überprüfung und Bewertung beeinflusst das künftige Lernen und Handeln der Studierenden in erheblichem Maße. Leistungsbeurteilungen dienen als Feedbackinstrumente, sie sollen Stärken und Schwächen aufzeigen und den Fokus der Studierenden auf bestimmte Aspekte ihres Handelns leiten, um sie zu motivieren, sich vor dem Hintergrund der Beurteilung gezielt weiterzuentwickeln (Fritschi and Roth 2014).

Es gibt zwei Arten der Beurteilung, die zu unterscheiden sind und die verschiedene Funktionen erfüllen: Die formative- und die summative Beurteilung (Orsmond, Merry et al. 2000).

Ziel der formativen Beurteilung ist es, laufend Informationen zu Lernfortschritt, Stärken und Schwächen zu sammeln und über regelmäßige Rückmeldungen an den Studierenden dessen Verbesserung zu erzielen. Diese Form des kursbegleitenden Feedbacks funktioniert nur dann, wenn der Lernende aufmerksam und motiviert ist, wenn er das Feedback einordnen und auf sich beziehen kann und wenn er es richtig interpretieren kann, das heißt, wenn er über genügend fachliche Vorkenntnisse verfügt.

Diese Art der Beurteilung zielt nicht auf formale Anerkennung, sondern auf eine Unterstützung des Lernprozesses ab.

Ergänzend dazu bedarf es aber ebenso summativer Beurteilung der studentischen Arbeiten: Hierbei erfolgt eine Bewertung des Wissensbeziehungsweise des Könnensstands und der Fortschritt der Studierenden wird formal bewertet und in der Regel in Form einer Note festgehalten (Fritschi and Roth 2014). Hier steht, anders als bei der formativen Beurteilung, nicht der Prozess des Lernens, sondern der Erfolg im Fokus. Es geht hierbei mehr um Bewertung und Selektion als um Beratung.

Qualitative Beurteilung von Leistungsüberprüfungen

Die Qualität einer Bewertung wird an den drei Hauptgütekriterien Objektivität, Reliabilität und Gültigkeit gemessen.

Die Objektivität zeigt an, inwieweit eine Bewertung unabhängig ist von der beurteilenden Person. Ein Test ist demnach objektiv, wenn verschiedene Prüfer gleiche Ergebnisse erhalten. Man unterscheidet Durchführungs-, Auswertungs- und Interpretationsobjektivität. Um eine Durchführungsobjektivität zu gewährleisten, ist eine einheitliche Prüfungssituation für alle Prüflinge zu schaffen. Dazu zählen beispielsweise die Aufgabenstellung, die zur Verfügung stehenden Hilfsmittel und die Dauer der Prüfung (Ingenkamp and Lissmann 2008). Beispielhaft hierfür sind moderne Prüfungstechniken wie objective structured clinical examination (OSCE) zu nennen.

Auswertungsobjektivität bedeutet, dass die Auswertung prüferunabhängig abläuft, das heißt, es wird anhand von einem Lösungsschema oder ähnlichem vorgegangen (Jürgens and Sacher 2008).

Die Komponente der Interpretationsobjektivität ist erfüllt, wenn die ausgewerteten Ergebnisse von verschiedenen Prüfern identisch interpretiert werden, zum Beispiel durch einen festen Bezugsrahmen (Jürgens and Sacher 2008).

Die Reliabilität, auch Zuverlässigkeit, ist dann gegeben, wenn ein Test bei mehrmaliger Wiederholung dasselbe Ergebnis liefert und damit frei von Messfehlern ist. Sie bestimmt also, wie präzise und reproduzierbar ein Merkmal gemessen wird (Pospeschill 2010).

Die Gültigkeit, auch Validität, bestimmt, in welchem Maß der Test das misst, was tatsächlich zu messen ist, also auf die Lernziele abgestimmt ist (Bortz 2006).

In der Forschung gibt es verschiedene Kenngrößen, um die Qualität einer Prüfung zu bestimmen. Üblicherweise wird sie anhand von Urteilsübereinstimmungen gemessen. Hier kann man beispielsweise zwischen inter- beziehungsweise intrapersoneller Übereinstimmung unterscheiden, auch Inter- beziehungsweise Intrarater-Reliabilität genannt.

Unter Interrater-Reliabilität versteht man die Urteilsübereinstimmung mehrerer verschiedener Rater über ein Objekt, während Intrarater-Reliabilität die Übereinstimmung mehrerer Bewertungen desselben Objekts durch denselben Rater bezeichnet.

Übereinstimmungen kann man mithilfe statistischer Verfahren ermitteln.

Beurteilung praktischer Fertigkeiten von vorklinischen Zahnmedizinstudierenden

Die manuellen Fertigkeiten der Studierenden werden in aller Regel im Rahmen von Präparationstestaten durch das zahnärztliche Personal der Kliniken beurteilt. Die lange Zeit weit verbreitete Methode „glance and grade“ (Jenkins, Dummer et al. 1998) erfolgt per Inaugenscheinnahme durch den Beurteiler und kann den oben genannten Gütekriterien für eine Leistungsbeurteilung nicht gerecht werden. Es handelt sich hierbei um eine seit langer Zeit bekannte Problematik. So wird bereits seit über 40 Jahren an einer Optimierung der Prüfungsbeurteilungen im vorklinischen Abschnitt des Zahnmedizinstudiums gearbeitet.

Bewertungsmethoden

Im Wesentlichen kommen inzwischen die drei Bewertungsmethoden „glance and grade“, „Checkliste“ oder „Definierte Beurteilungskriterien“ bei der Bewertung präparierter Zähne zum Einsatz (Vann, Machen et al. 1983, Schiffler 2007).

Bei der „glance and grade“-Bewertungsmethode basiert das Ergebnis auf Erfahrungswerten und persönlichen Schwerpunkten des Prüfers (Lilley, ten Bruggen Cate et al. 1968, Fuller 1972, Gaines, Bruggers et al. 1974, Feil 1982).

Eine Checkliste dient als Leitfaden und beinhaltet die für die Bewertung relevanten Teilaspekte, sodass die Prüfenden bei der Festlegung ihres Urteils keine essenziellen Aspekte außer Acht lassen.

Durch eine Erweiterung einer einfachen Checkliste um definierte Beurteilungskriterien erhält man einen detaillierten Bewertungsbogen, der für jeden Teilaspekt abgestufte Bewertungen mit assoziierten Kriterien vorgibt. So sind bestimmte begangene Fehler mit einem definierten Notenniveau verknüpft, um den Ermessensspielraum bei der Bewertung zu minimieren.

Evaluierung der gängigen Bewertungsmethoden

In einigen Studien zur Reliabilität von Beurteilungen praktischer studentischer Fertigkeiten konnte gezeigt werden, dass die Benotung per Augenschein

unzureichende Maße an Objektivität und Reproduzierbarkeit liefert (Sharaf, AbdelAziz et al. 2007, Renne, McGill et al. 2013). Es konnten Abweichungen sowohl zwischen den Bewertungen verschiedener Prüfer (Interrater) als auch zwischen denen desselben Prüfers zu verschiedenen Zeitpunkten (Intrarater) nachgewiesen werden (Lilley, ten Bruggen Cate et al. 1968, Fuller 1972, Jenkins, Dummer et al. 1998, Renne, McGill et al. 2013, Schepke, van Wulfften Palthe et al. 2020). Die Variabilität innerhalb desselben Prüfers fällt allerdings einigen Studien zufolge geringer aus als die zwischen verschiedenen Prüfern (Lilley, ten Bruggen Cate et al. 1968), bzw. ist gar nicht in signifikantem Maß nachzuweisen (Sharaf, AbdelAziz et al. 2007).

Diese unzuverlässigen, abweichenden Bewertungen führen zu Verwirrung und Frust unter den Studierenden (Natkin and Guild 1967). Im Kursalltag registrieren sie zügig, welche Betreuer eher milde oder streng bewerten. Der Fokus verschiebt sich durch die Subjektivität bei der Beurteilung hauptsächlich auf die Erreichung eines Testats beziehungsweise einer guten Note. Gleichzeitig sinkt die Lernmotivation durch das Gefühl von Willkür und möglicher Bevorzugung beziehungsweise Diskriminierung einzelner Studierender (Renne, McGill et al. 2013).

Davon ausgehend wurden verschiedene Ansätze verfolgt, um die Objektivität und Reliabilität bei der Beurteilung praktischer studentischer Arbeiten zu steigern. Beispielsweise versuchte man durch die Definition exakter Beurteilungskriterien eine Reliabilitätssteigerung zu erzielen. Dieser Ansatz führte zu verschiedenen, teils widersprüchlichen Ergebnissen:

In einigen Studien konnten durch die Methode mit Checkliste und Kriterien deutliche Erfolge verzeichnet werden (Gaines, Bruggers et al. 1974, Dhuru, Rypel et al. 1978, Goepferd and Kerber 1980, Bedi, Lo et al. 1987, Schmitt, Möltnner et al. 2016). Vor allem unter Beurteilenden mit wenig Berufserfahrung konnte eine höhere Urteilsübereinstimmung erzielt werden (Bedi, Lo et al. 1987). Andere Untersuchungen mit strukturiertem Bewertungsbogen berichten jedoch, keine Reliabilitätssteigerung erreicht zu haben (Fuller 1972, Vann, Machen et al. 1983, Sharaf, AbdelAziz et al. 2007).

Einflussfaktor Notenskala

Ein ebenfalls diskutierter Punkt in der Bewertung mittels Beurteilungskriterien ist die Gestaltung der Notenskala. Houpt und Kress untersuchten in diesem Zusammenhang, ob die Anzahl und Definition möglicher Notenabstufungen die Urteilsübereinstimmung beeinflussen (Haupt and Kress 1971). Verglichen wurden ein 2-Punkte- sowie ein 5-Punkte-Bewertungsbogen. Die Bewertung mit dem 2-Punkte-Bogen lieferte eine bessere Übereinstimmung. Es ist jedoch zu berücksichtigen, dass dies schon allein durch die verminderte Anzahl an Möglichkeiten bedingt sein kann und dass eine reine Entscheidung zwischen „erfüllt“ und „nicht erfüllt“ für praktische Fertigkeiten oft nicht getroffen werden kann.

Einflussfaktor Rater beziehungsweise Ratertraining

Sicherlich hat auch der Bewertende selbst einen großen Einfluss auf das Bewertungsergebnis. Hier konnte zum Beispiel ein größeres Maß an Berufserfahrung als positiver Einfluss auf die Reliabilität nachgewiesen werden (Dhuru, Rypel et al. 1978, Jenkins, Dummer et al. 1998). Dhuru et al. stellten insgesamt die geringste Reliabilität für die unerfahrenen Bewerter ohne Checkliste und die höchste Reliabilität für die erfahrenen Bewerter mit Checkliste fest. Um den Einfluss der Bewertenden zu minimieren, wird deshalb eine Kalibrierung der Mitarbeiter angestrebt. Nicht in allen Untersuchungen war dieser Versuch der Kalibrierung erfolgreich. Während eine Studie von Fuller, in der die Prüfer ein zweistündiges Training mit anschließender Diskussionsrunde absolvierten, keinen Erfolg zeigte (Fuller 1972), waren Diskussionstreffen nach einer Untersuchung von Natkin & Guild wohl erfolgreich (Natkin and Guild 1967). Auch in einer Untersuchung von Haj-Ali und Feil konnte gezeigt werden, dass sich die Interrater-Reliabilität durch eine Kalibrierung des Personals erhöht und dass der Effekt der Kalibrierung zudem auch nach einer Zeitspanne von zehn Wochen noch anhält (Haj-Ali and Feil 2006). Hier wurden die Präparationen dreimal durch die Prüfer beurteilt: Vor der Kalibrierung, unmittelbar nach der Kalibrierung und schließlich zehn Wochen nach der Kalibrierung. Die Kalibrierung umfasste eine Powerpoint-Präsentation, in der zu jedem Bewertungskriterium das Ideal bildlich demonstriert und von den Teilnehmern diskutiert wurde.

Computer-aided Learning im Zahnmedizinstudium

Im Klinikalltag ist es oft nicht realisierbar, allen Gütekriterien der Leistungsbewertung gerecht zu werden. Mithilfe der digitalen Technik wäre es heutzutage vielleicht aber möglich, nah an das angestrebte Optimum heranzukommen und der Forderung nachzukommen, die menschliche Komponente aus der Leistungsbeurteilung zu eliminieren.

Es gibt bereits seit längerer Zeit Geräte einiger Hersteller auf dem Markt, die für die Anwendung im Zahnmedizinstudium entwickelt wurden.

PREPassistant bzw. Dental Teacher (KaVo, Biberach, Deutschland)

Der PREPassistant ermöglicht die digitale Darstellung und Bewertung einer Präparation am Modellzahn im Vergleich zu einer Idealpräparation. Nach dem Scanvorgang, für den der einzelne Zahn in einem bestimmten Halter im Scanner positioniert wird, werden beide Präparationen einander gegenübergestellt und können sowohl in 3D als auch in zweidimensionalen Schnittbildern anhand von definierten Punkten, Strecken und Winkeln verglichen werden. Abweichungen vom Ideal werden numerisch angezeigt und werden zusätzlich durch eine überlagerte Darstellung der Präparationen veranschaulicht. Zudem wird auf der Basis zuvor festgelegter Toleranzbereiche ein Gesamturteil abgegeben (Arnetz and Dornhofer 2004). Untersuchungen zufolge liegen die größten Schwachstellen des Systems darin, dass die Beurteilung der Oberflächenbeschaffenheit sowie die Kontinuität des Präparationsrandes nicht mithilfe des PREPassistants erfolgen kann, sodass ein Teil einer umfassenden Präparationsbeurteilung weiterhin in der Hand des Lehrpersonals liegt (Kournetas, Jaeger et al. 2004, Cardoso, Barbosa et al. 2006).

Der PREPassistant wurde inzwischen von der Nachfolgesoftware Dental Teacher abgelöst, welche weiterhin Winkel und Strecken messen kann und sowohl Schnittbilder als auch dreidimensionale Überlagerungen von Präparationen mit der Masterpräparation und/ oder dem unpräparierten Zahn generieren (Nagy, Simon et al. 2018).

DentSim (Denx, Jerusalem, Israel)

DentSim wird in Verbindung mit einer Phantomkopf-Simulationseinheit verwendet. Dieses System veranschaulicht dem Studierenden nicht nur Fehler in der Präparation, sondern es gibt auch Feedback in Echtzeit, beziehungsweise der Prozess der Präparation wird computergestützt aufgezeichnet, sodass nachvollzogen werden kann, wie ein Fehler zustande gekommen ist (Welk, Splieth et al. 2004).

prepCheck (Dentsply Sirona, Bensheim, Deutschland)

prepCheck funktioniert als App für die mobile Computereinheit CEREC AC und wird in Verbindung mit einer der systemeigenen digitalen Kameras (CEREC Primescan oder Omnicam) eingesetzt. Zur Analyse werden die am Phantomkopf erstellte Präparation, der Antagonist sowie eine Bissnahme eingescannt. Mithilfe der App wird die Präparation im Anschluss entweder in Bezug zu einer Normpräparation mit durch den Lehrenden definierten Normwerten oder zu einer sogenannten Masterpräparation gesetzt und anhand dessen analysiert. Abweichungen werden farblich dargestellt. Ist eine Masterpräparation hinterlegt, so kann diese mit der zu beurteilenden Präparation überlagert dargestellt werden (Lenherr and Marinello 2014).

Einbindung von Computer-aided Learning (CAL) in den Kursalltag

Der Einsatz dieser Geräte sowie deren Einbindung in die vorklinischen Kurse wurden bereits getestet.

Die Einführung eines solchen digitalen Systems in den Lehrplan bedeutet zunächst einen zeitlichen Mehraufwand (Welk, Splieth et al. 2006), da sowohl Lehrpersonal als auch Studierende erst mit Geräten und Software vertraut gemacht werden müssen und das Erlernen des Scannens und der Anwendung zeitintensiv sein können (Wolgin, Grabowski et al. 2018, Schepke, van Wulfften Palthe et al. 2020). Die Implementierung der digitalen Lehre ist außerdem mit hohen Kosten verbunden. Langfristig verspricht man sich davon jedoch die objektive Beurteilung von studentischen Arbeiten in der Vorklinik, Kosteneinsparungen im Personalbereich und geringere Betreuungszeiten von

Studierenden durch Kurstutoren (Kournetas, Jaeger et al. 2004, Cardoso, Barbosa et al. 2006).

Laut der Metaanalyse von Rosenberg et al. kann computer-aided learning (CAL) in der Ausbildung von Zahnmedizinierenden zur Ergänzung der traditionellen Lehrmethoden oder zur Unterstützung der Studierenden beim Selbststudium erfolgreich zum Einsatz kommen, da es nachweislich genauso effektiv ist wie andere Unterrichtsmethoden (Rosenberg, Grad et al. 2003).

Hierbei hat sich gezeigt, dass CAL zum Beispiel im Bereich der Selbstbeurteilung sehr hilfreich für die Studierenden ist. Das sofortige und visualisierte Feedback ist gut nachzuvollziehen und führt unter den Studierenden zu einer gesteigerten Motivation, die Leistung zu verbessern sowie zu einem besseren Grundverständnis für die Lehrinhalte (Hamil, Mennito et al. 2014, Park, Sheinbaum et al. 2017). Für eine Studie der Universität in Krems (Österreich) wurden die Kavitätenpräparationen für eine Füllung von Studierenden verglichen, nachdem sich die eine Gruppe im Selbststudium mit Hilfe von prepCheck vorbereitet hatte und die andere Gruppe während der Vorbereitung traditionell von Kurstutoren betreut worden war. Es zeigte sich, dass die Ergebnisse beider Gruppen vergleichbar waren, sodass davon ausgegangen werden kann, dass eine Implementierung digitaler Analysesoftware funktionieren kann und vor allem in Hinblick auf steigende Studierendenzahlen eine hilfreiche Entlastung für das Klinikpersonal darstellen kann (Wolgin, Grabowski et al. 2018).

Fraglich ist jedoch, ob ein solches System auch zur alleinigen Leistungsbewertung herangezogen werden kann.

Zahnpräparation

Es stellt sich die Frage, warum die Leistungsbewertung von Zahnpräparationen in der vorklinischen Ausbildung ein so viel diskutiertes Thema ist.

Zähne, die aufgrund von kariösen Läsionen, Trauma oder anderen Ursachen, wie z.B. Bruxismus einen Hartschubstanzverlust erlitten haben, der eine direkte Restauration im Sinne einer Füllung nicht mehr zulässt, bedürfen einer indirekten Versorgung mit Inlay, Teilkronen oder Kronen - je nach Defektausdehnung und Defektlokalisation. Jede dieser Versorgungen erfordert zunächst die Präparation, also das Beschleifen des jeweiligen Zahnes in einer auf die spätere Restauration angepassten Weise. Damit macht die Präparation von Pfeilerzähnen ein Schwerpunktgebiet der täglichen Arbeit eines Zahnarztes aus, und ist deshalb auch eine der im Zahnmedizinstudium zu erlernenden Kernkompetenzen (Karl, Graef et al. 2011).

Außerdem ist die Präparationstechnik neben den Materialeigenschaften von entscheidender Wichtigkeit für den langfristigen Erfolg restaurativer Maßnahmen (Kern, Wolfart et al. 2022) und sollte daher sicher beherrscht werden.

Die materialspezifischen Anforderungen stehen oft im Widerspruch zu den biologischen Anforderungen (Edelhoff and Sorensen 2002): Die Vitalerhaltung der Pulpa sowie die Schonung des Parodontiums sind essenziell, aber nur durch ausreichende Reduktion der Zahnhartsubstanz können vorgegebene Materialstärken bei anatomisch korrektem Kronendesign erzielt werden (Kern, Wolfart et al. 2022). Daher sollte jede zahnärztlich-prothetische Präparationsmaßnahme das Ziel verfolgen, eine möglichst geringe Traumatisierung von Zahn, Gingiva und Parodont zu verursachen, gleichzeitig jedoch eine Versorgung mit ästhetisch und funktionell befriedigendem Zahnersatz zu ermöglichen.

Präparationsrichtlinien

In der Literatur sind Präparationsrichtlinien zu finden, deren Einhaltung eine langfristig funktionierende Versorgung gewährleisten soll. Diese Richtlinien befinden sich jedoch, entsprechend der zur Verwendung kommenden dentalen Werkstoffe, im Wandel: So werden für Kronen aus Gold andere Materialstärken

benötigt als für Metallkeramik und wiederum andere für Vollkeramik. Entsprechende abweichende Anforderungen sind ebenfalls in Bezug auf Konizität, Oberflächenbeschaffenheit und Präparationsgrenze zu finden.

Einige Punkte sind jedoch unabhängig von den spezifischen Materialanforderungen zu beachten:

Schutz der Pulpa

Die Pfeilerzahnpräparation stellt im Zusammenhang mit der Überkronung den Hauptrisikofaktor für eine mögliche Pulpaschädigung dar (Kern, Wolfart et al. 2022). Dabei kann zum einen das Maß des Zahnhartsubstanzabtrags entscheidend sein, wobei das Risiko einer irreversiblen Pulpaschädigung mit zunehmendem Abtrag ansteigt. Wird eine Restdentinstärke von 1-2 mm eingehalten, so können Pulpaschäden in der Regel vermieden werden (Podhorsky, Rehmann et al. 2015).

Des Weiteren kann die Pulpa thermisch geschädigt werden. Daher ist eine ausreichende Wasserkühlung von 50 ml/min während des Präparationsvorgangs essenziell zur Vermeidung irreversibler Pulpaschädigungen. Außerdem ist eine erhöhte Reibungswärme durch eine zu hohe Andruckkraft oder durch stumpfe Instrumente zu vermeiden (Kern, Wolfart et al. 2022).

Ein weiteres Risiko für die Pulpa besteht in der möglichen Austrocknung. Daher sollte auf wiederholtes, intensives Trockenblasen während und nach der Präparation verzichtet werden (Matthews, Showman et al. 1993).

Schutz des marginalen Parodontiums

Die beste Möglichkeit zum Schutz des marginalen Parodontiums bietet die supragingivale Präparation bzw. ein supragingival zu liegen kommender Restaurationsrand. Aus ästhetischen Gründen kann jedoch eine subgingivale Präparation nötig sein. In diesen Fällen ist ein ausreichender Abstand zwischen Kronenrand und Knochen, die sogenannte „biologische Breite“, einzuhalten, sodass genügend Platz für die Ausbildung eines physiologischen Sulkus und epithelialen und bindegewebigen Attachments bleibt (Wilson and Maynard 1981).

Dazu und um möglichst atraumatisch zu arbeiten, sollte vor der Präparation ein dünner Gingivaretraktionsfaden eingelegt werden und die Präparation soll dem Verlauf des Gingivasaums folgen.

Retentions- und Widerstandsform

Um eine Verankerung der späteren Restauration gegen abziehende und extraaxiale Kräfte zu gewährleisten, sollte man bei Präparationen an einem Stumpf eine ausreichende Retentions- und Widerstandsform erzielen (Kern, Wolfart et al. 2022).

Diese Merkmale sind definiert über:

- Den Umfang des Stumpfes: Je größer der Umfang, desto geringer ist bei gleicher Höhe und Konizität des Stumpfes die Widerstandskraft gegen kippende Kräfte.
- Die Höhe des Stumpfes: Die Mindesthöhe der vertikalen Wände soll 3 mm betragen.
- Den Präparationswinkel: Der ideale Konvergenzwinkel der Vertikalflächen zur Zahnachse beträgt drei Grad.

Kriterien, die je nach Ersatzmaterial und Konstruktion variieren, sind beispielsweise:

- Der Substanzabtrag
- Die Präparationsform

Richtlinien für festsitzende metallkeramische Restaurationen

Da sich diese Studie mit der Untersuchung von Präparationen für eine metallkeramische Versorgung beschäftigt, sollen die Präparationskriterien für diese Art der Versorgung im Folgenden genauer erläutert werden.

- Okklusale Reduktion
 - Okklusal werden 1,5 mm Platz zum Antagonisten benötigt (Motta, Pereira et al. 2014).
- Zirkuläre Reduktion
 - Es wird ein Abtrag von 1,2 mm empfohlen (Hobo and Shillingburg 1973).
- Präparationsgrenze
 - Im sichtbaren Bereich wird eine Stufenpräparation von 1,0-1,2 mm und im nicht sichtbaren Bereich von 0,8 mm empfohlen (Al-Dwairi, Bashatwa et al. 2015).
- Oberflächenbeschaffenheit
 - Es sollen keine Riefen oder scharfen Kanten vorhanden sein (Amaya-Pajares, Ritter et al. 2016).

Kontrolle der Präparation

Zur Kontrolle ihrer Präparationen verwenden die Studierenden der RWTH Aachen aufgeschnittene Silikonschlüssel, die sie zuvor auf den unbeschliffenen Frasacomodellen herstellen. Mithilfe einer Pradodontalsonde lässt sich so der Substanzabtrag abschätzen.

Im Allgemeinen sollten Präparationen auch immer durch Überprüfung der Schichtstärken des eingeschliffenen Provisoriums beurteilt werden.

Computergestützte Zahnmedizin

Die computergestützte Zahnmedizin ist auf dem Vormarsch und hält Einzug in immer mehr Anwendungsgebiete (Rekow 2018). Elektronische Patientenakten und digitales Röntgen kommen schon lange zum Einsatz, Intraoralscanner werden immer populärer, die CAD/CAM-Fertigung (computer-aided design/computer-aided manufacturing) von Zahnersatz und auch der 3D-Druck erlangen einen immer höheren Bekanntheits- und Verwendungsgrad (Schlenz, Rehmann et al. 2019). Im Folgenden soll auf die CAD/CAM-Technologie genauer eingegangen werden. Sie gewann in der Zahnmedizin ab der Einführung des von Mörmann und Brandestini entwickelten CEREC-Systems (Dentsply Sirona, Bensheim, Deutschland) in den 1980er Jahren zunehmend an Bedeutung: Dieses System ermöglichte als erstes kommerziell vertriebenes seiner Art die chairside-Fertigung von vollkeramischem Zahnersatz an nur einem Tag mittels der sogenannten CAD/CAM-Technologie (Liu and Essig 2008). Der Begriff „CAD/CAM“ (computer-aided design/ computer-aided manufacturing) kommt ursprünglich aus dem technischen Maschinenbau und bezeichnet die computergestützte, dreidimensionale Planung eines Werkstücks und dessen anschließende automatische Herstellung (Tinschert, Natt et al. 2004).

In der Zahnmedizin bestehen CAD/CAM-Systeme in der Regel aus drei Komponenten (Tinschert, Natt et al. 2004, Beuer, Schweiger et al. 2008):

- Die erste Komponente ist eine Digitalisierungseinheit zur Generierung des digitalen Datensatzes, der im Anschluss computergestützt weiterverarbeitet wird. Ein präparierter Zahn und je nach Restaurationsart auch die Nachbarzähne, die antagonistische Zahnreihe oder auch der gesamte Zahnbogen werden mittels Intraoralscanner digital abgeformt. So können die konventionelle Abformung umgangen, und mit ihr verknüpfte Fehlerquellen eliminiert werden.
- Als zweite Komponente folgt der CAD-Bereich. Es kommt eine Software zur Planung und Konstruktion des Zahnersatzes zum Einsatz. Am digitalen, dreidimensionalen Modell, welches auf dem zuvor generierten Datensatz basiert, erfolgt das virtuelle Design des Zahnersatzes. Die physiologische Okklusionsposition kann ebenfalls mit in die Berechnung eingeschlossen werden, indem der entsprechende Gegenbiss ebenfalls

abgeformt wird. Der Zahnarzt kann seine Präparation nun direkt am Bildschirm stark vergrößert kontrollieren und bei eventuellen Aufnahmefehlern selektiv ohne aufwendige Zweitabformung nachscannen.

- Schließlich folgt der CAM-Bereich: Hierbei handelt es sich um eine Fertigungstechnologie zur Überführung der virtuellen Konstruktion auf einen dentalen Werkstoff. Grundlage für die Fertigung bildet der hergestellte Konstruktionsdatensatz. Derzeit kommen zur Umsetzung vor allem Schleif- und Fräsmaschinen zum Einsatz, die den Zahnersatz subtraktiv aus Rohlingen herausarbeiten. Alternativ kommen zunehmend auch additive Fertigungsverfahren zum Einsatz, auch bekannt als 3D-Druck.

Die beschriebene Prozesskette kann in vielen Bereichen der Zahnmedizin zum Einsatz kommen, beispielsweise zur Fertigung von feststehendem Zahnersatz im Sinne von Inlays, Teilkronen, Kronen oder Brücken, von implantatgetragenen Restaurationen oder von Schablonen für den Bereich der Implantatchirurgie.

Zielsetzung

Vor dem in der Einleitung dargestellten Problemhintergrund ermittelt die vorliegende Arbeit den Mehrwert der Analysesoftware prepCheck für die Bewertung von Kronenpräparationen in Bezug auf Präparationswinkel, Oberflächenrauigkeit, Oberflächenbeschaffenheit, Unterschnitte und okklusalen Substanzabtrag. Es wird untersucht, ob diese digitale Analysesoftware eine Vereinfachung und Objektivierung der Bewertung ermöglicht.

Hierzu sollen die folgenden Fragen beantwortet werden:

1. Gibt es signifikante Unterschiede zwischen den Beurteilungen derselben Präparationen, die die verschiedenen Zahnärzte in ihrer Rolle als Prüfer an der Uniklinik Aachen rein visuell unter Zuhilfenahme einer standardisierten Checkliste vornehmen? (Interrater-Reliabilität)
2. Gibt es signifikante Unterschiede in der Bewertung von Zahnpräparationen durch dieselben Prüfer unter Verwendung der Software prepCheck gegenüber der rein visuellen Bewertung anhand einer standardisierten Checkliste?
3. Wie konsistent sind die Bewertungen derselben Kronenpräparation durch einen Prüfer, wenn diese mit einer standardisierten Checkliste zu zwei verschiedenen Zeitpunkten durchgeführt werden? (Intrarater-Reliabilität)

Die entsprechenden Nullhypothesen lauten:

1. Es gibt keine signifikanten Unterschiede zwischen den Beurteilungen derselben Präparationen, die verschiedene Prüfer rein visuell unter Zuhilfenahme einer standardisierten Checkliste vornehmen.
2. Es gibt keine signifikanten Unterschiede in der Bewertung von Zahnpräparationen durch denselben Prüfer unter Verwendung der Software prepCheck gegenüber der rein visuellen Bewertung anhand einer standardisierten Checkliste.
3. Die Bewertungen derselben Kronenpräparation durch denselben Prüfer sind zu verschiedenen Zeitpunkten vollständig konsistent.

.

Material und Methoden

Die zur Beantwortung der oben gestellten Fragen durchgeführten Versuche werden im folgenden Kapitel im Detail dargestellt.

Es wurden in zwei Sitzungen 20 Zahnpräparationen für eine vollverblendete Krone durch zwölf approbierte Zahnärzte anhand einer Checkliste, beziehungsweise unter Zuhilfenahme einer Analysesoftware beurteilt.

Studiendesign

Die Studie war monozentrisch, prospektiv und komparativ geplant.

Ethische Aspekte

Nach Prüfung des Studiendesigns durch die zuständige Ethikkommission wurde festgestellt, dass für diese Studie kein Ethikvotum erforderlich ist.

Probanden

An der Untersuchung nahmen zwölf approbierte Zahnärzte aus der Abteilung für Zahnärztliche Prothetik der Uniklinik Aachen teil, von denen sechs weiblichen und sechs männlichen Geschlechts sind. Die Probanden nahmen in der Untersuchung die Rolle der Prüfer im vorklinischen Studentenkurs ein. Sechs Probanden hatten mehr als zehn Jahre Berufserfahrung, sechs Probanden hatten zwischen einem und sechs Jahren Berufserfahrung. Das Mindestalter der Probanden betrug 25 Jahre. Fünf Probanden waren zwischen 25 und 35 Jahren alt, sieben Probanden waren über 35 Jahre alt. Alle Teilnehmer hatten Erfahrung in der Bewertung studentischer Präparationen im Vorklinik- beziehungsweise Klinikalltag.

Eingeschlossene Präparationen

Um reale Beurteilungen zu schaffen wurden 20 von Studierenden präparierte Modellzähne der Firma Frasaco (frasaco GmbH, Deutschland) für die Untersuchungen verwendet. Die präparierten Zähne stammten aus der Prothetik-Vorwoche, die die Studierenden der RWTH Aachen als Zugangsvoraussetzung

zum ersten klinischen Behandlungskurs in der Abteilung für Zahnärztliche Prothetik, also vor dem achten Semester, absolvieren müssen. Die Präparationen (n=20) wurden zufällig anonym ausgewählt. Alle Präparationen wurden an Zahn 16 für eine vollverblendete Metallkeramikkrone nach den bekannten Präparationsrichtlinien und unter einheitlichen Prüfungsverhältnissen vorgenommen. Das heißt, alle Studierenden hatten dasselbe Zeitintervall zur Verfügung und konnten kein Feedback während der Durchführung einholen. Bei der Auswahl der präparierten Modellzähne war das Prüfungsurteil über die Präparationen weder bekannt noch relevant, die Qualität der Präparationen spielte also keine Rolle.

Den für die Studie ausgewählten Zahnpräparationen wurden zufällig die Nummern 1-20 zugeteilt. Die Nummern wurden mit einem walzenförmigen Diamanten in die „Wurzel“ eingraviert, sodass diese Nummer nur für die Studienleitung ersichtlich war. Anschließend wurden die Zähne in einen Sortierkasten eingeordnet.

Bewertungsbogen

Die Beurteilung der Präparationen erfolgte anhand von definierten Qualitätskriterien, die in einer Checkliste festgehalten wurden.

Der verwendete Bewertungsbogen ist bereits zur Beurteilung einer Frontzahnpräparation publiziert worden (Gratton, Kwon et al. 2016) und wurde daher von der Doktorandin nur geringfügig für den Seitenzahnbereich modifiziert.

Tabelle 1: Bewertungsbogen aus der vorliegenden Studie nach Gratton, Kwon et. al. 2016

	Test	Sehr gut (3 Punkte)	Klinisch verwendbar (2 Punkte)	Unter dem Standard (0 Punkte)
Okklusaler Abtrag	Morphologische Reduktion Vergleich mit kontralateralem Zahn, Silikonschlüssel	Vollanatomische Präparation	Teilweise anatomische Präparation	Keine anatomische Präparation
	Dachfirst Silikonschlüssel, Sonde	Dachfirstförmiger Abtrag des tragenden Höckers korrekt	Dachfirstförmiger Abtrag des tragenden Höckers ausreichend	Dachfirstförmiger Abtrag des tragenden Höckers nicht vorhanden
	Okklusale Reduktion Silikonschlüssel & Parodontalsonde	1,5 mm	1,5 – 2 mm	<1,5 mm oder >2 mm
Axiale Wandreduktion	Zirkuläre Reduktion Silikonschlüssel & Parodontalsonde	1,5 mm	1,5 – 2 mm	<1,5 mm >2 mm
	Präparationswinkel Parodontalsonde	3°- 10°	10°- 20°	<3° oder > 20°
	Untersichgehende Stellen Sonde, visuelle & taktile Inspektion	Keine Unterschnitte		Unterschnitte > 0,2 mm an Präparationsgrenze angrenzend
	Einschubachse Parodontalsonde, Sonde, Spiegel, kontralateraler Zahn	Parallel zur Zahnachse	Mit Krone versorgbar, jedoch nicht in Zahnachse (Gefahr der Pulpenschädigung)	Aufgrund der Achse nicht mit Restauration zu versorgen
Präparationsgrenze	Vertikale Position Parodontalsonde	supragingival 0,1- 0,9 mm dem Gingivaverlauf folgend	Supra-/ epigingival 0,9 – 1,4 mm dem Gingivaverlauf folgend	Subgingival >/= 1,5 mm supragingival Nicht dem Gingivaverlauf folgend
	Präparationsform Sonde, Silikonschlüssel	Ausgeprägte Hohlkehle 1,0- 1,2 mm Keine Schanze	Ausgeprägte Hohlkehle / abgerundete Stufe zwischen 0,75 -1,5 mm	Abgerundete Stufe / Hohlkehle / Tangentialpräparation <0,75 mm > 1,5 mm Schanzenbildung
	Approximalabstand Parodontalsonde gleitet an Präparationsgrenze vorbei	>0,5 mm		<0,5 mm
Feinschliff	Glättung Sonde	Gleichmäßige Rauigkeit ohne Kanten	Ausreichend, Sonde bleibt jedoch haken	Horizontale oder vertikale Rillen / Stufen
	Oberflächenübergänge Sonde, Silikonschlüssel	Abgerundet	Scharf, jedoch klinisch akzeptabel	Zu abgerundet zu scharf

Die Präparationsrichtlinien waren in vier Hauptkategorien unterteilt:

1. Okklusaler Abtrag
2. Axiale Wandreduktion
3. Präparationsgrenze
4. Feinschliff.

Innerhalb dieser Kategorien fanden sich jeweils enger definierte Punkte, sodass die Präparationen insgesamt anhand von zwölf Aspekten beurteilt wurden.

Für die Punkte auf der Checkliste standen drei abgestufte Kriterien zur Verfügung, die den folgenden Urteilen entsprachen: „sehr gut“, „klinisch verwendbar“ oder „unter dem Standard“. Je nach Beurteilung konnten so pro Teilaspekt drei, zwei oder null Punkte erreicht werden, woraus sich eine maximale Punktzahl von 36 ergab.

Für jedes Kriterium war genau definiert, zum Beispiel anhand einer angegebenen Spanne in Millimetern für die okklusale Reduktion oder in Grad für den Präparationswinkel, wann welches Urteil zu vergeben war.

Für zwei Kriterien, „Untersichgehende Stellen“ und „Approximalabstand“, fiel das Urteil „klinisch verwendbar“ weg.

Die Checkliste umfasste außerdem für jeden Punkt Empfehlungen für die zu verwendenden Hilfsmittel. Diese waren jeweils unter der Bezeichnung des Teilaspektes aufgeführt.

Analysesoftware

Bei der verwendeten Analysesoftware handelt es sich um prepCheck Version 1.1 (Sirona Dental Systems GmbH 2014, Bensheim, Deutschland), eine Applikation für die CEREC-AC-Einheit (Dentsply Sirona, Bensheim, Deutschland), die mit den systemeigenen digitalen Kameras Bluecam und Omnicam kompatibel ist, und mit deren Hilfe Zahnpräparationen computergestützt dokumentiert,

analysiert und archiviert werden können. Die Präparationen werden dabei mit einer Normpräparation oder mit einer sogenannten Masterpräparation verglichen.

Analyseparameter

Im Menü „Analyse“ können verschiedene Parameter analysiert werden. Die Konfiguration der einzelnen Parameter wird durch den Anwender vorgenommen. Die Parameter werden im Folgenden erläutert.

Präparationsrand

Hier werden scharfe Kanten oder stark konkave Flächen im Bereich der Präparationsgrenze orange markiert dargestellt. Alle Bereiche innerhalb der eingestellten Norm erscheinen weiß.

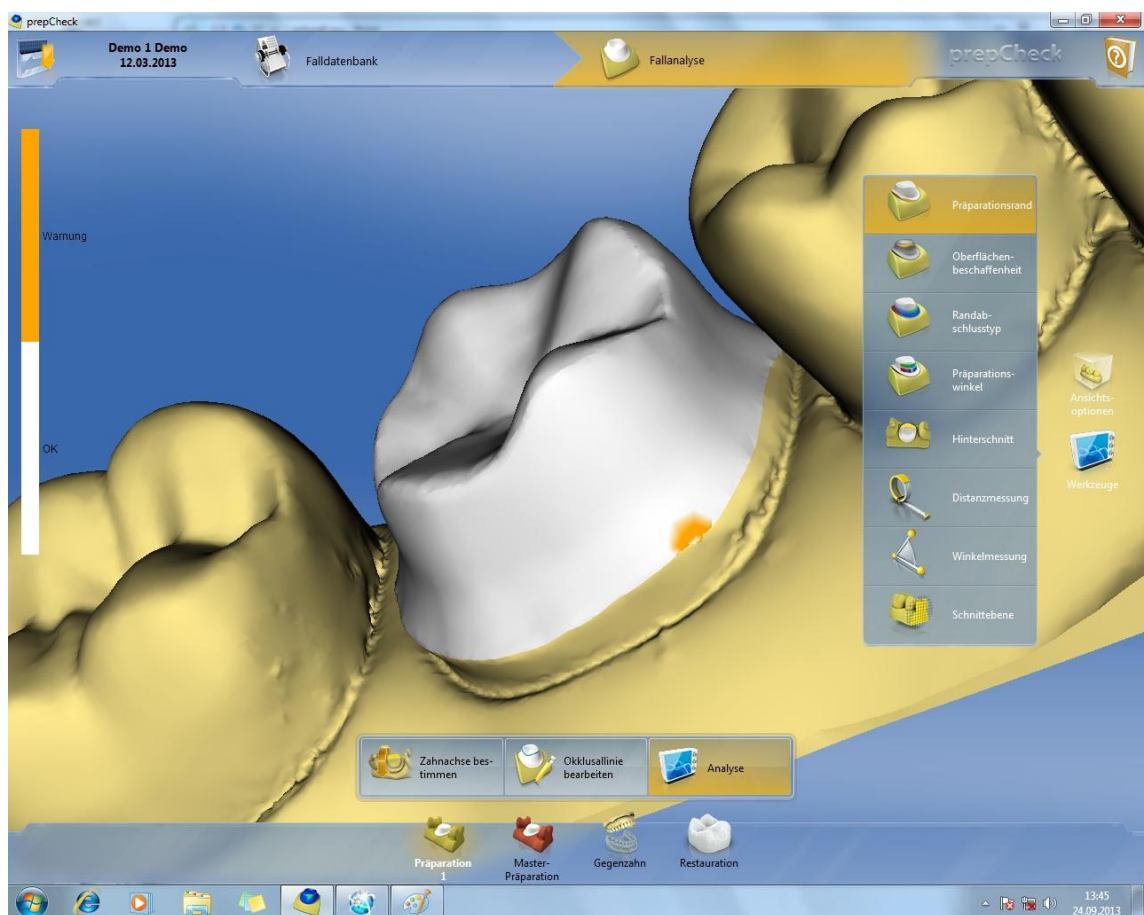


Abbildung 1: Screenshot aus prepCheck: Analyse des Präparationsrands

Oberflächenbeschaffenheit

Die Oberfläche wird hier auf den Krümmungsgrad hin untersucht. Dabei erscheint die Präparation in Grautönen: Weniger gekrümmte Bereiche sind hellgrau und stärker gekrümmte Bereiche dunkelgrau. Die Stufe des Grautons gibt also keinen Aufschluss über die Krümmungsrichtung, nur über deren Ausprägung. Scharfe Kanten (besonders stark gekrümmte Bereiche) werden auch hier orange dargestellt.

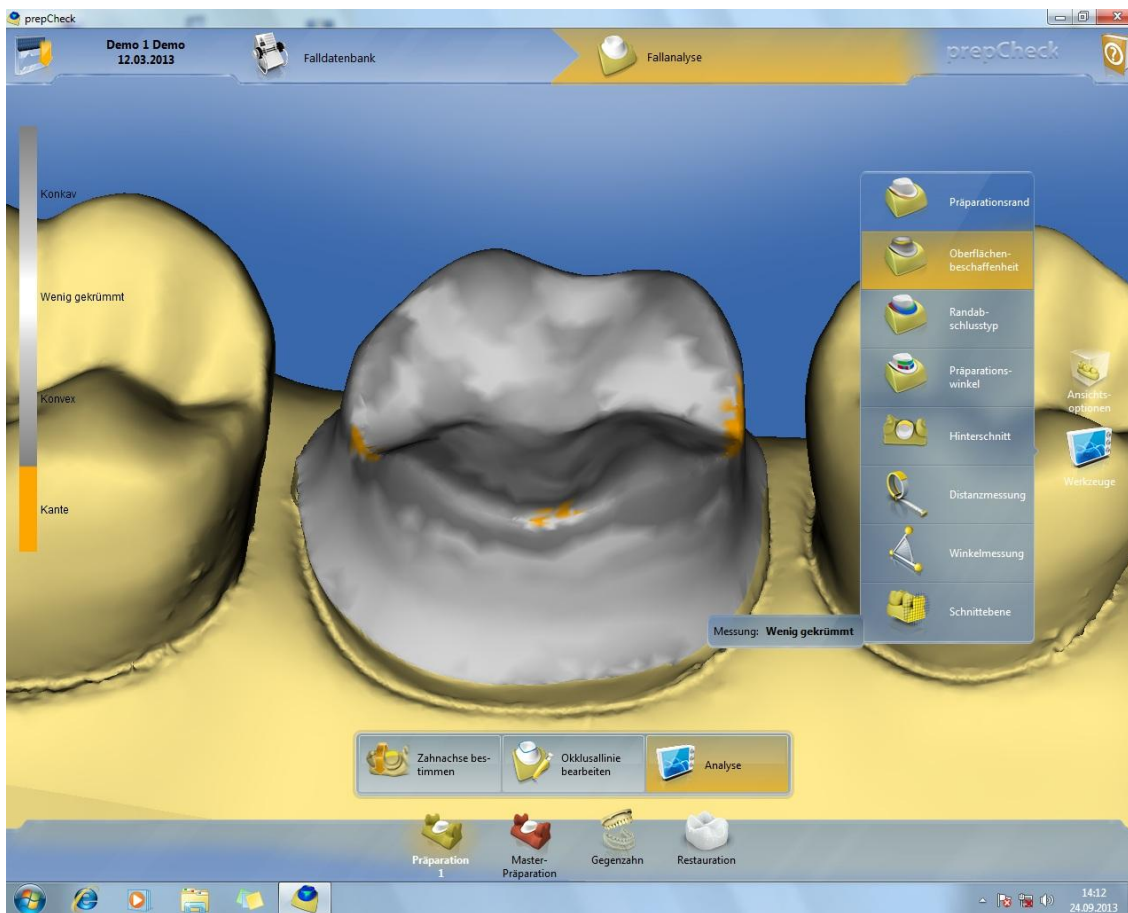


Abbildung 2: Screenshot aus prepCheck: Analyse der Oberflächenbeschaffenheit

Randabschlusstyp

Dieses Tool analysiert die Beschaffenheit der Präparationsgrenze anhand von zuvor definierten Kurvenverläufen und Toleranzbereichen. So kann in einem Koordinatensystem eine beliebige Kurve, entsprechend dem gewünschten Verlauf, angelegt werden, die zum Beispiel einer Hohlkehle oder Stufe entspricht. Die Analyse wird in einer dreistufigen Farbskala angezeigt, welche auch genauso

bei den Parametern Substanzabtrag und Präparationswinkel zur Verwendung kommt. Dabei zeigen die blauen Bereiche, dass die Präparation im Toleranzbereich liegt. Die grünen Bereiche zeigen an, dass mehr abgetragen werden muss und in rot markierten Bereichen wurde bereits zu viel abgetragen. Die Abweichung von der idealen Hohlkehle (sofern diese als Ideal hinterlegt ist), kann auch an jeder Stelle mit dem Cursor abgelesen werden.

Präparationswinkel

Wie oben beschrieben wird hier farblich abgestuft der zirkuläre Präparationswinkel dargestellt. Auch hier kann durch Bewegen des Cursors der Winkel an jeder beliebigen Stelle exakt angezeigt werden. Entscheidend für die Messwerte ist die zuvor durch den Anwender festgelegte Einschubachse.

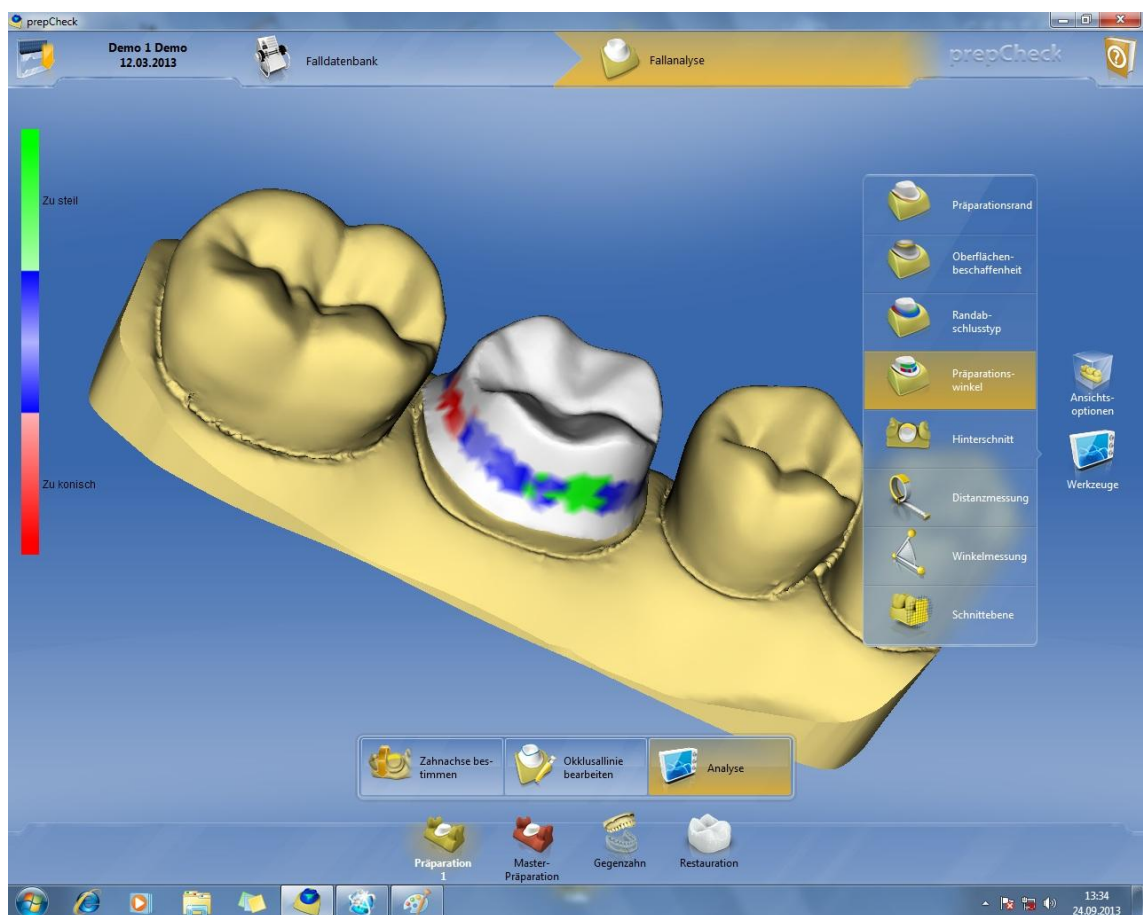


Abbildung 3: Screenshot aus prepCheck: Analyse des Präparationswinkels

Hinterschnitt

Dieser Parameter zeigt untersichgehende Bereiche in gelb. Auch hier wird der Bezug zur vorher eingestellten Einschubrichtung hergestellt.

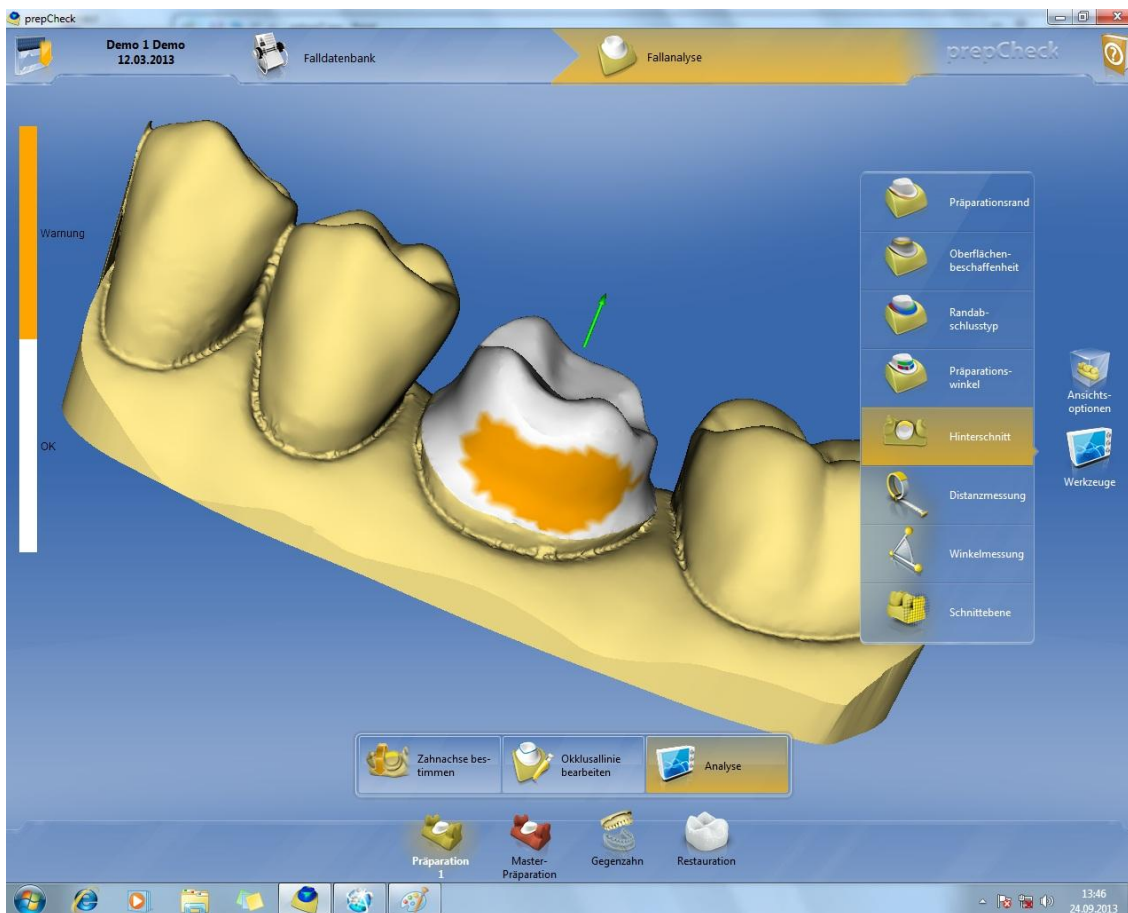


Abbildung 4: Screenshot aus prepCheck: Analyse des Hinterschnitts

Distanzmessung

Mit diesem Werkzeug können Distanzen zwischen zwei beliebigen Punkten, die durch einen Linksklick gesetzt werden, manuell vom Anwender bestimmt werden.

Distanz zum Gegenkiefer

Um den Substanzabtrag von okklusal zu beurteilen, kann der Gegenkiefer eingeblendet werden, wobei dieser beliebig transparent dargestellt werden kann. Es wird an jeder Stelle die kürzeste Distanz zwischen den beiden Oberflächen (Präparation und Antagonist) gemessen. Die Werte werden farblich dargestellt.

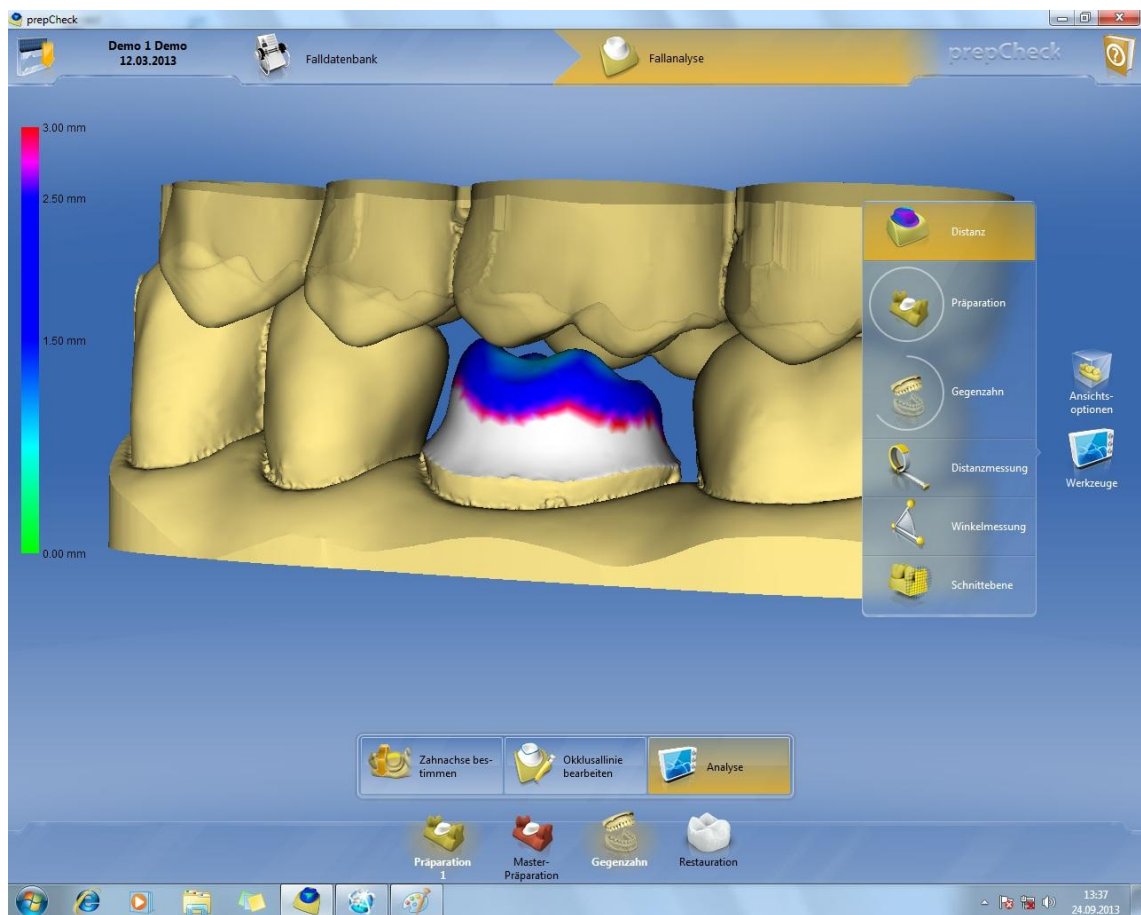


Abbildung 5: Screenshot aus prepCheck: Analyse der Distanz zum Gegenkiefer

Vergleich mit Masterpräparation

Außerdem kann die vorliegende Präparation mit der Masterpräparation überlagert werden. Hierbei handelt es sich in der Regel um eine zuvor eingespeiste ideale Präparation, die als Referenz für die zu beurteilenden Präparationen hinterlegt ist.

In der vorliegenden Studie wurde hingegen der unbeschiffene Zahn 16 als „Masterpräparation“ hinterlegt. So war es möglich, bei Überlagerung mit der Präparation, den Substanzabtrag mit Hilfe des Cursors an jeder beliebigen Stelle zu beurteilen. Die angezeigte und gewohnte Farbskalierung konnte jedoch nicht zu Rate gezogen werden. Im Gegenteil zeigten blaue Bereiche nun an, dass die Präparation an dieser Stelle nicht vom ursprünglichen Zahn abweicht.

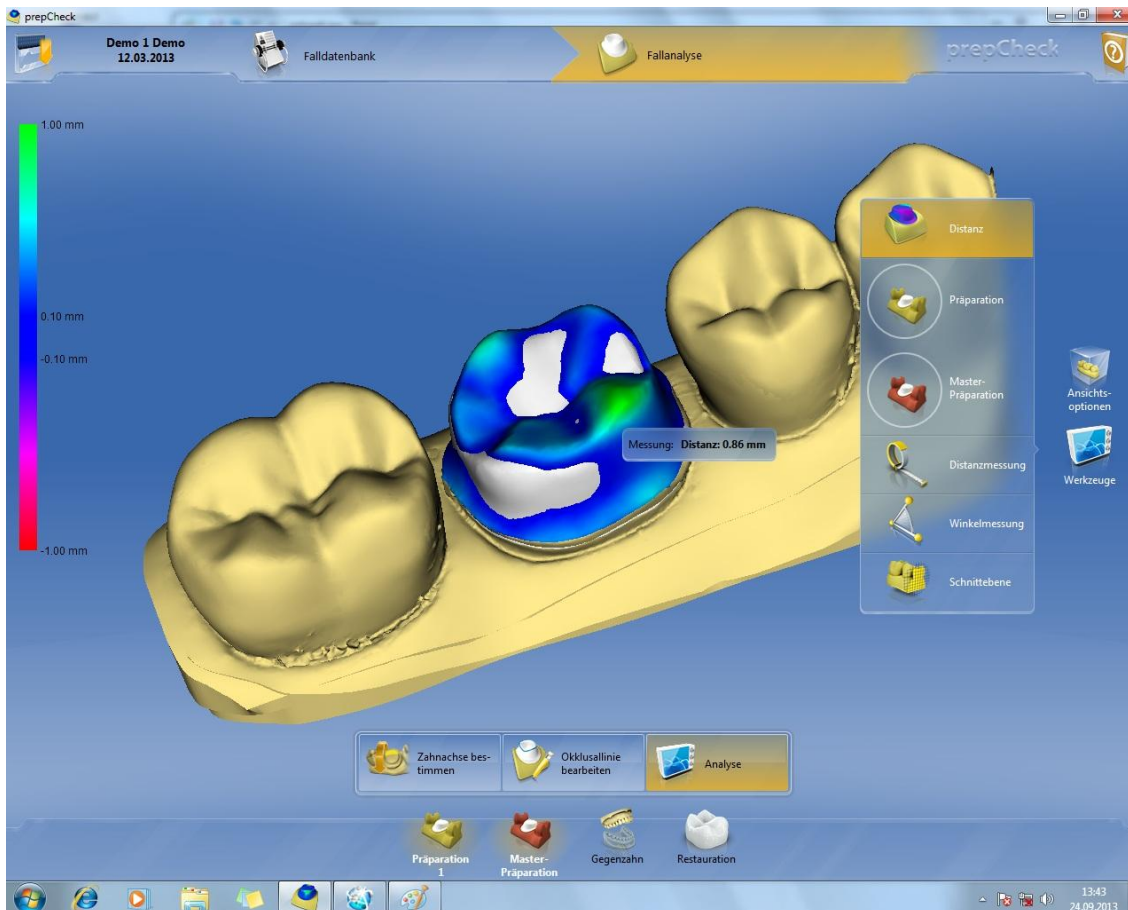


Abbildung 6: Screenshot aus prepCheck: Vergleich mit Masterpräparation

Ablauf

Alle Probanden führten die Versuche unabhängig voneinander, also zeitlich und räumlich getrennt, durch.

Erste Sitzung (Praetest)

In der ersten Sitzung wurden alle Präparationen ($n=20$) mithilfe des Bewertungsbogens visuell bewertet.

Die Checkliste wurde den Probanden zu Versuchsbeginn jeweils durch die Studienleitung erläutert. Während der Versuchsdurchführung mussten alle Entscheidungen allein von den Probanden getroffen werden.

Zur Bewertung standen den Probanden der in ein vollbezahntes Frasacomodell eingespannte, präparierte Zahn 16 mit dem dazugehörigen Gegenkiefer, sowie eine zahnärztliche Sonde, eine Parodontalsonde und zwei zuvor auf dem

unpräparierten Zahn 16 angefertigte Silikonschlüssel zur Verfügung. Die Zuhilfenahme der Hilfsmittel war optional.

Die Studienleitung war während der gesamten Versuchsdauer anwesend, um die Durchführung zu beaufsichtigen, Zeitmessungen während der Bewertungen vorzunehmen und um die präparierten Zähne auszuwechseln. Es wurden zwei Oberkiefermodelle im Wechsel verwendet, um Wartezeiten möglichst zu vermeiden. Für jeden Zahn wurde dem Probanden ein neuer Bewertungsbogen vorgelegt, der mit Datum, Zahn- und Probandennummer versehen war und auf dem nachträglich die für die Präparationsbeurteilung benötigte Zeit und die erreichte Punktzahl eingetragen wurden. Der Proband musste für jeden der zwölf Aspekte ein Kreuz in das Feld des seiner Meinung nach zutreffenden Urteils setzen.

Die Zeitmessung erfolgte für jeden Zahn einzeln und wurde jeweils gestartet, sobald das Frasacomodell an den Probanden ausgehändigt wurde und gestoppt, nachdem das letzte Kreuz gesetzt war. Der ausgefüllte Bewertungsbogen wurde dann an die Studienleitung zurückgegeben.

PrepCheck-Schulung

Vor Beginn der zweiten Messreihe wurde in der Abteilung eine Schulung zur Nutzung der Analysesoftware prepCheck im Sinne eines Hands-on Kurses abgehalten, um alle Mitarbeiter und insbesondere die Probanden mit dem Programm vertraut zu machen und auf einen einheitlichen Kenntnisstand zu bringen. Die Anwendung wurde anhand einer PowerPoint-Präsentation sowie einer Demonstration am Gerät erläutert.

Vorbereitung für zweite Sitzung

Um eine Präparationsanalyse mit prepCheck vorzunehmen, wurden Scans beider Kiefer sowie des Bukkalbisses benötigt. Für die Scans wurde die CEREC Omnicam verwendet. Die Scans wurden, um im Wesentlichen die für die Bewertung benötigte Zeit zu messen, vor Versuchsbeginn durch die Studienleitung vorgenommen. Hierzu wurden die Frasacomodelle in ein

Simulationskiefergelenk eingespannt, um eine reproduzierbare Okklusion zu gewährleisten.

Da nur ein Quadrant einen präparierten Zahn enthielt, konnte man auf einen Ganzkieferscan verzichten. Vom Hersteller wird jedoch gefordert, den Scan eines Quadranten bis zum zweiten Frontzahn jenseits der Mittellinie zu erweitern. Also wurden im Oberkiefer der Bereich von Zahn 17 bis Zahn 22 und im Unterkiefer der Bereich von Zahn 47 bis Zahn 32 gescannt.

Hierbei wurde der vorgegebene Scanpfad für die in der Studie verwendete Omnicam eingehalten: Der Scanvorgang wurde okklusal bei dem am weitesten distal gelegenen Zahn gestartet. Sobald dieser erkannt wurde, wurde die Kamera 45 Grad nach palatinal gekippt und aus diesem Winkel der gesamte Bereich gescannt. Bei Zahn 22 wurde die Kamera um weitere 45 Grad nach oral gekippt. Von hier aus wurde in umgekehrter Richtung bis Zahn 17 gescannt. Die gesamte Strecke wurde drei weitere Male abgescannt: Von okklusal, 45 Grad nach vestibulär, weitere 45 Grad nach vestibulär. Im Anschluss wurde der Scan betrachtet. Hierzu konnte das entstandene digitale 3D-Modell mithilfe des Trackballs und der Maustasten vergrößert oder in jede beliebige Richtung geschwenkt werden.

Bislang nicht abgebildete Bereiche wurden erneut gescannt, bis ein lückenloser Scan vorlag.

Analog wurde im Unterkiefer verfahren.

Zuletzt wurde der Bukkalbiss gescannt. Hierzu wurde das Simulationsgelenk geschlossen und ein Bereich von etwa zwei Zähnen beidseits des präparierten Zahns zuzüglich der Antagonisten im Unterkiefer aufgenommen. Waren alle Scans vollständig wurde mit Hilfe des Tools „Ausschneiden“ der präparierte Zahn im Oberkiefer- und Bukkalbiss-Scan wieder ausgeschnitten.

Zweite Sitzung (Posttest)

In der zweiten Sitzung, die in einem zeitlichen Abstand von mindestens zehn Wochen stattfand, wurden erneut dieselben Präparationen bewertet. Darüber wussten die Probanden während der ersten Sitzung nicht Bescheid.

Die Zähne mit den Nummern 1-10 wurden, wie gehabt, visuell mit der Checkliste und den optional zur Verfügung stehenden Hilfsmitteln der Sitzung 1 bewertet. Die Nummern 11-15 wurden in dieser Sitzung ebenfalls anhand der Checkliste, allerdings unter Zuhilfenahme der Software prepCheck, bewertet. Die Zähne 16-20 wurden in der zweiten Sitzung nicht bewertet.

Bei Versuchsbeginn wurde den Probanden, die bereits aus der ersten Sitzung bekannte Checkliste mit den Bewertungskriterien erneut erläutert.

Es wurde mit der Beurteilung der Zähne 11-15 unter Zuhilfenahme von prepCheck begonnen. Bevor die prepCheck-Applikation zum Einsatz kam, erfolgten die ersten notwendigen Arbeitsschritte in der CEREC-Software 4.0. Dazu gehörten der Scan, die Modellzuordnung, die Modellausrichtung und die Festlegung der Präparationsgrenze. Erst nach Eingabe dieser Informationen in die CEREC-Software konnte darauf aufbauend die prepCheck-Analyse erfolgen.

Die Probanden scannten jeweils den ausgeschnittenen Zahn und den ausgeschnittenen Bereich im Bukkalbiss. Dies erfolgte in Eigenregie, das heißt, es lag im Ermessen des Probanden, wann der Scan alle relevanten Informationen enthielt. Mit abgeschlossenem Scanvorgang wurde das Frasacomodell beiseitegelegt, um zu vermeiden, dass die Probanden auf eine visuelle Beurteilung umschwenken. Nach der Berechnung der Modelle erfolgte deren Zuordnung zueinander: Es wurden das Präparationsmodell, der transparent dargestellte Bukkalbiss und das Antagonistenmodell untereinander abgebildet. Durch Überlagerung der beiden Modelle mit dem Bukkalbiss konnte eine korrekte räumliche Modellzuordnung erreicht werden. Des Weiteren wurden die Modelle in der Vertikalen, Sagittalen und Horizontalen ausgerichtet.

Im nächsten Schritt wurde die Präparationsgrenze festgelegt. Hierzu konnte zwischen dem automatischen Kantenfinder und manueller Eingabe gewählt werden. Im Automatikmodus findet die CEREC-Software automatisch den Präparationsrand, wenn der Cursor an der Präparationsgrenze entlanggeführt wird. Der Verlauf wird als grüne Linie angezeigt. Korrekt detektierte Bereiche können mit Einzelclicks fixiert werden und erscheinen dann blau.

Bei der manuellen Eingabe definiert der Anwender durch Einzelclicks mit der linken Maustaste den Verlauf der Präparationsgrenze.

In beiden Modi wird die Eingabe gestartet und beendet mit einem Doppelklick auf die linke Maustaste. Der Start- bzw. Endpunkt wird als roter Punkt angezeigt.

Durch Drehen des Modells in alle Richtungen kann der Verlauf kontrolliert werden. Ungenauigkeiten können manuell ausgebessert werden, indem man erneut mit Doppelklick die Korrektur beginnt, den Verlauf durch Einzelklicks festlegt und die Korrektur nach Fehlerausbesserung mit einem Doppelklick abschließt.

Diese Schritte führten die Probanden eigenständig durch, sie lagen also in ihrem Ermessen. Nach Beendigung dieser Arbeitsschritte wurde die prepCheck-Applikation geöffnet. Zunächst erfolgte die Festlegung der Einschubrichtung durch den Probanden.

Schlussendlich konnte der Proband nun mithilfe des oben beschriebenen Analysetools die Beurteilung der Präparation durchführen. Dazu wurde erneut der bekannte Bewertungsbogen benutzt. Lediglich die bisher genutzten Hilfsmittel zur visuellen Beurteilung sowie das Simulationsmodell standen jetzt nicht zur Verfügung.

Auch während dieses Versuchs wurde eine Zeitmessung durchgeführt, die jeweils mit dem Scan durch den Probanden begann und gestoppt wurde, nachdem der Bewertungsbogen vollständig ausgefüllt war.

Im zweiten Versuchsteil wurden die Zähne 1-10 wie im Abschnitt „Erste Sitzung“ beschrieben bewertet.

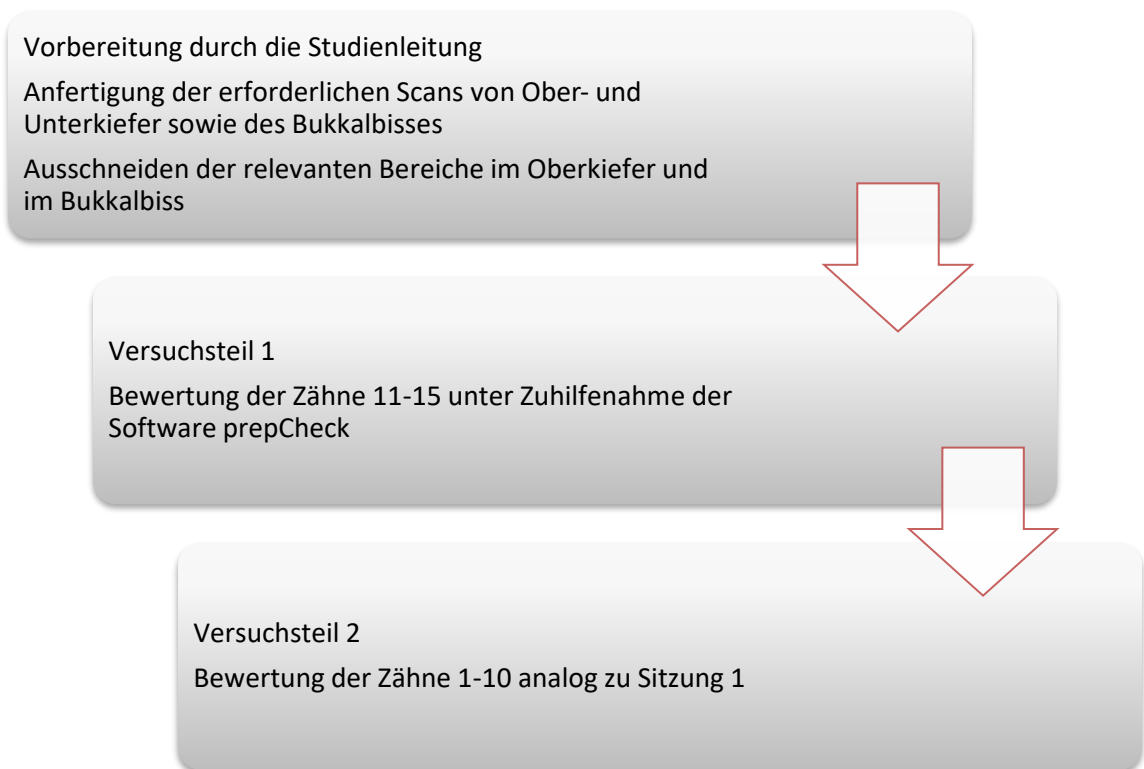


Abbildung 7: Vorbereitung und Ablauf der zweiten Sitzung

Statistische Auswertung

Alle erhobenen Daten wurden zunächst in eine Excel-Tabelle übertragen.

Die statistische Auswertung erfolgte mit der Software SPSS 25 (IBM).

Die Daten wurden zuerst deskriptiv analysiert.

Es folgte ein Test auf Normalverteilung sowohl grafisch als auch analytisch mittels Shapiro-Wilk Test. Der Vergleich der Präparationsparameter zwischen den Untersuchern erfolgte mittels Mann-Whitney U Test und einfaktorieller Varianzanalyse.

Der Vergleich der Ergebnisse der Bewertungen zwischen beiden Testzeitpunkten bezogen auf die Einzelwerte und die Gesamtpunktzahl erfolgte mittels Wilcoxon Signed rank test und einfaktorieller Varianzanalyse. Die Unterschiede bezogen auf die einzelnen Parameter innerhalb des Posttestes unter Zuhilfenahme von prepCheck gegenüber der herkömmlichen Variante erfolgte über den Mann Whitney U Test.

Die Interrater-Reliabilität wurde mittels Interklassenkorrelationskoeffizienten berechnet.

Die Interpretation der Reliabilitätskoeffizienten erfolgte nach der Einteilung nach Landis und Koch (Landis and Koch 1977).

Tabelle 2: Interpretation der Reliabilitätskoeffizienten nach Landis und Koch (Landis and Koch 1977)

Kappa	Übereinstimmung
≤ 0	Keine
0 – 0,2	Sehr geringe
0,2 – 0,4	Geringe
0,4 – 0,6	Mittlere
0,6 – 0,8	Hohe
0,8 - 1	Sehr hohe

Ergebnisse

In diesem Kapitel werden die Ergebnisse der zuvor beschriebenen Versuchsreihe zusammengefasst. Zunächst werden die Untersuchungen zu der ersten Sitzung (Praetest) dargestellt. Anschließend folgen die Untersuchungen zu der zweiten Sitzung (Posttest) und im Anschluss werden die Unterschiede zwischen den beiden Sitzungen beschrieben.

Untersuchungen zum Praetest

Tabelle 3: Gesamtpunktzahl Praetest (konventionelle Bewertung)

	Mittelwert	Standard- abweichung	N
Proband 1	21,00	0,837	20
Proband 2	21,90	1,110	20
Proband 3	25,35	0,877	20
Proband 4	15,90	1,076	20
Proband 5	19,40	1,360	20
Proband 6	25,90	1,355	20
Proband 7	23,25	1,100	20
Proband 8	26,25	1,081	20
Proband 9	19,35	1,367	20
Proband 10	21,75	1,256	20
Proband 11	19,25	1,497	20
Proband 12	22,40	0,916	20

Tabelle 3 zeigt zunächst deskriptiv die im Mittel durch die Probanden vergebenen Gesamtpunkte. Die minimal mögliche Punktzahl betrug 0 und die maximal mögliche Punktzahl lag bei 36 Punkten. Dabei wurden alle 20 Präparationen berücksichtigt, die im Praetest bewertet wurden.

Proband 4 vergab mit einem Mittelwert von 15,90 die wenigsten Punkte. Dies entspricht 44,2% der maximal erreichbaren Punktzahl. Proband 8 vergab durchschnittlich 26,25 Punkte (72,9% der maximalen Punktzahl) und damit die meisten unter allen Probanden.

Die Standardabweichung variiert für die Probanden zwischen 0,837 (Rater 1) und 1,497 (Rater 11).

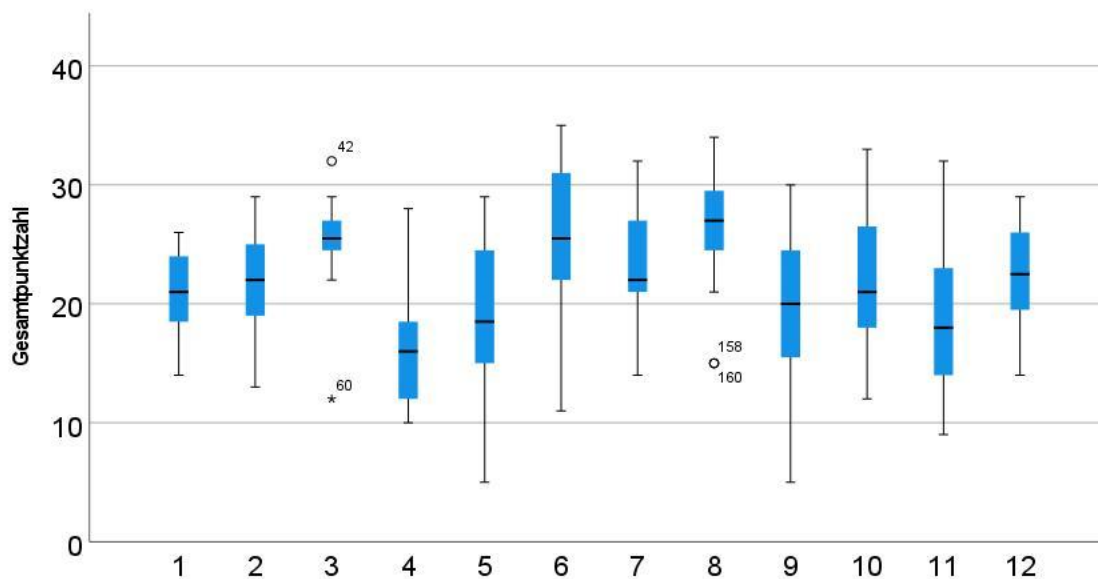


Abbildung 8: Vergebene Gesamtpunkte im Praetest

Man erkennt in Abbildung 8, die die deskriptiven Ergebnisse der vorigen Tabelle grafisch darstellt, dass es in der Bewertung Unterschiede gab. Die Probanden 3, 6 und 8 vergaben im Mittel die höchsten Punktzahlen, Proband 4 die wenigsten. Proband 4 war 35 Jahre alt, weiblich und hatte 12 Jahre Berufserfahrung. Proband 3 war 26 Jahre alt, männlich und hatte zwei Jahre Berufserfahrung. Proband 6 war 29 Jahre alt, weiblich und hatte 3,5 Jahre Berufserfahrung. Proband 8 war 29 Jahre alt, männlich und hatte 3,5 Jahre Berufserfahrung.

Bei sieben Probanden lag die im Mittel vergebene Punktzahl oberhalb der anzunehmenden Bestehensgrenze von 21,6 Punkten (60% der maximal erreichbaren Punktzahl von 36).

Die folgende Abbildung 9 zeigt die signifikanten Unterschiede in der vergebenen Gesamtpunktzahl im Praetest zwischen den einzelnen Bewertern. Die blau markierten Felder zeigen an, dass es zwischen der mittleren vergebenen Gesamtpunktzahl durch die beiden jeweils verglichenen Probanden signifikante

Unterschiede gab. Die letzte Spalte zeigt auf, wie viele Probanden signifikant verschiedene Punkte vergaben, als der jeweilige Proband.

Die meisten signifikanten Unterschiede zu anderen Probanden wies Proband 4 auf: 7 Probanden bewerteten signifikant anders.

Darauf folgen mit je 5 signifikanten Abweichungen die Probanden 7 und 9.

Proband	1	2	3	4	5	6	7	8	9	10	11	12	Anzahl sig. Unterschiede
1	■						■		■				2
2		■		■									1
3			■	■		■				■			3
4		■	■	■	■		■	■	■		■		7
5				■	■								1
6			■			■	■		■				3
7	■			■		■	■			■		■	5
8				■				■					1
9	■			■		■			■	■		■	5
10			■				■		■	■			3
11				■							■		1
12							■		■			■	2

Abbildung 9: Varianz der Bewertungen durch die Probanden

Tabelle 4 zeigt die Mittelwerte der im Praetest vergebenen Punkte sowie die Korrelation unter den Probanden, sowohl für die einzelnen Parameter als auch für die Gesamtpunkte. Es konnten pro Parameter 0, 2 oder 3 Punkte vergeben werden. Daraus ergibt sich die bereits erwähnte maximale Punktzahl von 36.

Mit einem Mittelwert von 0,71 wurde die „okklusale Reduktion“ im Praetest am schlechtesten bewertet. Geringfügig höher fiel das mittlere Urteil über die zirkuläre Reduktion aus (0,88).

Am besten hingegen wurde die Einschubachse bewertet (Mittelwert 2,83). Darauf folgte mit 2,81 das Kriterium „Approximalkontakte“.

Die vergebenen Gesamtpunkte lagen im Mittel bei 21,81. Der Korrelationskoeffizient betrug 0,883.

Die höchste Übereinstimmung unter den Ratern konnte für das Kriterium „Approximalkontakte“ erzielt werden und lag bei 0,884.

Mit einem Korrelationskoeffizient von 0,489 war die Übereinstimmung für das Kriterium „Glättung“ am geringsten.

Tabelle 4: Mittelwerte für die einzelnen Parameter und die Gesamtpunktzahl im Praetest mit n=240 (12 Probanden x 20 Präparationen) + Korrelation

	Mittelwert Praetest (konventionell)	Korrelationskoeffizient Praetest (konventionell)	Cronbachs Alpha
Morphologische Reduktion	1,99	0,782	0,830
Dachfirst	1,23	0,695	0,756
Okklusale Reduktion	0,71	0,778	0,798
Zirkuläre Reduktion	0,88	0,63	0,686
Präparationswinkel	1,78	0,727	0,771
Unterschnitte	1,93	0,797	0,853
Einschubachse	2,83	0,592	0,615
Vertikale Position	2,08	0,851	0,878
Präparationsform	1,32	0,817	0,836
Approximalkontakte	2,81	0,884	0,888
Glättung	2,17	0,489	0,630
Oberflächenübergänge	2,06	0,690	0,759
Gesamtpunkte	21,81	0,883	0,927

Untersuchungen zum Posttest

Die folgende Tabelle 5 zeigt die einzelnen Interrateranalysen für die im Posttest konventionell bewerteten Präparationen 1-10 (n=120).

Für die Gesamtpunktzahl lag der Korrelationskoeffizient bei 0,857. Dies entspricht einer sehr hohen Übereinstimmung. Betrachtet man die einzelnen Parameter, so stellt man fest, dass die höchste Übereinstimmung mit 0,895 für das Kriterium „Präparationsform“ erreicht wurde. Ebenfalls sehr hohe Übereinstimmungen ($>0,8$) lagen für vier weitere Kriterien vor: „Okklusale Reduktion“ (0,841), „Präparationswinkel“ (0,852), „Vertikale Position“ (0,837) und „Approximalkontakte“ (0,853). Für das Kriterium „Oberflächenübergänge“ lag der Korrelationskoeffizient mit 0,175 im sehr geringen Bereich. Für die restlichen Kriterien lag eine mittlere bis hohe Übereinstimmung vor.

Tabelle 5: Korrelationskoeffizienten für die konventionelle Bewertung der Präparationen 1-10 durch alle zwölf Probanden im Posttest; n=120 (12 Probanden x 10 Präparationen)

Konventionell Posttest (n=10)		
	Korrelationskoeffizient	Cronbachs Alpha
Morphologische Reduktion	0,753	0,810
Dachfirst	0,504	0,660
Okklusale Reduktion	0,841	0,833
Zirkuläre Reduktion	0,718	0,739
Präparationswinkel	0,852	0,895
Unterschnitte	0,512	0,649
Einschubachse	0,587	0,578
Vertikale Position	0,837	0,854
Präparationsform	0,895	0,951
Approximalkontakte	0,853	0,875
Glättung	0,720	0,800
Oberflächenübergänge	0,175	0,270
Gesamtpunkte	0,857	0,894

Die folgende Tabelle 6 zeigt die einzelnen Interrateranalysen für die im Posttest mit prepCheck bewerteten Präparationen 11-15.

Der Korrelationskoeffizient für die Gesamtpunktzahl lag mit 0,784 im hohen Bereich. Betrachtet man die einzelnen Parameter, so erreichten die Kriterien „Einschubachse“ und „Approximalkontakte“ sehr hohe Korrelationskoeffizienten ($> 0,8$). Alle Kriterien erreichten zumindest mittlere Werte für die Korrelation. Der geringste Korrelationskoeffizient (0,560) lag für das Kriterium „Präparationswinkel“ vor.

Tabelle 6: Korrelationskoeffizienten für die Bewertung der Präparationen 11-15 durch alle zwölf Probanden mit prepCheck im Posttest; n=60 (12 Probanden x 5 Präparationen)

prepCheck Posttest (n=5)		
	Korrelationskoeffizient	Cronbachs Alpha
Morphologische Reduktion	0,712	0,821
Dachfirst	0,593	0,667
Okklusale Reduktion	0,779	0,777
Zirkuläre Reduktion	0,665	0,761
Präparationswinkel	0,560	0,601
Unterschnitte	0,587	0,578
Einschubachse	0,856	0,842
Vertikale Position	0,648	0,689
Präparationsform	0,741	0,744
Approximalkontakte	0,873	0,843
Glättung	0,734	0,712
Oberflächenübergänge	0,587	0,568
Gesamtpunkte	0,784	0,843

Unterschiede zwischen erstem und zweitem Durchgang

Bezogen auf die Gesamtpunktzahl gab es, wie die Tabellen 4 und 5 zeigen, ohne prepCheck sowohl im Praetest als auch im Posttest eine sehr hohe Übereinstimmung bzw. Interraterreliabilität (jeweils > 0.8). Die Übereinstimmung unter der Nutzung von prepCheck (vgl. Tabelle 6) war mit einem Wert von 0,78 geringfügig schlechter, ist jedoch weiterhin als hoch einzustufen.

Vergleicht man die Korrelationskoeffizienten für die konventionelle Bewertung im Prae- bzw. Posttest miteinander (Tabellen 4 und 5) so kann festgestellt werden, dass für drei Parameter („Vertikale Position der Präparationsgrenze“, „Approximalkontakte“ und „Präparationsform“) in beiden Durchgängen sehr hohe Übereinstimmungen gefunden werden konnten. Die Parameter „Okklusale Reduktion“ und „Präparationswinkel“ wurden im Posttest mit sehr hoher Übereinstimmung bewertet, im Praetest lagen die Werte im Bereich guter Übereinstimmung. Hier wurde also eine Steigerung der Interrater-Reliabilität erreicht. Auch für die „Zirkuläre Reduktion“ verbesserte sich die Übereinstimmung von 0,63 auf 0,71. Diese Tendenz zur gesteigerten Korrelation ist aber nicht für alle Parameter zu finden: Beim Kriterium „Dachfirst“ herrschte im Posttest weniger Einigkeit. Der Korrelationskoeffizient sank von 0,69 auf 0,50.

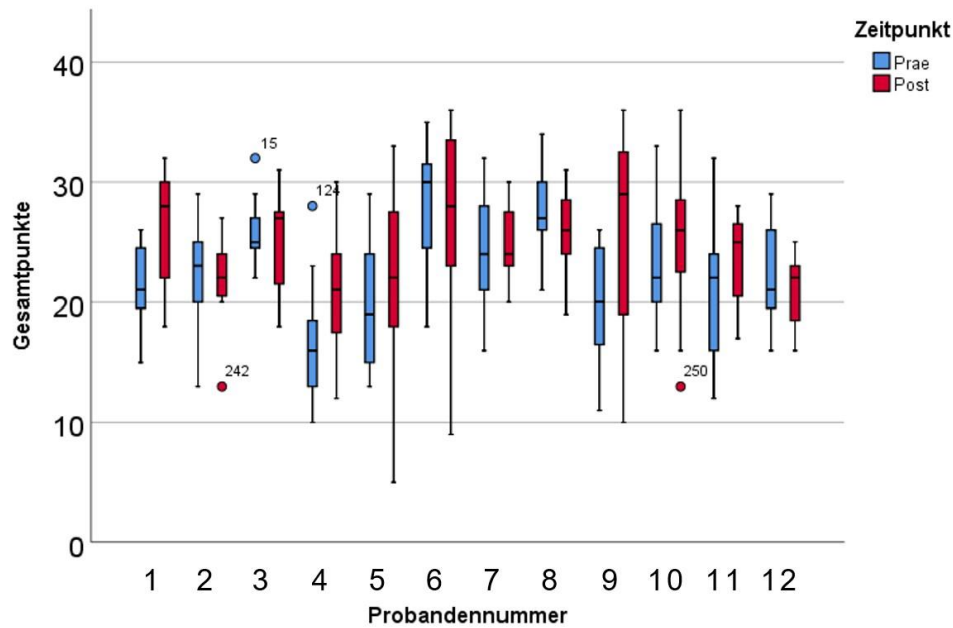


Abbildung 10: Durch die Probanden vergebene Gesamtpunkte im Prae- vs. Posttest (jeweils mit $n=15$)

Abbildung 10 stellt die durch die Probanden vergebenen Gesamtpunkte im Prae- und Posttest gegenüber. Hierbei wurden jeweils 15 Zahnpräparationen berücksichtigt, d.h. im Posttest sowohl die konventionell als auch die mit prepCheck bewerteten. Acht von zwölf Probanden vergaben im Posttest durchschnittlich mehr Punkte als im Praetest.

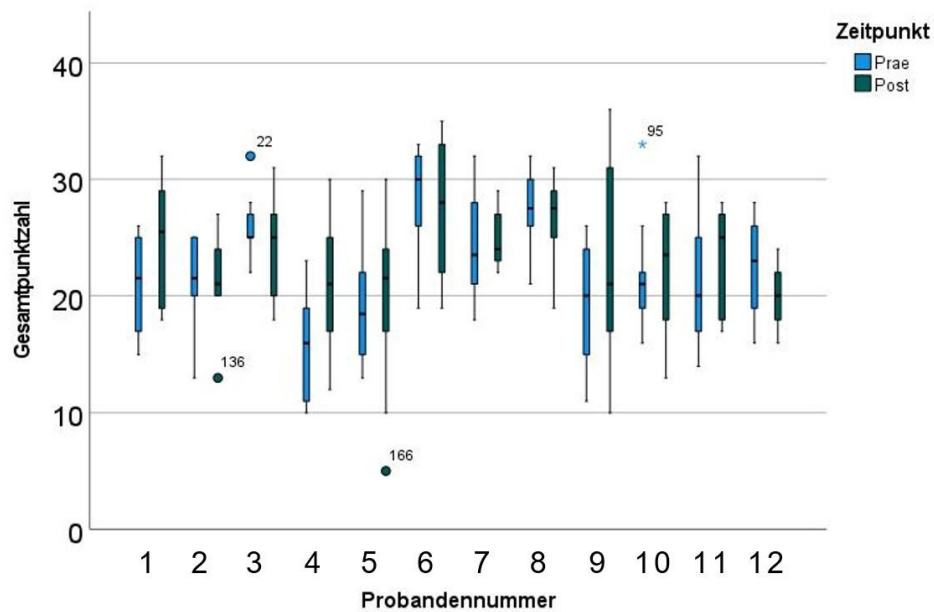


Abbildung 11: Vergebene Gesamtpunktzahl für die Zahnpräparationen 1-10 (konventionelle Bewertung) im Prae- bzw. Posttest durch die 12 Probanden

Abbildung 11 zeigt die durch die Probanden vergebenen Gesamtpunkte für die Zahnnummern 1-10, die in beiden Durchgängen ohne prepCheck bewertet wurden. Dabei werden Prae- und Posttest einander gegenübergestellt. Für sieben von zwölf Probanden lag der Median des Posttests über dem des Praetests, das heißt, es wurden höhere Gesamtpunktzahlen vergeben.

Tabelle 7: Gegenüberstellung der Zähne 1-10 (ohne prepCheck) im ersten beziehungsweise zweiten Durchgang. Mit * versehene und fett gedruckte Werte wiesen signifikante Abweichungen auf.

Zähne 1-10						
		Praetest (konventionell)		Posttest (konventionell)		
	n	Mittelwert	SD	Mittelwert	SD	p-Wert
Morphologische Reduktion	120	2,01	,874	2,02	,907	0,996
Dachfirst	120	1,23	1,156	1,27	1,242	0,866
Okklusale Reduktion	120	,91	1,257	1,12	1,323	0,105
Zirkuläre Reduktion	120	,92	1,213	1,02	1,240	0,360
Präparationswinkel	120	1,72	1,210	1,88	1,317	0,326
Unterschnitte	120	1,87	1,461	1,70	1,493	0,239
Einschubachse	120	2,78	,419	2,85	,423	0,068
Vertikale Position	120	2,24	1,021	2,35	,795	0,176
Präparationsform	120	1,43*	1,314	1,80*	1,274	0,009
Approximalkontakte	120	2,98	,274	2,95	,386	0,564
Glättung	120	2,17	,999	2,20	1,009	0,854
Oberflächenübergänge	120	2,17	,947	2,03	1,069	0,163
Gesamtpunkte	120	22,42	5,487	23,23	5,555	0,100
Zeit [min:sek:ms]	120	3:11:54*	1:33:51	2:33:00*	1:17:41	<0,001

Tabelle 7 stellt die Beurteilung der Zahnnummern 1-10 im ersten beziehungsweise zweiten Durchgang gegenüber. Die Zähne wurden in beiden Durchgängen konventionell anhand der Checkliste bewertet. Dafür wurde im zweiten Durchgang signifikant weniger Zeit benötigt.

Die mittlere Gesamtpunktzahl war im zweiten Durchgang tendenziell höher als im Ersten (vergleiche hierzu auch Abb. 12).

Betrachtet man die einzelnen Parameter, so wurde die Präparationsform im zweiten Durchgang signifikant besser bewertet ($p=0,009$).

Insgesamt wurden im zweiten Durchgang neun von zwölf Kriterien mit einer höheren Punktzahl beurteilt.

Für die Standardabweichung konnten im zweiten Durchgang tendenziell höhere Werte ermittelt werden. Das bedeutet, es gab eine größere Streuung der vergebenen Punkte.

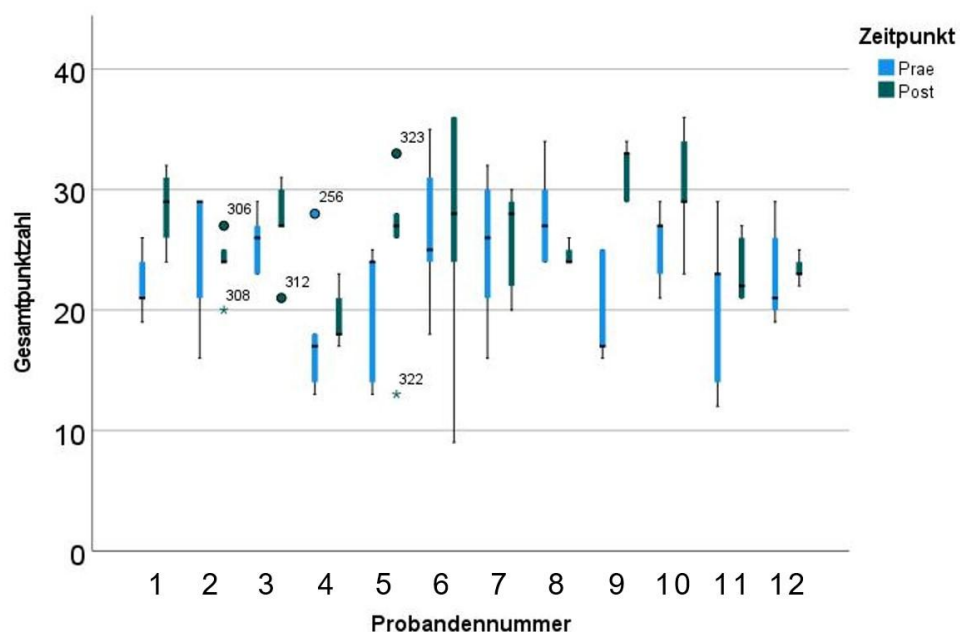


Abbildung 12: Vergebene Gesamtpunktzahl für die Zahnpräparationen 11-15 (mit prepcheck) im Prae- bzw. Posttest durch die 12 Probanden

Abbildung 12 zeigt die durch die Probanden vergebenen Gesamtpunkte für die Zahnnummern 11-15, die im ersten Durchgang ohne prepCheck und im zweiten Durchgang unter Zuhilfenahme der Software beurteilt wurden. Dabei werden Prae- und Posttest einander gegenübergestellt. Für neun von zwölf Probanden lag der Median des Posttests über dem des Praetests, das heißt, es wurden unter der Verwendung von prepCheck signifikant höhere Gesamtpunktzahlen vergeben (vergleiche hierzu auch Tabelle 8).

Tabelle 8: Gegenüberstellung der vergebenen Teilpunkte nach Beurteilungsparametern für die Zahnpräparationen 11-15 im Prae- beziehungsweise Posttest

Zähne 11-15						
		Prätest (konventionell)		Posttest (prepCheck)		
	n	Mittelwert	SD	Mittelwert	SD	p-Wert
Morphologische Reduktion	60	2,12	0,783	2,63	0,486	<0,001
Dachfirst	60	1,62	1,180	1,82	1,186	0,263
Okklusale Reduktion	60	0,78	1,236	0,80	1,219	0,930
Zirkuläre Reduktion	60	1,10	1,349	0,78	1,195	0,090
Präparationswinkel	60	1,98	1,308	1,97	1,301	0,908
Unterschnitte	60	2,20	1,338	2,80	0,755	0,005
Einschubachse	60	2,93	0,252	2,93	0,252	1,000
Vertikale Position	60	2,15	1,162	2,43	0,810	0,041
Präparationsform	60	1,50	1,295	2,25	1,198	0,001
Approximalkontakte	60	2,90	0,543	3,00	0,000	0,157
Glättung	60	2,25	1,099	2,20	0,879	0,567
Oberflächen- übergänge	60	1,72	1,166	2,33	0,795	<0,001
Gesamtpunkte	60	23,22	5,582	25,83	5,400	0,002
Zeit [min:sek:ms]	60	2:01:06	0:46:37	14:49:51	3:59:46	<0,001

In Tabelle 8 wird die Beurteilung der einzelnen Parameter für die Zahnnummern 11-15 durch 12 Probanden (n=60) im ersten beziehungsweise zweiten Durchgang gegenübergestellt. Im ersten Durchgang erfolgte die Beurteilung konventionell anhand der Checkliste, wobei sie im zweiten Durchgang unter Zuhilfenahme von prepCheck durchgeführt wurde.

Hierbei gab es für fünf Parameter, „Morphologische Reduktion“, „Unterschnitte“, „Vertikale Position“, „Präparationsform“ und „Oberflächenübergänge“, signifikante Unterschiede in der Bewertung: Alle fünf wurden mit prepCheck signifikant besser beurteilt.

Die Standardabweichung war im zweiten Durchgang für zehn von zwölf Parameter, darunter auch alle fünf signifikant besser bewerteten Parameter, niedriger, woraus sich eine höhere Einigkeit unter den Ratern ablesen lässt.

Auch die Gesamtpunktzahl mit prepCheck war signifikant erhöht.

Die für die Beurteilung benötigte Zeit lag im zweiten Durchgang bei 14:49 Minuten im Vergleich zu 2:01 Minuten im ersten Durchgang und war mit mehr als der siebenfachen Dauer somit signifikant höher.

Tendenziell, aber nicht signifikant schlechter wurden unter Verwendung von prepCheck die Parameter „Zirkuläre Reduktion“ und „Glättung“ bewertet.

Keine statistisch signifikante Differenz ($p > 0,9$) gab es für die Parameter „Präparationswinkel“ und „Okklusale Reduktion“.

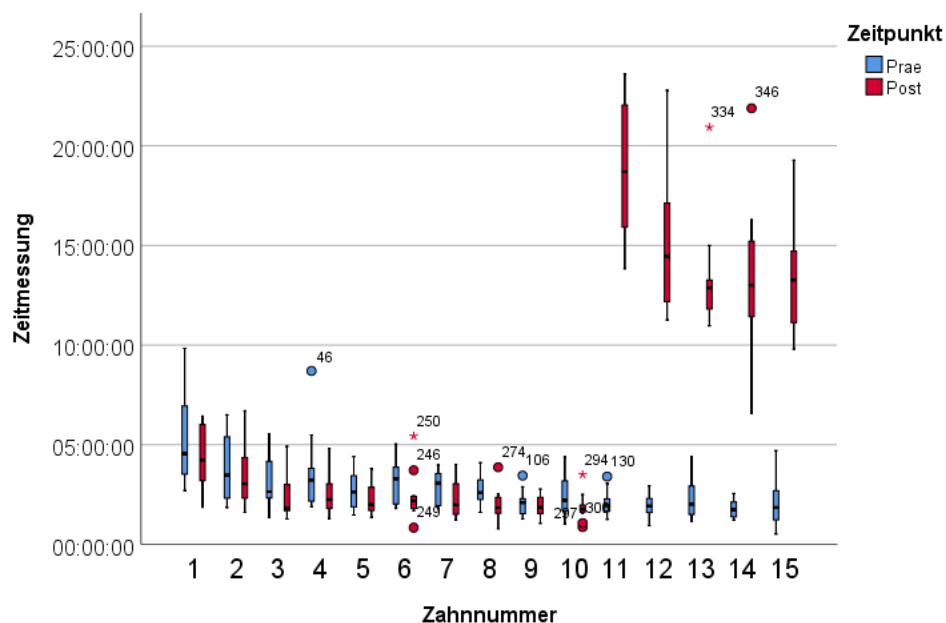


Abbildung 13: Die benötigte Zeit für die Bewertung der Zahnpräparationen 1-15 im Prae- beziehungsweise Posttest

Abbildung 13 veranschaulicht die Ergebnisse der Zeitmessung. Dabei werden die Ergebnisse für Prae- und Posttest einander gegenübergestellt. Im Posttest benötigten die Probanden für die konventionell beurteilten Zähne (Zahnnummern

1-10) weniger Zeit als im Prätest. Für die Bewertung mit prepCheck (Zahnnummern 11-15) benötigten alle jedoch fast die siebenfache Zeit.

Intrarater-Reliabilität

Die Intrarater-Reliabilität ist das Gütekriterium, das die Konsistenz beziehungsweise Zuverlässigkeit von Beurteilungen durch denselben Probanden zu verschiedenen Zeitpunkten misst. Zur Ermittlung der Intrarater-Reliabilität wurden in Tabelle 9 die Bewertungen der Zähne 1-10 aus dem Prae- und Posttest miteinander verglichen, also die konventionelle Bewertungsmethode. Diese Betrachtung wurde für jeden Probanden nach den zwölf Bewertungskriterien aufgeschlüsselt.

So lässt sich zum einen analysieren, welche Probanden konsistente Bewertungen vornahmen. Zum anderen kann aber auch eine Aussage darüber getroffen werden, ob für bestimmte Parameter gehäuft inkonsistente Bewertungen vorlagen.

Für das Kriterium „Approximalkontakte“ konnte für elf von zwölf Probanden eine sehr hohe, beziehungsweise für zehn Probanden sogar eine perfekte Übereinstimmung ihrer Beurteilungen gefunden werden. Auch für das Kriterium „Einschubachse“ erreichten sieben Probanden eine sehr hohe bis perfekte Übereinstimmung. Für beide Parameter wurden durchweg sehr hohe Punktzahlen vergeben. Bei allen anderen Parametern erreichten nie mehr als zwei Probanden sehr gute Werte für die Korrelation ihrer Urteile. Es gibt mehrere Parameter, für die alle Probanden maximal mittlere Werte für die Intrarater-Reliabilität erreichten: „Oberflächenübergänge“, „Präparationsform“, „Präparationswinkel“ und „Dachfirst“.

Vergleicht man die einzelnen Probanden, kann man folgende Ergebnisse festhalten:

Am besten schnitt Proband 13 ab, der für fünf von zwölf Kriterien hohe oder sehr hohe Werte für die Übereinstimmung erzielt. Darauf folgte Proband 7, der bei vier Parametern sehr gute bis perfekte Urteilsübereinstimmungen erzielte.

Alle anderen Probanden erzielten für maximal zwei Parameter sehr hohe bis perfekte Übereinstimmungen.

Proband 4 erzielt für neun Kriterien eine sehr geringe bis geringe Übereinstimmung.

Tabelle 9: Intrarater-Reliabilität nach Parametern, Gesamtpunkten und Zeit für Probanden 1-12

		morph.						okkl.			zirk.			Präparations						uners.			geh.			Einschub											
		Reduktion			κ			Dachfirst			κ			Reduktion			κ			winkel			κ			Stellen			κ			achse			pos		
Zahn 1-10		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		t		κ			
Probanden nummer	1	23	18	-0,21	2	8	0,29	10	13	0,62	12	13	0,43	23	23	0,34	0	6	0,00	25	30	0,20															
	2	17	19	0,32	10	10	0,29	10	12	-0,14	10	13	0,33	13	24	"-0,2	27	18	0,29	24	24	0,17															
	3	22	18	0,31	12	21	0,21	20	11	0,12	18	15	0,08	22	13	-0,06	15	9	0,57	29	27	0,41															
	4	18	17	0,14	2	9	0,23	4	10	0,06	0	4	0,09	13	15	0,27	18	18	0,17	30	30	0,81															
	5	15	15	0,38	14	12	0,14	3	5	0,13	2	8	-0,19	18	17	0,29	21	18	0,78	29	26	0,11															
	6	19	25	0,17	23	24	0,32	19	15	0,40	20	17	0,58	19	18	0,49	27	27	1,00	30	29	1,00															
	7	30	28	0,81	20	16	0,42	13	13	0,35	6	7	0,04	9	14	0,62	24	21	0,74	28	29	0,15															
	8	23	22	0,21	17	17	0,03	16	20	0,09	21	20	0,13	25	28	0,40	12	9	0,39	28	30	0,80															
	9	14	14	0,24	11	16	0,08	6	14	0,14	9	8	0,11	17	16	0,13	18	24	0,55	27	30	0,80															
	10	24	26	0,23	16	10	0,19	5	9	0,30	4	13	0,24	11	14	0,46	12	12	0,58	26	27	0,78															
	11	21	20	0,44	12	9	0,13	3	12	0,19	8	5	0,38	12	18	0,49	21	24	0,21	27	30	0,82															
	12	16	19	0,52	8	0	0,63	0	0	1,00	0	0	1,00	24	26	0,48	27	18	0,29	30	30	1,00															

		vert.			Präparations			approx.			Glättung			Oberflächen			κ			Gesamtpunkte			Zeitmessung														
		Position			κ			form			κ			κ			κ			κ			κ			κ			κ			κ					
Zahn 1-10		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post		prae	post				
Probanden nummer	1	22	22	0,38	18	24	0,20	30	30	1,00	25	28	0,00	27	28	0,21	212	248	26:33:00,0	20:09:00,0																	
	2	16	20	0,09	14	17	0,29	30	30	1,00	19	19	-0,11	21	9	0,02	213	213	37:47:00,0	27:17:00,0																	
	3	24	28	"-0,3	18	18	0,49	30	30	1,00	26	23	"-0,3	23	24	0,33	260	241	20:47:00,0	20:10:00,0																	
	4	20	21	0,20	3	25	0,05	30	30	1,00	12	20	0,26	12	15	0,55	159	215	43:28:00,0	25:23:00,0																	
	5	9	13	0,21	11	15	0,06	27	27	0,11	24	19	0,20	21	22	0,25	194	198	21:39:00,0	32:30:00,0																	
	6	28	30	0,81	8	15	0,50	30	30	1,00	22	27	0,38	29	26	0,19	282	277	42:04:00,0	40:55:00,0																	
	7	27	26	0,35	14	25	0,14	30	27	0,91	25	21	0,48	17	17	0,31	241	250	31:57:00,0	20:05:00,0																	
	8	29	24	0,23	22	22	0,36	30	30	1,00	26	20	0,17	28	24	0,07	276	267	31:12:00,0	22:56:00,0																	
	9	26	27	0,35	12	8	0,51	30	30	1,00	7	15	0,40	20	22	0,39	197	224	23:27:00,0	13:15:00,0																	
	10	23	24	0,67	18	16	0,47	30	30	1,00	22	21	0,21	23	24	0,33	216	224	53:19:00,0	41:49:00,0																	
	11	19	20	0,55	16	16	0,46	30	30	1,00	26	28	0,20	22	18	0,29	215	231	31:59:00,0	22:21:00,0																	
	12	26	27	0,09	18	15	0,07	30	30	1,00	26	23	0,81	17	15	0,11	225	200	19:37:00,0	19:12:00,0																	

Diskussion

Die vorliegende Diskussion behandelt die zentralen methodischen Aspekte, die wichtigsten Ergebnisse sowie die Implementierbarkeit der Analysesoftware prepCheck in den vorklinischen Kursalltag der Zahnmedizin.

Diskussion der Methode

Für die Untersuchung kamen die industriell gefertigten Zahnmodelle der Firma Frasaco (frasaco GmbH, Tettmang, Deutschland) zum Einsatz, an denen auch in den vorklinischen Kursen an der RWTH Aachen gelehrt wird. Da die Gingivamanschetten durch Alterung ihre Gestalt verändern können, wurden beide verwendeten Modelle mit unbenutzten Manschetten versehen, um zu gewährleisten, dass es nicht zu Veränderungen der Position der Präparationsgrenze und damit eihergehend zu abweichenden Urteilen kommt.

Um den Scan des Bukkalbisses für jeden Probanden reproduzieren zu können und damit die Durchführungsobjektivität zu steigern, kam im Posttest ein spezielles Simulationskiefergelenk zum Einsatz, in welches Ober- und Unterkiefer eingespannt werden.

Die Bedingungen während der unter Aufsicht durchgeführten Messung waren für alle Probanden identisch, um eine hohe Durchführungsobjektivität zu gewährleisten. Dies bedeutete, die gleiche Anzahl an zu bewertenden Präparationen, die gleiche Reihenfolge der Zähne, die Verwendung der gleichen Bewertungsbögen. Die Bewertungen der einzelnen Präparationen erfolgten direkt aufeinander folgend, um eine typische Situation zu simulieren, wie man sie während der Leistungsüberprüfungen in den vorklinischen Kursen findet, bei denen die Tutoren viele Beurteilungen in kurzer Zeit durchführen. So werden Rückschlüsse auf die Qualität der Benotung in den vorklinischen Behandlungskursen zugelassen. Es ist davon auszugehen, dass Stress und Ermüdung einen Einfluss auf die Ergebnisse ausüben (Gurung, Botelho et al. 2022).

Um bei der konventionellen Beurteilung ein hohes Niveau an Auswertungs- und Interpretationsobjektivität zu erzielen, fand im Vorhinein eine Kalibrierung der Probanden im Sinne einer halbtägigen Schulung zum Thema

Präparationsbeurteilung statt. Verschiedene Studien konnten belegen, dass eine solche Kalibrierung zu höheren Übereinstimmungen der Urteile, also zu einer gesteigerten Interrater-Reliabilität, führt (Haj-Ali and Feil 2006, Albino, Young et al. 2008).

Für die Beurteilungen wurde zusätzlich ein Bewertungsbogen mit definierten Kriterien und Schwellenwerten beziehungsweise Abstufungen herangezogen. Auch dieses Hilfsmittel wurde entwickelt und hat sich in der Bewertung von studentischen Arbeiten im vorklinischen Zahnmedizinstudium etabliert, um die Objektivität und Reproduzierbarkeit der Bewertungen zu steigern. Einige Untersuchungen konnten in der Vergangenheit positive Effekte feststellen (Gaines, Bruggers et al. 1974, Goepferd and Kerber 1980, Bedi, Lo et al. 1987, Kellersmann 2007, Schmitt, Möltner et al. 2016).

Bei dem in dieser Studie verwendeten Bewertungsbogen wurden die Präparationen anhand von zwölf Parametern bewertet. Es wurden, für die Qualität einer Präparation entscheidende, Aspekte in einer Checkliste zusammengefasst, die so bereits in einer anderen Studie (Gratton, Kwon et al. 2016) zum Einsatz gekommen war und nur geringfügig modifiziert wurde. Für jeden der zwölf Teilaspekte standen drei abgestufte und genau definierte Bewertungen zur Verfügung, um die Notenwahl zu erleichtern.

Vor dem Hintergrund der „Zentralen Tendenz“, die besagt, dass Beurteiler dazu neigen, die Extreme einer Notenskala auszusparen und ihre Noten schwerpunktmäßig in der Mitte zu vergeben (Jürgens and Sacher 2008), wäre eine vierstufige Bewertungsskala gegebenenfalls von Vorteil gewesen.

Der Bewertungsbogen enthielt teilweise Aspekte, die nicht durch die prepCheck-Applikation analysiert werden konnten. Es sollten gegebenenfalls Versuchsreihen durchgeführt werden, die ausschließlich Parameter berücksichtigen, die die Analysesoftware auswertet. Dadurch ließe sich in noch aussagekräftigerem Maß beurteilen, ob Bewertungen auf Basis von prepCheck einheitlicher sind als konventionell durchgeführte Bewertungen. Andererseits handelt es sich um Aspekte, die zu einer umfassenden Präparationsbeurteilung nötig sind, sodass es wünschenswert wäre, dass zukünftige Versionen von prepCheck diese Parameter evaluieren könnten.

Von den Probanden wurde an einzelnen Stellen Kritik an Bewertungskriterien geäußert. So wurde beispielsweise bemängelt, dass unter dem Punkt „Dachfirst“ kein Kriterium für „Dachfirst vorhanden, aber nicht ausreichend“ zur Auswahl stand. Dadurch kam es vereinzelt zu Unsicherheiten bei der Entscheidung zwischen „ausreichend“ für das Urteil „klinisch verwendbar“ und „nicht vorhanden“ für das Urteil „unter dem Standard“.

Der Bewertungsbogen umfasste nicht die Beurteilung der Nachbarzähne auf eventuelle Verletzungen hin, obwohl dies natürlich ein essenzieller Bestandteil der Präparationsbeurteilung ist. Es lässt sich jedoch aus dem Kriterium „Zirkuläre Reduktion“ ableiten, ob approximal ausreichend Zahnschubstanz abgetragen wurde, um nachbarzahnschonend zu präparieren. Dass das Kriterium nicht Bestandteil der Checkliste ist, ist dem Studiendesign geschuldet, da während der Versuchsdurchführung jeweils nur der präparierte Zahn 16 gewechselt wurde und das Frasacomodell ansonsten vollbezahlt blieb.

Außerdem standen den Probanden für die konventionelle Bewertung Hilfsmittel wie eine Parodontalsonde mit genormten Millimetermarkierungen und Silikonschlüssel vom unpräparierten Zahn zur Verfügung. Deren Nutzung war bewusst optional, da auch im Kursalltag davon auszugehen ist, dass nicht alle Prüfer Gebrauch davon machen.

Für die Präparationsbeurteilung unter Zuhilfenahme der Analysesoftware prepCheck kann man davon ausgehen, dass eine hohe Auswertungsobjektivität erreicht wurde. Allerdings ist nicht garantiert, dass bei der Beurteilung mit der Analysesoftware ein ebenso hohes Maß an Interpretationsobjektivität erzielt werden konnte, da es weiterhin dem Anwender unterliegt ein abschließendes Urteil zu fällen. So wertet die Software zwar genau aus, an welchen Stellen beispielsweise der Schubstanzabtrag außerhalb der definierten Toleranz liegt, der Anwender entscheidet jedoch, ob die Präparation dennoch dem klinischen Standard entspricht oder ob sie für dieses Kriterium unter dem Standard liegt.

Analysesoftware prepCheck

Es ist aktuell nicht möglich, Präparationen umfassend und vollständig durch prepCheck bewerten zu lassen. Präparationsgrenze sowie Okklusallinie müssen

vom Anwender eingezeichnet werden (Lenherr and Marinello 2014). Auch die Einschubachse legt der Benutzer fest. Als potenzielle Fehlerquellen haben diese anwenderabhängigen Punkte Einfluss auf die spätere Bewertung. Ebenfalls ist zu berücksichtigen, dass bereits die Qualität des Scans erheblichen Einfluss auf das Analyseergebnis hat. Bei Nichteinhaltung des Scanpfads oder Artefakten kann es zu eingeschränkter Aussagekraft des Analyseergebnisses kommen.

Die verwendete Softwareversion prepCheck 1.1 kann eine eventuelle Schädigung der Nachbarzähne nicht identifizieren oder bewerten. Lediglich über einen zu geringen Substanzabtrag approximal können Rückschlüsse auf Nachbarzahnschädigungen gezogen werden, wobei prepCheck dies in grün, also als „verbesserungswürdig“ einstufen würde. Hier liegt ein Nachteil der eingesetzten Software, da dieses Kriterium essenziell für die Beurteilung von Zahnpräparationen im vorklinischen Studentenkurs ist. Die Verletzung von Nachbarzähnen und das damit einhergehende erhöhte Kariesrisiko weisen eine hohe Inzidenz auf (Moopnar and Faulkner 1991, Strübig and Opitz 2000). Umso wichtiger ist es, die Studierenden bereits in der vorklinischen Ausbildung dafür zu sensibilisieren und darin zu schulen, nachbarzahnschonend zu präparieren. Da eine aktuelle Version der prepCheck-Applikation Nachbarzahnverletzungen messen kann, wäre es inzwischen möglich, eine Aussage diesbezüglich zu treffen, sowohl in einer Studie mit entsprechendem Studiendesign als auch im Kursalltag.

In der vorliegenden Studie haben Scan, Wartezeiten und Analyse sehr viel Zeit in Anspruch genommen, wie die Ergebnisse für die Zeitmessung bei der prepCheck Analyse im Posttest zeigen. Es kam bei der Menge an Aufnahmen und Daten gehäuft zum Absturz der CEREC Software sowie der Analysesoftware prepCheck. Daher war es aus Zeitgründen im Berufsalltag der Probanden nicht realisierbar, wie ursprünglich geplant, zehn Präparationen mithilfe von prepCheck zu bewerten und die Anzahl wurde durch die Studienleitung auf n=5 reduziert. Hierzu ist zu beachten, dass eine solche Problematik durch den technischen Fortschritt und die inzwischen verbesserte Rechenleistung heute nicht mehr zu erwarten wäre.

In der vorliegenden Studie wurde der unpräparierte Zahn als „Masterpräparation“ hinterlegt. Der Vergleich mit dem unbeschiffenen Zahn hilft zum einen bei der

Beurteilung der morphologischen Reduktion, also bei der Überprüfung der Erhaltung der ursprünglichen Zahnform. Außerdem lässt der direkte Vergleich mit dem unpräparierten Zahn eine bessere Einschätzung über die Entfernung zur Pulpa zu. Außerdem kann so an jeder Stelle der exakte Substanzabtrag abgelesen werden.

Es wäre jedoch sicher sinnvoll, sowohl eine Idealpräparation als auch den unpräparierten Zahn mit der Präparation überlagern beziehungsweise vergleichen zu können. Vor allem bei der Anwendung einer Analysesoftware zum Selbststudium kann die Überlagerung mit der Idealsituation für die Studierenden hilfreich sein, da ihnen so visualisiert wird, an welchen Stellen ihre Arbeit vom Optimum abweicht und verbesserungswürdig ist (Gratton, Kwon et al. 2016).

Die allgemeine Gültigkeit der Ergebnisse aus der digitalen Präparationsanalyse basiert auf der Präzision des verwendeten Intraoralscanners und der Software. Mit der Validierung der Präzision der CEREC Omnicam haben sich einige Studien befasst, die zu dem Ergebnis kamen, dass die CEREC Omnicam akzeptable Maße an Genauigkeit liefert (Boeddinghaus, Breloer et al. 2015, Ender, Zimmermann et al. 2016).

Umgang mit der CEREC Omnicam

Aus der Literatur ist bekannt, dass einige Zeit verstreicht, bis der Umgang mit digitalen Analysesystemen sicher beherrscht wird und die Nutzung effizient wird (Welk, Splieth et al. 2006, El Tantawi and Saleh 2008). Da nicht alle Probanden im Umgang mit der CEREC Omnicam und der Analysesoftware prepCheck vertraut waren, wurde im Vorfeld der zweiten Messung eine Schulung mit praktischen Übungen für die Abteilung der Zahnärztlichen Prothetik abgehalten. Zudem wurden die Scans von Gegenkiefer, Bukkalbiss und Oberkiefer bis auf den präparierten Zahn bereits im Vorhinein durch die Studienleitung vorgenommen, um Verzerrungseffekte in Bezug auf die benötigte Zeit durch Ungeübtheit der Probanden möglichst zu minimieren. Trotzdem zeigte sich auch in dieser Studie, dass die Probanden für die Beurteilung unter Zuhilfenahme von prepCheck die siebenfache Zeit gegenüber der konventionellen Methode benötigten. Dieser zeitliche Mehraufwand im Vergleich zur konventionellen

Präparationsbeurteilung entsteht nicht nur durch die Ungeübtheit der Probanden im Scanvorgang, sondern sicherlich auch durch andere Faktoren wie Unsicherheit in der Anwendung der Analysesoftware prepCheck und durch Wartezeiten während der Modellberechnung und während des Starts der prepCheck-Applikation. Diese Wartezeiten können Frustration seitens des Anwenders auslösen, genau wie nicht eigens lösbare technische Probleme, die während dieser Studie wiederholt auftraten und auch in der Literatur Erwähnung finden (El Tantawi, Abdelsalam et al. 2015).

Die beschriebenen Verzerrungseffekte müssen bei der Betrachtung der Ergebnisse der vorliegenden Studie berücksichtigt werden und könnten umgangen werden, indem in zukünftigen Untersuchungen intensivere Schulungen im Vorhinein abgehalten werden, beziehungsweise nur im Umgang mit CEREC geübtes Personal in das Probandenkollektiv aufgenommen wird.

Zudem wäre, um Wartezeiten zu verringern, eine verbesserte Leistungsfähigkeit der CEREC Software wünschenswert, wobei zu beachten ist, dass in dieser Studie mit der Version 4.0 CEREC Software und der Version 1.1 der prepCheck-Applikation gearbeitet wurde.

In einer anderen Untersuchung wurde die Aufgabe des Scans in die Hand der Studierenden übertragen, welche die Datei im Anschluss auf einem USB-Stick speicherten, sodass die Prüfer lediglich die prepCheck-Analyse durchführen mussten (Schepke, van Wulfften Palthe et al. 2020). Dieses Vorgehen würde im Vergleich zu der vorliegenden Studie zu einer Zeitersparnis für die Prüfer führen. Dieses Vorgehen wäre nicht nur für ein potenzielles zukünftiges Studiendesign interessant, sondern auch für eine Implementierung in den Kursalltag. In der beschriebenen Untersuchung wurden auch Einschubachse, Präparationsgrenze und Okklusallinie durch die Studierenden festgelegt. Neben Zeiteinsparung für die Prüfer bietet dieses Verfahren die Möglichkeit für die Studierenden, Routine in der Anwendung der Geräte zu erlangen und ihre erbrachte Leistung zu reflektieren. Es birgt aber auch die Gefahr, dass die Studierenden durch bewusste oder unbewusste falsche Festlegung der Parameter das Analyseergebnis manipulieren, indem sie beispielsweise die Präparationsgrenze oberhalb von eventuellen scharfen Kanten markieren und diese so aus der Bewertung herausfallen. Außerdem könnte eine Präparationsanalyse bei

mangelhafter Qualität des Scans erschwert oder unmöglich sein. Hier wäre die Validität des Tests zu überdenken, das heißt, ob der Test tatsächlich nur das zu messende Merkmal, in diesem Fall die Qualität der Kronenpräparation, überprüft.

Soll nur die Qualität der Präparation beurteilt werden, könnte der eigens durch die Studierenden angefertigte Scan zu Verzerrungen des Ergebnisses führen, da die Scanqualität das Analyseergebnis beeinflussen würde.

Angesichts der zunehmenden Relevanz digitaler Abformungen in der Zahnmedizin wäre es sicherlich wichtig, den Scan schon in einem vorklinischen Setting zum Bestandteil einer Leistungsüberprüfung zu machen (Zitzmann, Kovaltschuk et al. 2017, Cicciù, Fiorillo et al. 2020). Es erscheint jedoch sinnvoller, dies in einer separaten Prüfung zu tun, um nicht zu viele Fähigkeiten im selben Test zu überprüfen. So ließen sich am ehesten oben beschriebene Verzerrungseffekte vermeiden und die Validität der Tests maximieren.

Die Studierenden aus der oben beschriebenen Untersuchung von Schepke et al. berichteten zudem, dass das Erlernen des Scanvorgangs und des Umgangs mit dem Gerät so zeitintensiv war, dass sie das Gefühl hatten, nicht genug Zeit zur Prüfungsvorbereitung zu haben. Zudem sei es nach der Prüfung, als die Scans angefertigt wurden, zu Hektik und Chaos gekommen, was sich durch mehr zur Verfügung stehende Geräte hätte vermeiden lassen (Schepke, van Wulfften Palthe et al. 2020). Die Anzahl der vorhandenen Geräte und die mit der Anschaffung verbundenen Kosten sind somit limitierende Faktoren für die erfolgreiche Implementierung digitaler Abformungen und Präparationsanalysen in den Kursalltag.

Diskussion der Ergebnisse

Interrater-Reliabilität

Betrachtet man die Interrater-Reliabilität für den Praetest, so ist die Korrelation der Gesamtpunkte mit einem Korrelationskoeffizienten von 0,88 als sehr hoch einzustufen. Faktoren, die zu diesem Ergebnis beigetragen haben, könnten sein: Die regelmäßige Kalibrierung der Mitarbeiter der Zahnärztlichen Prothetik oder die zur Verfügung gestellte Checkliste als Maßstab. Der positive Einfluss dieser

Aspekte, die im Laufe der Zeit zur Steigerung der Objektivität eingeführt wurden, wird auch in einigen anderen Studien erwähnt (Gaines, Bruggers et al. 1974, Dhuru, Rypel et al. 1978, Goepferd and Kerber 1980, Jenkins, Dummer et al. 1998, Haj-Ali and Feil 2006, Schmitt, Möltner et al. 2016). Es konnte gezeigt werden, dass eine strukturierte Benotung anhand von definierten Parametern der „glance and grade“- Methode überlegen ist. Zudem wirkte sich die Kalibrierung der Tutoren positiv auf die Korrelation ihrer Beurteilungen aus. Die Studien konnten also zeigen, dass diese Maßnahmen geeignet sind, um die Bewertung von Präparationen, verglichen mit der Bewertung per in Augenscheinnahme, objektiver zu gestalten.

Dem entgegengesetzt stehen deutlich weniger Untersuchungen, die keinen positiven Effekt durch eine Checkliste feststellen konnten (Fuller 1972, Sharaf, AbdelAziz et al. 2007).

Trotz der, in Bezug auf die Gesamtpunkte, insgesamt sehr guten Interrater-Reliabilität, war die Variabilität zwischen Prüfern präsent: Es soll an dieser Stelle auf Proband 4 eingegangen werden, dessen Bewertungen signifikant von sieben anderen Probanden abwichen. Er vergab gleichzeitig im Mittel die wenigsten Punkte und erreichte nur sehr geringe Werte für die Intrarater-Reliabilität. Auch wenn es sich hier um einen Ausreißer handelt, sollte bedacht werden, dass eine solche Situation, in der das Ergebnis einer Leistungsüberprüfung einem extremen subjektiven Einfluss unterliegt, sich durchaus im Kurs zutragen kann.

Alle Probanden bewerteten signifikant anders als mindestens ein weiterer Proband. Diese Beobachtung zeigt, dass subjektive Einflüsse in der Leistungsbewertung bestehen bleiben, auch wenn standardisierte Checklisten und Schulungen zum Einsatz kommen. Damit ist die Nullhypothese 1 widerlegt, denn es gibt signifikante Unterschiede zwischen den Beurteilungen derselben Präparationen, die verschiedene Prüfer unter Zuhilfenahme einer standardisierten Checkliste vornehmen.

Für die einzelnen Parameter zeigen sich Abweichungen unter den Probanden:

Für zwei Parameter liegt die Übereinstimmung im mittleren Bereich („Glättung“ und „Einschubachse“). Für die „Glättung“ ist der Unterschied zwischen rein visueller Beurteilung und Beurteilung unter Zuhilfenahme eines Hilfsmittels

(fühlen mit Finger oder Sonde) womöglich ausschlaggebend für verschiedene Urteile. Es mangelt also an Auswertungsobjektivität. Man könnte diskutieren, ob die Ergebnisse bei obligatorischer Nutzung der Hilfsmittel höhere Übereinstimmungslevel zeigen würden, in dieser Untersuchung wurde jedoch, wie oben beschrieben, bewusst versucht, die Situation nachzustellen, wie sie im vorklinischen Kurs stattfindet. Das Kriterium „Einschubachse“ kann bei der konventionellen Methode nur geschätzt und nicht mit Hilfsmitteln gemessen werden, was zu einer verringerten Übereinstimmung der Urteile beigetragen haben kann.

Dass im Praetest die höchste Übereinstimmung für das Kriterium „Approximalabstand“ erzielt wurde und sich die sehr hohe Korrelation der Urteile im Posttest bestätigte, könnte möglicherweise darauf zurückzuführen sein, dass dieses Kriterium in der Checkliste keine mittlere Bewertungsstufe vorsah und somit lediglich zwei Auswahloptionen zur Verfügung standen. Es handelt sich hierbei folglich um ein im Vergleich zu den anderen Parametern klar abgrenzbares und weniger interpretationsanfälliges Kriterium. Dieses Ergebnis steht im Einklang mit den Ausführungen von Cohen, der beschreibt, dass dichotome Bewertungsskalen aufgrund des reduzierten Interpretationsspielraums tendenziell zu einer höheren Beurteilerübereinstimmung (Interrater-Reliabilität) führen (Cohen 1960).

Bezüglich des Gütekriteriums „Interrater-Reliabilität“ ist jedoch grundsätzlich zu hinterfragen, ob sich aus einer hohen Übereinstimmung der Urteile eine hohe Objektivität ableiten lässt. Nur, weil unter den Probanden eine gute Interrater-Reliabilität erreicht wird, ist nicht gesagt, dass die Probanden gemeinsam zum richtigen Urteil finden. Es könnte sich auch um eine gemeinsame Fehleinschätzung handeln.

Intrarater-Reliabilität

In dieser Studie wurde die Intrarater-Reliabilität, also die Übereinstimmung der Bewertungen desselben Prüfers zu unterschiedlichen Zeitpunkten, für die konventionelle Beurteilung unter Zuhilfenahme der Checkliste mit definierten Kriterien untersucht. Die ermittelten Reliabilitätskoeffizienten für die einzelnen

Probanden zeigen große Abweichungen auf. Einige Probanden waren in der Lage, konsistent zu bewerten, während andere in ihren Urteilen teils erheblich abwichen. Proband 13 erreichte für fünf Kriterien Reliabilitätskoeffizienten von $>0,8$. Das bedeutet, seine Bewertungen waren für diese Kriterien reproduzierbar. Für Proband 4 hingegen lagen die Koeffizienten für neun Kriterien bei $<0,3$. Die starken Abweichungen hinsichtlich der Reproduzierbarkeit der Urteile zeigen auf, dass die Probanden, die hier die Rolle eines Prüfers übernahmen, selbst einen großen Einfluss auf die Bewertung haben, auch wenn eine Kalibrierung stattfindet und obwohl die Beurteilung anhand einer Checkliste durchgeführt wird. Diese Schwankung zwischen einzelnen Prüfern hinsichtlich der Konsistenz ihrer Bewertungen lässt sich in anderen Untersuchungen ebenfalls nachweisen. So zeigten Miyazono, Shinozaki et al. in ihrer Studie, dass bei der dreimaligen konventionellen Bewertung durch denselben Prüfer keine perfekte Konsistenz erreicht wurde (Miyazono, Shinozaki et al. 2019).

Es konnte festgestellt werden, dass es den Probanden für einige Kriterien besser gelang, ihr Urteil zu reproduzieren als für andere: Beim Kriterium „Approximalabstand“ erreichten elf von zwölf Probanden sehr hohe Maße an Intrarater-Reliabilität. Auch die „Einschubachse“ wurde von sieben Probanden mit hoher bis sehr hoher Reliabilität benotet. Das Kriterium „Approximalabstand“ erforderte, wie oben bereits beschrieben, eine Entscheidung zwischen nur zwei Niveauabstufungen, die von den Probanden in den unterschiedlichen Tests eindeutig zu reproduzieren waren. Deutlich schlechter gelang den Probanden eine reproduzierbare Bewertung von komplexeren Kriterien wie „Okklusale Reduktion“ oder „Zirkuläre Reduktion“. Hier erreichten 75% der Probanden nur ein sehr geringes oder geringes Maß an Reliabilität. Eine Erklärung hierfür ist, dass es zwischen den abgestuften Bewertungen um Unterschiede von Bruchteilen von Millimetern geht, die trotz der genau definierten Bereiche und der Hilfsmittel wie Silikonschlüssel und Parodontalsonde schwierig reproduzierbar zu beurteilen sind. Zudem standen die Hilfsmittel optional zur Verfügung und einige Probanden machten keinen Gebrauch davon. Diese Ergebnisse, dass Kriterien wie der Approximalabstand eine höhere Reproduzierbarkeit aufweisen als komplexere, fein differenzierbare Parameter, erhielten auch Kwon, Restrepo-Kennedy et al., in deren Untersuchung sich zeigte, dass klar abgrenzbare Kriterien häufiger zu konsistenten Bewertungen führten, während komplexe oder

subjektiv schwerer bewertbare Parameter wie die okklusale Reduktion oder Präparationsform zu variableren Ergebnissen führten (Kwon, Restrepo-Kennedy et al. 2014). Auch eine weitere Studie, die die konventionelle Bewertung mit Checkliste, insbesondere bei komplexen Kriterien, häufig nicht ausreicht, um eine konsistente Bewertung sicherzustellen (Sharaf, AbdelAziz et al. 2007). Diese Untersuchung zeigte, dass auch bei erfahrenen Prüfern Variabilität besteht und bestätigt damit ebenfalls die in der vorliegenden Studie festgestellten Schwächen der konventionellen Methode.

Auch in der vorliegenden Studie konnte kein eindeutiger Zusammenhang zwischen Berufserfahrung und Reliabilität der Beurteilungen festgestellt werden. Dieses Ergebnis entspricht anderen Untersuchungen aus der Literatur (Sharaf, AbdelAziz et al. 2007, Al Amri, Sherfudhin et al. 2018, Grabowski 2019).

Betrachtet man die vergebenen Gesamtpunktzahlen und ruft sich erneut vor Augen, dass für die simulierte Prüfungssituation die Bestehensgrenze von 60% bei 21,6 Punkten gelegen hätte, fällt auf, dass im Praetest sechs Probanden im Mittel weniger Punkte vergaben, als zum Bestehen notwendig gewesen wären, während im Posttest vier Probanden im Mittel <21,6 Punkte vergaben. An dieser Beobachtung lassen sich die Konsequenzen aus subjektiven Bewertungseinflüssen ableiten.

Dieser, für die konventionelle Bewertung mit Checkliste, beobachtete Mangel an Reliabilität beziehungsweise Zuverlässigkeit, scheint besorgniserregend, wenn man bedenkt, dass auf gleiche Weise in den studentischen Kursen der Vorklinik über Bestehen oder Nichtbestehen einer Prüfung oder sogar eines Semesters entschieden wird. Die Untersuchungen haben gezeigt, dass es den Probanden trotz standardisierter Checkliste nicht zuverlässig gelingt, die Beurteilung zu reproduzieren.

Aufgrund der geringen Fallzahl kann allerdings nicht von einer Allgemeingültigkeit der Werte ausgegangen werden. Eine höhere Fallzahl war für die vorliegende Studie nicht zu realisieren, da nur eine begrenzte Anzahl an approbierten Zahnärzten in der Klinik für Zahnärztliche Prothetik beschäftigt ist. Zukünftig wäre es sicher interessant, eine ähnliche Untersuchung als Multicenter-Studie gemeinsam mit anderen deutschen Universitäten durchzuführen.

In anderen Untersuchungen, die die Intrarater-Reliabilität für die Bewertung mit der Analyse durch die prepCheck-Software untersucht haben, hat sich gezeigt, dass die Analysesoftware stets höhere Maße an Intrarater-Reliabilität erreicht (Renne, McGill et al. 2013, Kwon, Restrepo-Kennedy et al. 2014, Kateeb, Kamal et al. 2017, Grabowski 2019). Auch in der bereits zitierten Studie des Fukuoka Dental College in Japan, in der fünf Rater das Kollektiv an Präparationen drei Mal konventionell und drei Mal mit der Analysesoftware beurteilten, konnte gezeigt werden, dass kein Rater mit der konventionellen Methode eine perfekte Übereinstimmung in seinen Urteilen erzielen konnte. Mithilfe von prepCheck gelang es den Ratern hingegen häufig (Miyazono, Shinozaki et al. 2019).

Einfluss der Analysesoftware prepCheck auf die Beurteilung

Vergleicht man die konventionellen Beurteilungen für die Zähne 11-15 mit den prepCheck-basierten Beurteilungen, war die Standardabweichung mit prepCheck für zehn von zwölf Kriterien geringer, das heißt die vergebenen Punkte waren weniger breit gestreut. Daraus kann man eine höhere Übereinstimmung der Urteile ableiten. Dies steht im Einklang mit den Ergebnissen aus anderen Untersuchungen, die ebenfalls eine gesteigerte Urteilsübereinstimmung unter den Prüfern feststellen konnten (Kwon, Restrepo-Kennedy et al. 2014, Kateeb, Kamal et al. 2017). Außerdem wurden fünf Kriterien unter der Anwendung von prepCheck signifikant besser bewertet.

Also kann die Nullhypothese 2, es gebe keine Unterschiede in der Bewertung von Zahnpräparationen durch denselben Prüfer unter Verwendung der Analysesoftware gegenüber der visuellen Bewertung, ebenfalls widerlegt werden.

Die gehäuft signifikant bessere Beurteilung unter Anwendung von prepCheck gibt Grund zur Annahme, dass die Probanden bei der konventionellen Methode in einigen Fällen strenger und negativer über die studentischen Arbeiten urteilen. Hier stellt sich erneut die Frage, inwieweit die sehr hohen Werte für die Interrater-Reliabilität bei der konventionellen Methode überhaupt aussagekräftig sind, wenn man die Qualität der Beurteilungen erheben möchte. Es könnte sich hierbei auch um gemeinsame Fehleinschätzungen halten.

Betrachtet man nur den Posttest, wurde von den Parametern unter der Anwendung von prepCheck für das Kriterium „Präparationswinkel“ die höchste Standardabweichung gefunden. Eine höhere Standardabweichung bedeutet eine höhere Varianz in den vergebenen Punkten. Eine mögliche Erklärung in diesem Fall wäre, dass die Analyse des Präparationswinkels entscheidend von der zuvor festgelegten Einschubachse abhängt. Diese Einstellung wurde jeweils durch die Probanden vorgenommen und unterliegt somit dem subjektiven Einfluss des Anwenders.

Implementierbarkeit in den Kursalltag

Die Implementierung von prepCheck im vorklinischen Kursalltag bringt praktische Herausforderungen mit sich, insbesondere hinsichtlich des Zeitaufwands und der entstehenden Kosten.

Würde man in Prüfungssituationen wie in der vorliegenden Studie den Prüfer scannen lassen, so würden die Studierenden weiterhin nach Prüfungsende ihre Präparation im Simulationsmodell abgeben. In dem Fall würde der Prüfer ein Vielfaches der Zeit benötigen, verglichen mit einer konventionellen Bewertung. In der vorliegenden Studie wurden etwa 15 Minuten pro Präparation benötigt. Dies entsprach der siebenfachen Dauer im Vergleich zur konventionellen Beurteilung. Dabei ist zu berücksichtigen, dass die Zeitmessung in der Studie die Wartezeit, die vom Scanvorgang bis zur Darstellung der Modelle und bis zur Darstellung der Ergebnisse in prepCheck verging, mit einbezog. Auch andere Studien verzeichneten große Unterschiede im Zeitaufwand und kamen zu dem Ergebnis, dass die Wartezeiten ein akzeptables Maß überschreiten (Wolgin, Grabowski et al. 2018, Grabowski 2019).

Alternative Modelle, bei denen Studierende den Scan vornehmen und die Scandatei dem Prüfer übergeben, könnten die Arbeitslast der Prüfer erheblich reduzieren. Dieses Vorgehen setzt jedoch eine ausreichende Anzahl von Scannern voraus, um Wartezeiten und Frustration unter den Studierenden zu minimieren (Schepke, van Wulfften Palthe et al. 2020).

Stärken der Studie

Die Kalibrierung der Prüfer und die Anwendung der Checkliste trugen zur Konsistenz und Objektivität der Bewertungen bei und ermöglichten, subjektive Bewertungsfaktoren so weit wie möglich zu reduzieren. Die Methodik, die sowohl eine konventionelle als auch eine digitale Bewertung untersuchte, bietet wertvolle Einblicke in die Vorteile und Schwächen beider Systeme.

Limitationen der Studie

Die Ergebnisse der vorliegenden Untersuchungen sind aufgrund der begrenzten Teilnehmerzahl und der genutzten Softwareversion prepCheck 1.1 möglicherweise nur eingeschränkt auf andere Kontexte übertragbar. Für eine stärkere Allgemeingültigkeit wäre sowohl ein größeres Probandenkollektiv als auch eine größere Zahl an Proben, also präparierten Zähnen, wünschenswert.

Aufgrund der begrenzten Anzahl an Mitarbeitern in der Klinik war ein größeres Probandenkollektiv nicht realisierbar. Da die Versuche schon mit der eingesetzten Zahl an Proben viel Zeit in Anspruch nahmen, wäre eine größere Zahl nicht in den Arbeitsalltag der Teilnehmer zu integrieren gewesen.

Die in der Arbeit beschriebenen, möglichen Verzerrungseffekte durch Ungeübtheit der Probanden im Umgang mit der CEREC Omnicam, hätten durch stärkere Selektion der Probanden minimiert werden können. Da das Probandenkollektiv dadurch weiter geschrumpft wäre, wurde darauf verzichtet.

Schlussfolgerungen

Zusammenfassend kann festgestellt werden, dass die Intrarater-Reliabilität für die konventionelle Bewertungsmethode, trotz Checkliste mit definierten Kriterien und Hilfsmitteln, oft keine akzeptablen Werte annimmt und man somit davon ausgehen muss, dass es in den vorklinischen Kursen, in denen immer noch standardmäßig nach konventioneller Methode bewertet wird, zu subjektiven, inkonsistenten Bewertungen der studentischen Arbeiten kommt.

Die hohe Interrater-Reliabilität für die Gesamtpunktzahl deutet darauf hin, dass Maßnahmen wie regelmäßige Kalibrierung der Mitarbeiter und die Checkliste zu einer gewissen Objektivität beitragen können.

Dass einheitliche Bewertungen aber auch durch eine gemeinsame Fehleinschätzung zustande kommen können, könnte dadurch bestätigt werden, dass unter Anwendung von prepCheck gehäuft signifikant besser, also weniger streng bewertet wurde. Auch niedrigere Standardabweichungen bei den Bewertungen mit prepCheck im Vergleich zu den konventionellen Bewertungen, zeigen die Tendenz zu einer Objektivierung der Beurteilungen.

Die Interpretationsobjektivität bleibt jedoch ein Punkt, der optimiert werden sollte, entweder durch die direkte Notenvergabe durch die Software oder durch detaillierte prozentuale Angaben.

Der Zeitaufwand bleibt ein limitierender Faktor hinsichtlich des Mehrwerts. Hier ist zu erwarten, dass zukünftige Software-Versionen sowie intensivere Schulungen zu einer verbesserten Effizienz beitragen könnten.

Trotz der zeitlichen und technischen Herausforderungen zeigt die zunehmende Einbindung einer Analysesoftware in die vorklinischen Kurse des Zahnmedizinstudiums Potenzial, die Objektivität und Konsistenz in der Leistungsbewertung zu steigern. Sei es zum Selbststudium während der Vorbereitung auf Prüfungen, tatsächlich zur Präparationsbeurteilung durch die Prüfer oder nur um die Studierenden in der digitalen Zahnheilkunde, die auch außerhalb der Universität immer weiter an Bedeutung gewinnt, zu schulen, eine verstärkte Anwendung der Analysesoftware in der zahnmedizinischen Ausbildung ist zu begrüßen.

Zusammenfassung

Mehrwert eines digitalen Analysesystems für die Bewertung von Zahnpräparationen für eine vollverblendete Krone in der vorklinischen Ausbildung

Ziel: Diese Studie untersucht den Mehrwert einer digitalen Analysesoftware für die Beurteilung von studentischen Kronenpräparationen gegenüber der konventionellen visuellen Beurteilung mittels Checkliste.

Material und Methode: Zwanzig von Studierenden präparierte Kunststoffzähne wurden von zwölf Zahnärzten beurteilt. Die Beurteilung in der ersten Sitzung erfolgte konventionell anhand einer Checkliste mit zwölf Kriterien. Nach einer Kalibrierung der Prüfer und Schulung zur Nutzung der Analysesoftware wurde in der zweiten Sitzung ein Teil der Präparationen erneut visuell beurteilt (n=10) und ein anderer Teil (n=5) mit der Software „prepCheck“ (Sirona) bewertet. Die statistischen Auswertungen bezogen sich auf Zeitaufwand sowie Inter- und Intrarater-Reliabilität.

Ergebnisse: Die Ergebnisse zeigen, dass die konventionelle Methode mit Checkliste eine hohe Interrater-Reliabilität aufweist (Korrelationskoeffizient 0,88), jedoch Schwankungen in der Intrarater-Reliabilität. Die Bewertung einzelner Parameter wie Approximalabstand war konsistenter, während komplexere Kriterien wie okklusale und zirkuläre Reduktion signifikante Variabilität aufwiesen. Mit „prepCheck“ wurden fünf der zwölf Kriterien signifikant höher bewertet, bei tendenziell geringerer Standardabweichung. Der Zeitaufwand für prepCheck lag jedoch etwa siebenmal höher.

Schlussfolgerungen: Die konventionelle Methode bietet trotz Checkliste oft keine konsistente Intrarater-Reliabilität, was zu subjektiven Bewertungen in vorklinischen Kursen führen kann. Die hohe Interrater-Reliabilität deutet jedoch darauf hin, dass Kalibrierungen und Checklisten Objektivität fördern. Die Analysesoftware prepCheck führte zu objektiveren, konsistenteren Bewertungen mit weniger strengen Urteilen und geringerer Streuung, was eine höhere Bewertungsgenauigkeit nahelegt. Trotz des hohen Zeitaufwands ist der Einsatz solcher Software im zahnmedizinischen Unterricht vielversprechend, besonders zur Förderung digitaler Kompetenz und Leistungsbewertung.

Summary

Added value of a digital analysis system for the evaluation of tooth preparations for a full-coverage crown in preclinical education

Objective: This study examines the added value of a digital analysis software for evaluating student crown preparations compared to conventional visual assessment using a checklist.

Materials and Methods: Twenty typodont teeth prepared by students were assessed by twelve dentists. In the first session, assessments were conducted conventionally using a checklist with twelve criteria. After calibrating the examiners and training them to use the analysis software, in the second session, part of the preparations was re-evaluated visually (n=10), while another part (n=5) was evaluated using the software “prepCheck” (Sirona). Statistical analyses focused on time required as well as inter- and intra-rater reliability.

Results: The results show that the conventional method with a checklist demonstrated high inter-rater reliability (correlation coefficient 0.88), yet revealed fluctuations in intra-rater reliability. Assessment of individual parameters, such as proximal distance, was more consistent, while more complex criteria, such as occlusal and circumferential reduction, showed significant variability. Using “prepCheck,” five of the twelve criteria were rated significantly higher, with a generally lower standard deviation. However, the time required for prepCheck was approximately seven times higher.

Conclusions: Despite the checklist, the conventional method often lacks consistent intra-rater reliability, which can lead to subjective evaluations in preclinical courses. However, the high inter-rater reliability suggests that calibrations and checklists help foster objectivity. The prepCheck analysis software resulted in more objective and consistent evaluations, with less stringent judgments and reduced variability, suggesting greater accuracy in assessment. Despite the high time investment, using such software in dental education is promising, particularly for promoting digital competence and consistent performance evaluation.

Abbildungsverzeichnis

ABBILDUNG 1: SCREENSHOT AUS PREPCHECK: ANALYSE DES PRÄPARATIONSRANDS.....	23
ABBILDUNG 2: SCREENSHOT AUS PREPCHECK: ANALYSE DER OBERFLÄCHENBESCHAFFENHEIT	24
ABBILDUNG 3: SCREENSHOT AUS PREPCHECK: ANALYSE DES PRÄPARATIONSWINKELS	25
ABBILDUNG 4: SCREENSHOT AUS PREPCHECK: ANALYSE DES HINTERSCHNITTS	26
ABBILDUNG 5: SCREENSHOT AUS PREPCHECK: ANALYSE DER DISTANZ ZUM GEGENKIEFER	27
ABBILDUNG 6: SCREENSHOT AUS PREPCHECK: VERGLEICH MIT MASTERPRÄPARATION	28
ABBILDUNG 7: VORBEREITUNG UND ABLAUF DER ZWEITEN SITZUNG	33
ABBILDUNG 8: VERGEBENE GESAMTPUNKTE IM PRAETEST	36
ABBILDUNG 9: VARIANZ DER BEWERTUNGEN DURCH DIE PROBANDEN	37
ABBILDUNG 10: DURCH DIE PROBANDEN VERGEBENE GESAMTPUNKTE IM PRAE- VS. POSTTEST (JEWEILS MIT N=15)	42
ABBILDUNG 11: VERGEBENE GESAMTPUNKTZAHL FÜR DIE ZAHNPRÄPARATIONEN 1-10 (KONVENTIONELLE BEWERTUNG) IM PRAE- BZW. POSTTEST DURCH DIE 12 PROBANDEN	43
ABBILDUNG 12: VERGEBENE GESAMTPUNKTZAHL FÜR DIE ZAHNPRÄPARATIONEN 11-15 (MIT PREPCHECK) IM PRAE- BZW. POSTTEST DURCH DIE 12 PROBANDEN	45
ABBILDUNG 13: DIE BENÖTIGTE ZEIT FÜR DIE BEWERTUNG DER ZAHNPRÄPARATIONEN 1-15 IM PRAE- BEZIEHUNGSWEISE POSTTEST	47

Tabellenverzeichnis

TABELLE 1: BEWERTUNGSBOGEN AUS DER VORLIEGENDEN STUDIE NACH GRATTON, KWON ET. AL. 2016	21
TABELLE 2: INTERPRETATION DER RELIABILITÄTSKOEFFIZIENTEN NACH LANDIS UND KOCH (LANDIS AND KOCH 1977)	34
TABELLE 3: GESAMTPUNKTZAHL PRAETEST (KONVENTIONELLE BEWERTUNG)	35
TABELLE 4: MITTELWERTE FÜR DIE EINZELNEN PARAMETER UND DIE GESAMTPUNKTZAHL IM PRAETEST MIT N=240 (12 PROBANDEN X 20 PRÄPARATIONEN) + KORRELATION	38
TABELLE 5: KORRELATIONSKOEFFIZIENTEN FÜR DIE KONVENTIONELLE BEWERTUNG DER PRÄPARATIONEN 1-10 DURCH ALLE ZWÖLF PROBANDEN IM POSTTEST; N=120 (12 PROBANDEN X 10 PRÄPARATIONEN)	39
TABELLE 6: KORRELATIONSKOEFFIZIENTEN FÜR DIE BEWERTUNG DER PRÄPARATIONEN 11-15 DURCH ALLE ZWÖLF PROBANDEN MIT PREPCHECK IM POSTTEST; N=60 (12 PROBANDEN X 5 PRÄPARATIONEN)	40
TABELLE 7: GEGENÜBERSTELLUNG DER ZÄHNE 1-10 (OHNE PREPCHECK) IM ERSTEN BEZIEHUNGSWEISE ZWEITEN DURCHGANG. MIT * VERSEHENE UND FETT GEDRUCKTE WERTE WIESEN SIGNIFIKANTE ABWEICHUNGEN AUF.	44
TABELLE 8: GEGENÜBERSTELLUNG DER VERGEBENEN TEILPUNKTE NACH BEURTEILUNGSPARAMETERN FÜR DIE ZAHNPRÄPARATIONEN 11-15 IM PRAE- BEZIEHUNGSWEISE POSTTEST	46
TABELLE 9: INTRARATER-RELIABILITÄT NACH PARAMETERN, GESAMTPUNKTEN UND ZEIT FÜR PROBANDEN 1-12.....	49

Literatur

- Al-Dwairi, Z. N., M. M. Bashatwa and E. Lynch (2015). "Assessment of Posterior Teeth Preparations for Metal-Ceramic Crowns." The European Journal of Prosthodontics and Restorative Dentistry **23**(3): P141-149.
- Al Amri, M. D., H. R. Sherfudhin and S. R. Habib (2018). "Effects of Evaluator's fatigue and level of expertise on the global and analytical evaluation of preclinical tooth preparation." Journal of Prosthodontics **27**(7): 636-643.
- Albino, J. E., S. K. Young, L. M. Neumann, G. A. Kramer, S. C. Andrieu, L. Henson, B. Horn and W. D. Hendricson (2008). "Assessing dental students' competence: best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education." Journal of dental education **72**(12): 1405-1435.
- Amaya-Pajares, S. P., A. V. Ritter, C. Vera Resendiz, B. R. Henson, L. Culp and T. E. Donovan (2016). "Effect of finishing and polishing on the surface roughness of four ceramic materials after occlusal adjustment." Journal of Esthetic and Restorative Dentistry **28**(6): 382-396.
- Arnetz, G. and R. Dornhofer (2004). "PREPAssistant: A System for Evaluating Tooth Preparations PREPAssistant–ein System zur Evaluierung von Präparationsvorgaben." International journal of computerized dentistry **7**: 187-197.
- Bedi, R., E. Lo, N. King and T. Chan (1987). "The effect of pictorial criteria upon the reliability of assessments of cavity preparations." Journal of dentistry **15**(5): 222-224.
- Beuer, F., J. Schweiger and D. Edelhoff (2008). "Digital dentistry: an overview of recent developments for CAD/CAM generated restorations." British dental journal **204**(9): 505-511.
- Boeddinghaus, M., E. S. Breloer, P. Rehmann and B. Wöstmann (2015). "Accuracy of single-tooth restorations based on intraoral digital and conventional impressions in patients." Clinical oral investigations **19**(8): 2027-2034.
- Bortz, J. (2006). Statistik: Für Human-und Sozialwissenschaftler, Springer-Verlag.
- Cardoso, J., C. Barbosa, S. Fernandes, C. Silva and A. Pinho (2006). "Reducing subjectivity in the evaluation of pre-clinical dental preparations for fixed prosthodontics using the Kavo PrepAssistant®." European journal of dental education **10**(3): 149-156.
- Cicciù, M., L. Fiorillo, C. D'Amico, D. Gambino, E. M. Amantia, L. Laino, S. Crimi, P. Campagna, A. Bianchi and A. S. Herford (2020). "3D digital impression systems compared with traditional techniques in dentistry: a recent data systematic review." Materials **13**(8): 1982.
- Cohen, J. (1960). "A coefficient of agreement for nominal scales." Educational and psychological measurement **20**(1): 37-46.
- Dhuru, V. B., T. S. Rypel and W. M. Johnston (1978). "Criterion-oriented grading system for preclinical operative dentistry laboratory course." Journal of Dental Education **42**(9): 528-531.
- Edelhoff, D. and J. A. Sorensen (2002). "Tooth structure removal associated with various preparation designs for anterior teeth." The Journal of prosthetic dentistry **87**(5): 503-509.
- El Tantawi, M. and S. Saleh (2008). "Attitudes of dental students towards using computers in education-a mixed design study." EMHJ-Eastern Mediterranean Health Journal, **14** (3), 675-685, 2008.

El Tantawi, M. M., M. M. Abdelsalam, A. M. Mourady and I. M. Elrifae (2015). "E-assessment in a limited-resources dental school using an open-source learning management system." Journal of dental education **79**(5): 571-583.

Ender, A., M. Zimmermann, T. Attin and A. Mehl (2016). "In vivo precision of conventional and digital methods for obtaining quadrant dental impressions." Clinical oral investigations **20**: 1495-1504.

Feil, P. H. (1982). "An analysis of the reliability of a laboratory evaluation system." Journal of Dental Education **46**(8): 489-494.

Fritschi, E. M. and C. Roth (2014). "Leistungsbewertung in der Praxisausbildung." Praxisausbildung konkret: Am Beispiel des Bachelor in Sozialer Arbeit der Fachhochschule Nordwestschweiz FHNW: 81.

Fuller, J. L. (1972). "The effects of training and criterion models on interjudge reliability." Journal of Dental Education **36**(4): 19-22.

Gaines, W. G., H. Bruggers and R. H. Rasmussen (1974). "Reliability of ratings in preclinical fixed prosthodontics: effect of objective scaling." Journal of Dental Education **38**(12): 672-675.

Goepferd, S. J. and P. E. Kerber (1980). "A comparison of two methods for evaluating primary class II cavity preparations." Journal of Dental Education **44**(9): 537-542.

Grabowski, H. M. L. (2019). Untersuchungen zur möglichen Steigerung der Beurteilungsqualität von Präparationsergebnissen mittels einer Analysesoftware im Vergleich zur klassischen Beurteilung durch wissenschaftliche Mitarbeiter in der studentischen Ausbildung.

Gratton, D. G., S. R. Kwon, D. Blanchette and S. A. Aquilino (2016). "Impact of Digital Tooth Preparation Evaluation Technology on Preclinical Dental Students' Technical and Self-Evaluation Skills." Journal of Dental Education **80**(1): 91-99.

Gurung, A., A. Botelho, R. Thompson, A. Sales, S. Baral and N. Heffernan (2022). Considerate, unfair, or just fatigued? examining factors that impact teacher. Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.

Haj-Ali, R. and P. Feil (2006). "Rater reliability: short-and long-term effects of calibration training." Journal of dental education **70**(4): 428-433.

Hamil, L. M., A. S. Mennito, W. G. Renné and J. Vuthiganon (2014). "Dental students' opinions of preparation assessment with E4D compare software versus traditional methods." Journal of Dental Education **78**(10): 1424-1431.

Hobo, S. and H. T. Shillingburg (1973). "Porcelain fused to metal: tooth preparation and coping design." The Journal of prosthetic dentistry **30**(1): 28-36.

Haupt, M. I. and G. Kress (1971). "Accuracy of measurement of clinical performance in dentistry."

Ingenkamp, K. and U. Lissmann (2008). Lehrbuch der pädagogischen Diagnostik, Beltz Verlag.

Jenkins, S. M., P. M. Dummer, A. S. Gilmour, D. H. Edmunds, R. Hicks and P. Ash (1998). "Evaluating undergraduate preclinical operative skill; use of a glance and grade marking system." Journal of Dentistry **26**(8): 679-684.

Jürgens, E. and W. Sacher (2008). Leistungserziehung und pädagogische Diagnostik in der Schule: Grundlagen und Anregungen für die Praxis, Kohlhammer.

Karl, M., F. Graef, M. Wichmann and N. Beck (2011). "Evaluation of tooth preparations—A comparative study between faculty members and pre-clinical students." European Journal of Dental Education **15**(4): 250-254.

Kateeb, E., M. Kamal, A. Kadamani, R. Abu Hantash and M. Abu Arqoub (2017). "Utilising an innovative digital software to grade pre-clinical crown preparation exercise." European Journal of Dental Education **21**(4): 220-227.

Kellersmann, C. T. (2007). "Zur Reliabilität der Beurteilung vorklinischer Phantomarbeiten bei Einsatz eines strukturierten Bewertungsbogens." Inaugural-Dissertation. Münster: Westfälischer Wilhelms-Universität Münster.

Kern, M., S. Wolfart, G. Heydecke, S. Witkowski, J. C. Türp and J. R. Strub (2022). Curriculum Prothetik: Band 2, Quintessenz Verlag.

Kournetas, N., B. Jaeger, D. Axmann, M. Groten, S. Lachmann, H. Weber and J. Geis-Gerstorfer (2004). "Assessing the reliability of a digital preparation assistant system used in dental education." Journal of Dental Education **68**(12): 1228-1234.

Kwon, S. R., N. Restrepo-Kennedy, D. V. Dawson, M. Hernandez, G. Denehy, D. Blanchette, D. G. Gratton, S. A. Aquilino and S. R. Armstrong (2014). "Dental anatomy grading: comparison between conventional visual and a novel digital assessment technique." Journal of dental education **78**(12): 1655-1662.

Landis, J. R. and G. G. Koch (1977). "The measurement of observer agreement for categorical data." biometrics: 159-174.

Lenherr, P. and C. P. Marinello (2014). "prepCheck computergestützte objektive Beurteilung von Zahnpräparationen im Simulationslabor." Swiss Dental Journal **124**(10): 1085-1092.

Lilley, J. D., H. J. ten Bruggen Cate, P. J. Holloway, J. K. Holt and K. B. Start (1968). "Reliability of practical tests in operative dentistry." Br Dent J **125**(5): 194-197.

Liu, P.-R. and M. E. Essig (2008). "Panorama of dental CAD/CAM restorative systems." Compendium of continuing education in dentistry (Jamesburg, NJ: 1995) **29**(8): 482, 484, 486-488 passim.

Matthews, W., C. Showman and D. H. Pashley (1993). "Air blast-induced evaporative water loss from human dentine, in vitro." Archives of oral biology **38**(6): 517-523.

Miyazono, S., Y. Shinozaki, H. Sato, K. Isshi and J. Yamashita (2019). "Use of Digital Technology to Improve Objective and Reliable Assessment in Dental Student Simulation Laboratories." Journal of Dental Education **83**(10): 1224-1232.

Moopnar, M. and K. Faulkner (1991). "Accidental damage to teeth adjacent to crown-prepared abutment teeth." Australian dental journal **36**(2): 136-140.

Motta, A. B., L. C. Pereira, F. P. Duda and K. J. Anusavice (2014). "Influence of substructure design and occlusal reduction on the stress distribution in metal ceramic complete crowns: 3D finite element analysis." Journal of Prosthodontics **23**(5): 381-389.

Nagy, Z., B. Simon, Z. Tóth and J. Vág (2018). "Evaluating the efficiency of the Dental Teacher system as a digital preclinical teaching tool." European Journal of Dental Education **22**(3): e619-e623.

Natkin, E. and R. E. Guild (1967). "Evaluation of preclinical laboratory performance: a systematic study." J Dent Educ **31**(2): 152-161.

Orsmond, P., S. Merry and K. Reiling (2000). "The use of student derived marking criteria in peer and self-assessment." Assessment & Evaluation in Higher Education **25**(1): 23-38.

Park, C. F., J. M. Sheinbaum, Y. Tamada, R. Chandiramani, L. Lian, C. Lee, J. Da Silva and S. Ishikawa-Nagai (2017). "Dental Students' Perceptions of Digital Assessment Software for Preclinical Tooth Preparation Exercises." J Dent Educ **81**(5): 597-603.

Podhorsky, A., P. Rehmann and B. Wöstmann (2015). "Tooth preparation for full-coverage restorations—a literature review." Clinical oral investigations **19**(5): 959-968.

Pospeschill, M. (2010). Testtheorie, Testkonstruktion, Testevaluation, UTB.

Rekow, E. D. (2018). Digital dentistry: a comprehensive reference and preview of the future, Quintessence Publishing Company Limited.

Renne, W. G., S. T. McGill, A. S. Mennito, B. J. Wolf, N. M. Marlow, S. Shaftman and J. R. Holmes (2013). "E4D compare software: an alternative to faculty grading in dental education." J Dent Educ **77**(2): 168-175.

Rosenberg, H., H. A. Grad and D. W. Matear (2003). "The effectiveness of computer-aided, self-instructional programs in dental education: a systematic review of the literature." J Dent Educ **67**(5): 524-532.

Schepke, U., M. E. van Wulfften Palthe, E. W. Meisberger, W. Kerdijk, M. S. Cune and B. Blok (2020). "Digital assessment of a retentive full crown preparation - an evaluation of prepCheck in an undergraduate preclinical teaching environment." Eur J Dent Educ.

Schiffler, C. P. (2007). "Zur inter-und intraindividuellen Reliabilität der Beurteilung vorklinischer Zahnersatzarbeiten mittels Checkliste." Med. Diss. Duisburg 2007.

Schlenz, M. A., P. Rehmann and B. Wöstmann (2019). "Auf dem Weg zur digitalen Prothetik: Einblicke in den Praxisalltag." ZWR-Das Deutsche Zahnärzteblatt **128**(05): 214-220.

Schmitt, L., A. Möltner, S. Rüttermann and S. Gerhardt-Szép (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." GMS journal for medical education **33**(4).

Sharaf, A. A., A. M. AbdelAziz and O. A. El Meligy (2007). "Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures." J Dent Educ **71**(4): 540-544.

Strübig, W. and J. Opitz (2000). "Präparationsdefekte an Nachbarzähnen bei Inlay-und Kronenversorgung." Dtsch Zahnärztl(55): 101-103.

Tinschert, J., G. Natt, S. Hassenpflug and H. Spiekermann (2004). "Status of current CAD/CAM technology in dental medicine." Int J Comput Dent **7**(1): 25-45.

Vann, W. F., J. B. Machen and P. B. Hounshell (1983). "Effects of criteria and checklists on reliability in preclinical evaluation." J Dent Educ **47**(10): 671-675.

Welk, A., C. Splieth, M. Rosin, B. Kordass and G. Meyer (2004). "DentSim-a future teaching option for dentists." International journal of computerized dentistry **7**(2): 123-130.

Welk, A., C. Splieth, D. Seyer, M. Rosin, M. Siemer and G. Meyer (2006). "German dental faculty attitudes towards computer-assisted simulation systems correlated with personal and professional profiles." European Journal of Dental Education **10**(2): 87-95.

Welk, A., C. Splieth, E. Wierinck, R. O. Gilpatrick and G. Meyer (2006). "Computer-assisted learning and simulation systems in dentistry--a challenge to society." Int J Comput Dent **9**(3): 253-265.

Wilson, R. D. and G. Maynard (1981). "Intracrevicular restorative dentistry." Int J Periodontics Restorative Dent **1**(4): 34-49.

Wolgin, M., S. Grabowski, S. Elhadad, W. Frank and A. M. Kielbassa (2018). "Comparison of a prepCheck-supported self-assessment concept with

conventional faculty supervision in a pre-clinical simulation environment." Eur J Dent Educ **22**(3): e522-e529.

Zitzmann, N. U., I. Kovaltschuk, P. Lenherr, P. Dedem and T. Joda (2017). "Dental students' perceptions of digital and conventional impression techniques: A randomized controlled trial." Journal of dental education **81**(10): 1227-1232.

Danksagung

Ich durfte die Arbeit an dieser Dissertation während meines Studiums an der RWTH Aachen beginnen und sie nun als approbierte Zahnärztin, tätig in einer Gemeinschaftspraxis mit meinem Mann und meinen Schwiegereltern, fertigstellen.

Mein besonderer Dank gilt Herrn Prof. Dr. Stefan Wolfart für die Überlassung des Themas der Dissertation und ihm und Frau Dr. Anne Rittich für ihre Betreuung.

Zudem möchte ich mich recht herzlich bei meinen Probanden, den Zahnärzten der Klinik für Zahnärztliche Prothetik an der RWTH Aachen, für die aufgebrachte Zeit und Teilnahme an der Versuchsreihe dieser Arbeit bedanken. Auch bedanken möchte ich mich bei den weiteren Mitarbeitern der Klinik für Zahnärztliche Prothetik, den zahnmedizinischen Fachangestellten und dem Zahntechniker Herrn Rübben.

Von ganzem Herzen bedanke ich mich bei meinem Mann Max für die Unterstützung und Motivation.

Erklärung § 5 Abs. 1 zur Datenaufbewahrung

Hiermit erkläre ich, dass die dieser Dissertation zu Grunde liegenden Originaldaten

- In der **Klinik für Zahnärztliche Prothetik und Biomaterialien, Zentrum für Implantologie** des Universitätsklinikums Aachen

hinterlegt sind.

Eidesstattliche Erklärung gemäß § 5 Abs. (1) und § 11 Abs. (3)

12. der Promotionsordnung

Hiermit erkläre ich, **Katharina Svenja Heinen, geborene Merkel** an Eides statt, dass ich folgende in der von mir selbstständig erstellten Dissertation „**Mehrwert eines digitalen Analysesystems für die Bewertung von Zahnpräparationen für eine vollverblendete Krone in der vorklinischen Ausbildung**“ dargestellten Ergebnisse erhoben habe: Bei der Durchführung der Arbeit hatte ich folgende Hilfestellungen, die in der Danksagung angegeben sind:

	Doktorandin	Prakt. Betreuerin Dr. med. dent. Anne Barbara Rittich MME	Doktorvater Univ. -Prof. Dr. med. dent. Stefan Wolfart	Summe (%)
Studienüberwachung		30	70	100
Studiendesign/Konzeption	60	20	20	100
Datenerhebung	80	20		100
Datenauswertung	100			100
Durchführung der Experimente	100			100
Statistische Auswertung	50	50		100
Bereitstellung von Materialien			100	100
Interpretation der Datenauswertung	60	20	20	100

Unterschrift der Doktorandin Katharina Svenja Heinen, geborene Merkel

Als Betreuer der obigen Dissertation bestätige ich die Angaben von Katharina Svenja Heinen

(Univ.-Prof. Dr. med. dent. Stefan Wolfart)