

Aus dem Institut für Zahn-, Mund- und Kieferheilkunde, Abteilung für
Zahnärztliche Prothetik, Alterszahnmedizin und Funktionslehre
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Untersuchungen zur möglichen Steigerung der Beurteilungsqualität von
Präparationsergebnissen mittels einer Analysesoftware im Vergleich zur
klassischen Beurteilung durch wissenschaftliche Mitarbeiter in der studentischen
Ausbildung

zur Erlangung des akademischen Grades

Doctor medicinae dentariae (Dr. med. dent.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Hannah Mirja Leonie Grabowski

aus Hamm

Datum der Promotion: 01.03.2019

INHALTSVERZEICHNIS

1. VERZEICHNISSE	5
1.1. Tabellenverzeichnis	5
1.2. Abbildungsverzeichnis	6
2. ZUSAMMENFASSUNG	8
2.1. Abstrakt	8
2.2. Abstract	10
3. EINLEITUNG	11
3.1. Hinführung zum Thema	11
3.2. Literaturübersicht	13
3.2.1. Mangelhafte Beurteilung	13
3.2.2. Lösungsansätze	14
3.2.3. Direkter Vergleich Analysesoftware und wissenschaftliche Mitarbeiter	16
3.2.4. Vergleich zwischen Schule und vorklinisch-universitärer Ausbildung	18
3.2.5. Das Problem der gerechten Beurteilung.....	18
3.2.6. Gütekriterien für die schulische Leistungsmessung und -beurteilung	19
3.2.7. Systematische Beeinflussungstendenzen der Beurteilung	20
3.2.8. Beobachtungs- und Beurteilungsfehler.....	22
3.2.9. Gegenmaßnahmen zur Verringerung systematischer Beurteilungsfehler.....	24
3.3. Eigene Fragestellung	26
3.3.1. Hypothesen	26
4. METHODIK	27
4.1. Versuchsteilnehmer	27
4.2. Material	29
4.3. Präparationsdurchführung	31
4.4. Versuchsdurchführung	33
4.5. Fragebögen	33
4.5.1. Fragebögen der Studenten	34
4.5.2. Fragebögen der Zahnärzte.....	35
4.6. Beurteilungsbögen	37
4.7. Beurteilungsdurchführung	41
4.8. Vorgehensweise/Bedienung der Analysesoftware	41
4.8.1. CEREC	41

4.8.2. prepCheck 1.1.....	42
4.9. Statistische Auswertung.....	47
5. ERGEBNISSE	49
5.1. Reliabilität der Beurteiler in der Notengebung/Intrarater-Reliabilität.....	49
5.2. Reliabilität der zahnärztlichen Beurteiler in der Notengebung/Interrater-Reliabilität.....	55
5.3. Reliabilität der Beurteiler hinsichtlich des Bestehens/Intrarater-Reliabilität	55
5.4. Reliabilität der zahnärztlichen Beurteiler hinsichtlich des Bestehens/Interrater-Reliabilität.....	56
5.5. Reliabilität der Beurteiler in den Notenkategorien/Intrarater-Reliabilität	56
5.6. Reliabilität der zahnärztlichen Beurteiler in den Notenkategorien/Interrater-Reliabilität.....	60
5.7. Reliabilität der zahnärztlichen Beurteiler in den Notenkategorien hinsichtlich des Bestehens/Interrater-Reliabilität	61
5.8. Fragebögen der Studenten	62
5.9. Fragebögen der Zahnärzte	64
6. DISKUSSION	66
6.1. Material und Auswahl der Präparationsform.....	66
6.1.1. Zahnmodelle	66
6.1.2. Präparation	67
6.1.3. Verwendung der CEREC-Aufnahmeeinheiten	67
6.2. Beurteilungssystem	67
6.2.1. Beurteilungsbogen prepCheck 1.1	67
6.2.2. Beurteilungsbogen zahnärztliche Beurteiler	68
6.2.3. Beurteilungsdurchführung.....	68
6.3. Statistische Auswertung.....	69
6.4. Diskussion der Ergebnisse	69
6.4.1. Vergleich der Intrarater-Reliabilität.....	69
6.4.2. Vergleich der Interrater-Reliabilität der zahnärztlichen Beurteiler	70
6.4.3. Bestehen/Nichtbestehen.....	70
6.4.4. Berufserfahrung.....	71
6.4.5. Schwierigstes Kriterium/Notenkategorie.....	71
6.4.6. Ermüdung	72
6.4.7. Beurteilungsmuster	73

6.5. Diskussion der Fragebögen der Studenten	74
6.6. Diskussion der Fragebögen der Zahnärzte	77
6.7. Fehler im Umgang mit CEREC/prepCheck 1.1	82
6.7.1. Folgen selbst eingestellter Parameter	82
6.7.2. Weitere Limitationen der computergestützten Beurteilung.....	83
6.8. Optimierungsoptionen für CEREC/prepCheck 1.1	86
6.8.1. Analyse der Hypothesen.....	88
6.9. Schlussfolgerung.....	89
7. LITERATURVERZEICHNIS	90
8. EIDESSTATTLICHE VERSICHERUNG	97
9. LEBENSLAUF	98
10. DANKSAGUNG.....	99
11. APPENDIX	100

1. VERZEICHNISSE

1.1. Tabellenverzeichnis

Tabelle 1: Berufserfahrung und Angestelltenzeit der Beurteiler	27
Tabelle 2: Vorgaben für die Vollgusspräparation	31
Tabelle 3: Beurteilungssystem prepCheck 1.1	40
Tabelle 4: Interpretationseinteilung der Reliabilitäts-Koeffizienten.....	48
Tabelle 5: Weighted Kappa-Koeffizienten aller Beurteiler	54
Tabelle 6: AC1-Koeffizienten der Notengebung der zahnärztlichen Beurteiler.....	55
Tabelle 7: Cohens Kappa-Koeffizienten aller Beurteiler hinsichtlich des Bestehens.....	55
Tabelle 8: AC1-Koeffizienten aller Beurteiler hinsichtlich des Bestehens.....	56
Tabelle 9: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „P“	56
Tabelle 10: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „A“	57
Tabelle 11: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „L“	57
Tabelle 12: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „K“	58
Tabelle 13: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „B“	58
Tabelle 14: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „M“	59
Tabelle 15: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien aller Beurteiler.....	59
Tabelle 16: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 1. Durchgang	60
Tabelle 17: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 2. Durchgang	60
Tabelle 18: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 1. Durchgang hinsichtlich des Bestehens	61
Tabelle 19: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 2. Durchgang hinsichtlich des Bestehens	61
Tabelle 20: Konfiguration der Präparationswinkel	113

1.2. Abbildungsverzeichnis

Abbildung 1: CEREC AC-Aufnahmeeinheit	28
Abbildung 2: Studiomodell Basic	29
Abbildung 3: Präparationsinstrumente.....	30
Abbildung 4: Markierte Zähne	30
Abbildung 5: Fertiggestellte Präparation (Goldstandard)	31
Abbildung 6: Vervielfältigter Goldstandard	32
Abbildung 7: Beurteilungsbogen für die zahnärztlichen Beurteiler.....	38
Abbildung 8: Okklusallinie	42
Abbildung 9: Präparationsrand.....	43
Abbildung 10: Oberflächenbeschaffenheit.....	43
Abbildung 11: Hinterschnitt.....	44
Abbildung 12: Präparationsrand.....	44
Abbildung 13: Randabschlusstyp.....	45
Abbildung 14: Substanzabtrag	46
Abbildung 15: Präparationswinkel.....	47
Abbildung 16: Absolute Häufigkeiten der Notengebung von „P“	49
Abbildung 17: Absolute Häufigkeiten der Notengebung von „A“	50
Abbildung 18: Absolute Häufigkeiten der Notengebung von „L“	51
Abbildung 19: Absolute Häufigkeiten der Notengebung von „K“	52
Abbildung 20: Absolute Häufigkeiten der Notengebung von „B“	53
Abbildung 21: Absolute Häufigkeiten der Notengebung von „M“	54
Abbildung 22: Falsch dargestellter zervikaler Interdentalraum	83
Abbildung 23: Versatz zum Goldstandard mit Distanzmessung	85
Abbildung 24: Versatz zum Goldstandard	85
Abbildung 25: Fragebogen der Studenten Gruppe prepCheck	103
Abbildung 26: Fragebogen der Studenten Gruppe Assistent	106
Abbildung 27: Fragebogen der Zahnärzte.....	111
Abbildung 28: Kalibrierungshilfe für die Note sehr gut	114
Abbildung 29: Kalibrierungshilfe für die Note akzeptabel	114
Abbildung 30: Kalibrierungshilfe für die Note verbesserungswürdig	115

Im Sinne einer besseren Lesbarkeit wurde auf die zeitgleiche Verwendung der weiblichen und männlichen Sprachformen verzichtet. Alle verwendeten Personenbezeichnungen gelten gleichermaßen für beide Geschlechter.

2. ZUSAMMENFASSUNG

2.1. Abstrakt

Einleitung: Ziel dieser Studie war es, die Beurteilungsqualität bei der Benotung von Präparationen – im Hinblick auf Reliabilität und Objektivität – im Rahmen der zahnmedizinischen universitären Ausbildung zu untersuchen. Dabei wurde die Reliabilität bei Präparationsergebnissen von fünf zahnärztlichen Beurteilern im Vergleich zu der prepCheck 1.1 (Sirona) Analysesoftware als wesentlicher Bestandteil einer fairen Beurteilung bestimmt. Weiterhin war es Gegenstand der Arbeit, mit Hilfe von Fragebögen die Rahmenbedingungen in der vorklinischen Ausbildung an der Zahnklinik Würzburg sowohl aus Studenten- als auch aus Betreuersicht darzustellen.

Methodik: 43 Studenten der Universität Würzburg erstellten eine Präparation für eine Vollgusskrone am Zahn 26. Diese wurde in jeweils vier Beurteilungskategorien (Präparationsgrenze, Oberfläche, Substanzabtrag und Präparationswinkel) von fünf zahnärztlichen Beurteilern und der Analysesoftware prepCheck 1.1 mit Hilfe eines vierstufigen Notensystems zweimal beurteilt. Zur Ermittlung der Reliabilität kamen Cohens Kappa (κ), Cohens Weighted Kappa (κ_w) sowie der Gwets AC1-Koeffizient (AC1) zum Einsatz. Dabei wurden die Notengebung, die Bestehensrate sowie die Beurteilungskategorien sowohl auf die Intrarater- als auch auf die Interrater-Reliabilität hin untersucht.

Ergebnisse: Notengebung: In der Intrarater-Reliabilität erreichte die Analysesoftware den höchsten Reliabilitätskoeffizienten ($\kappa_w = 0,817$), die zahnärztlichen Beurteiler erreichten Werte zwischen $\kappa_w = 0,42$ und $\kappa_w = 0,78$. Die Interrater-Reliabilität der zahnärztlichen Beurteiler lag bei $AC1 = 0,354 - 0,371$.

Bestehen: In der Intrarater-Reliabilität erreichte die Analysesoftware den höchsten Reliabilitätskoeffizienten (Cohens Kappa $\kappa = 0,9$), die zahnärztlichen Beurteiler erreichten Werte zwischen $\kappa = 0,486$ und $\kappa = 0,822$. Die Interrater-Reliabilität der zahnärztlichen Beurteiler lag bei $AC1 = 0,671 - 0,715$. Die Interrater-Reliabilität der zahnärztlichen Beurteiler war am höchsten in der Kategorie Oberfläche ($AC1 = 0,924 - 0,951$).

Notenkategorie: In der Intrarater-Reliabilität erreichte die Analysesoftware den höchsten Reliabilitätskoeffizienten in der Notenkategorie Substanzabtrag ($\kappa_w = 0,907$).

Drei zahnärztliche Beurteiler erlangten die höchsten Kappa-Koeffizienten in der Kategorie Präparationsgrenze ($\kappa_w = 0,6 - 0,679$). Ein Beurteiler erreichte den höchsten Wert in der Kategorie Oberfläche ($\kappa_w = 0,873$), ein weiterer in der Kategorie Substanzabtrag ($\kappa_w = 0,488$).

Die Interrater-Reliabilität der zahnärztlichen Beurteiler war am höchsten in der Kategorie Oberfläche ($AC1 = 0,516 - 0,546$).

Schlussfolgerungen: Die Ergebnisse dieser Studie zeigen, dass die Reliabilität der zahnärztlichen Beurteiler bei jeder Analyse hinter der Reliabilität der Analysesoftware zurückbleibt. Darüber hinaus unterliegt das Reliabilitätsniveau innerhalb der zahnärztlichen Beurteiler erheblichen Schwankungen. Die Auswertung der Fragebögen zeigt, dass sowohl die Studenten als auch die zahnärztlichen Beurteiler zum Teil mit der gegenwärtigen Ausbildungssituation unzufrieden sind.

2.2. Abstract

INTRODUCTION: This study investigates the quality of evaluation in studies of dental education, by examining the reliability in preparation deliverables by five faculty members in comparison to prepCheck 1.1 (Sirona) learning software as an essential element in fair appraisals.

Object of this study is also illustrating the frame conditions in preclinical training at dental clinic from the perspective of the students as well as the evaluator's perspective, by using questionnaires.

METHODS: 43 students from University of Würzburg created a preparation for a full crown on tooth 26. The preparations were evaluated twice on four categories (reduction, surface finish, taper and margin width) by five faculty members and the learning software, using a four-staged scale. For determination the reliability Cohen's kappa (κ), Cohen's weighted kappa (κ_w) and the Gwet's AC1-coefficient (AC1) were used.

The grading, the number of success as well as the rating category therefore were examined on their inter- and intra-rater reliability.

RESULTS: Grading: In case of reliability, the learning software achieved the highest reliability coefficient ($\kappa_w = 0,817$) while the faculty members valued $\kappa_w = 0,42$ to $\kappa_w = 0,78$. The inter-rater reliability of faculty members is located by $AC1 = 0,354 - 0,371$.

Passing: The learning software gained the highest coefficient in intra-rater reliability ($\kappa = 0,9$) while the member's value achieved $\kappa = 0,486$ and $\kappa = 0,822$. The supervisor's inter-rater reliability scored $AC1 = 0,671$ up to $AC1 = 0,715$. The faculty member's inter-rater reliability is the highest in the category surface finish ($AC1 = 0,924 - 0,951$).

Grading-criteria's: In intra-rater reliability the learning software achieved the highest reliability-coefficient in reduction ($\kappa_w = 0,907$). Three instructors achieved the highest kappa-coefficients in reduction ($\kappa_w = 0,6 - 0,679$). One of them gained the highest level in the category surface finish ($\kappa_w = 0,873$), another one in reduction ($\kappa_w = 0,488$). The faculty member's inter-rater reliability is the highest in the category surface finish ($AC1 = 0,516 - 0,546$).

CONCLUSION: This study illustrates that the reliability of faculty members is significantly lower than the learning software's reliability. Furthermore the level of reliability within the faculty members is subject to considerable fluctuations. The questionnaires' evaluation shows that the students and the evaluators as well are partly dissatisfied with the current educational situation.

3. EINLEITUNG

3.1. Hinführung zum Thema

Die grundlegende Herausforderung in der vorklinischen universitären Ausbildung angehender Zahnärzte ist die frühe Vermittlung praktischer Fertigkeiten als essentieller Bestandteil des späteren Berufsbildes. Die Vielfältigkeit des Berufes, welche viele unterschiedliche praktische Tätigkeiten aus verschiedenen Bereichen beinhaltet, wird den Studenten zum Großteil innerhalb der ersten fünf Semester im vorklinischen Abschnitt – bis zum Abschluss der zahnärztlichen Vorprüfung (Physikum) – demonstriert und vermittelt.

Die vorliegende Studie untersuchte ausschließlich die Situation der vorklinischen Ausbildung am Zentrum für Zahn-, Mund- und Kiefergesundheit der Julius-Maximilians-Universität Würzburg im Mai 2015.

Primär steht während des Studiums die Vermittlung der idealen Kriterien und Merkmale einer praktischen Arbeit im Vordergrund. Es handelt sich dabei sowohl um zahnärztliche Tätigkeiten, wie die Präparation von Zähnen, Abformübungen und Kieferrelationsbestimmungen, als auch um zahntechnische Übungen, wie das Ausgießen von Modellen, Aufwachsübungen sowie die Herstellung von Kronen, Brücken und unterschiedlichen Prothesen.

In der Vorklinik werden die Studenten zunächst theoretisch im Rahmen von Vorlesungen und unter Zuhilfenahme schriftlicher Aufzeichnungen bezüglich der Abläufe und Herstellungsprozesse der zu erlernenden Tätigkeiten unterrichtet. Anschließend trainieren sie in drei aufeinander aufbauenden praktischen Kursen – Technisch Propädeutischer Kurs, Phantomkurs der Zahnersatzkunde 1 und 2 – im vorklinischen Laboratorium ihre feinmotorischen Fähigkeiten durch Anwendung der vermittelten Merkmale an praktischen Arbeiten. Temporär begleitet durch zusätzliche Videodemonstrationen sowie direkte individuelle Demonstrationen an der Simulationseinheit und unterstützende Hinweise eines Lehrzahntechnikers, arbeiten die Studenten an der Universität Würzburg den Großteil der Zeit ohne direkte Betreuung an ihrem Laborarbeitsplatz. Zwischenschritte oder fertig gestellte praktische Arbeiten werden den wissenschaftlichen Mitarbeitern, bei denen es sich um approbierte Zahnärzte unterschiedlicher Erfahrungsstufen des Fachbereiches Zahnärztliche Prothetik handelt, vorgestellt. Es sind bis zu fünf wissenschaftliche Mitarbeiter im Wechsel für die Kurse zuständig, von denen immer zwei gleichzeitig präsent sind. Ihnen obliegt die Betreuung der etwa 50 Studenten pro Semester^a. Meist befinden sich die wissenschaftlichen

^a Information auf schriftliche Nachfrage bei der Stabsstelle für Presse- und Öffentlichkeitsarbeit der Julius-Maximilians-Universität Würzburg am 27.06.2018

Mitarbeiter während der Kurszeiten an Schreibtischen, da die Studenten sie nahezu permanent für Rückfragen aufsuchen. Nur selten ergibt sich für diese die Möglichkeit, selbst auf die Studenten zuzugehen und diese proaktiv zu betreuen. Entsprechende Wartezeiten und die regelmäßig wechselnde Betreuung erschweren zudem einen effizienten Ablauf. Darüber hinaus müssen sowohl die erbrachten Leistungen als auch die erforderlichen Zwischenschritte der Studenten in einem Testatheft schriftlich durch die wissenschaftlichen Mitarbeiter bestätigt werden. Diese können während des Herstellungsprozesses ihrer praktischen Arbeit nicht den nächsten Schritt beginnen, solange ein vorheriger Zwischenschritt nicht gegengezeichnet worden ist. Dadurch verkürzt sich wiederum die verfügbare Arbeitszeit für die häufig in ihrer Komplexität ansteigenden folgenden Arbeitsschritte. Diese Maßnahme soll zum einen dazu beitragen, dass die Studenten eine gewissenhafte, gründliche und saubere Arbeitsweise erlernen. Zum anderen soll so verhindert werden, dass praktische Arbeiten aus dem vorklinischen Laboratorium entfernt und nach externer Anfertigung wieder in die Zahnklinik gebracht werden. Diese Vorgehensweise bietet auch Spielraum für Willkür und Subjektivität, da die Beurteilungen überwiegend durch optische Begutachtung vorgenommen werden. Eine zahnärztliche Sonde kann zusätzlich zur taktilen Beurteilung herangezogen werden.

Die Benotung der vorab homogen und zufällig verteilten Arbeiten erfolgt in der Regel durch die fünf wissenschaftlichen Mitarbeiter, welche auch die vorklinisch praktischen Kurse betreuen, sowie durch den fachlichen Leiter (Oberarzt)¹. Die Beurteilung verläuft in chronologischer Reihenfolge, wobei die jeweils dazugehörigen Testathefte der Studenten eine namentliche Zuordnung und die Einsicht in den Notenspiegel ermöglichen¹. Ein Vergleich während des Benotungsprozesses mit anderen Arbeiten zur Einordnung der Leistung findet dabei nicht statt¹. Eine vorherige Absprache der wissenschaftlichen Mitarbeiter erfolgt in Bezug auf die wesentlichen Beurteilungskriterien im Sinne einer kurzen Einstimmung¹. Der komplette Beurteilungsablauf findet – wie auch schon bei der Betreuung im vorklinischen Kurs – unter Ausschluss einer vorher festgelegten schriftlichen Beurteilungsmatrix statt¹. Abschließend sichtet der Oberarzt als fachlicher Leiter noch einmal alle Noten und beurteilt sehr gute und sehr schlechte Arbeiten ein zweites Mal¹. Bei einer überdurchschnittlichen Durchfallquote ist es dem Oberarzt erlaubt, selbstständig eine Wiederholung anzuordnen¹.

3.2. Literaturübersicht

3.2.1. Mangelhafte Beurteilung

Die Beurteilung praktischer Arbeiten durch betreuende Personen stellt – im Hinblick auf mangelnde Objektivität und Reliabilität – ein bereits häufig beschriebenes Problem in der universitären Ausbildung dar. Zahlreiche Untersuchungen zeigen signifikante Unterschiede in der Interrater- und/oder der Intrarater-Übereinstimmungsrate²⁻⁹.

Ferner werden auch die negativen Auswirkungen der mangelhaften Beurteilungslage auf die Studenten von verschiedenen Autoren beschrieben. Die Problematik wurde bereits in den 1930er Jahren dargestellt und das Phänomen der defizitären Beurteilungslage als Ursache für Verwirrungen und Stress unter den Studenten bezeichnet¹⁰. Weitere Autoren beschreiben, dass vage und unzuverlässige Beurteilungen zu Unzufriedenheit bei Studenten und betreuenden Personen führt^{11, 12}. Darüber hinaus wird die willkürliche, unsichere und inkonstante Benotungsweise der Betreuer von den Studenten bewusst wahrgenommen und verleitet diese dazu, sich nur noch auf die Erreichung guter Noten zu fokussieren⁷. Die Studenten erkennen im Kursalltag schnell, welche Fakultätsangestellte sich durch eine strenge beziehungsweise milde Beurteilungsweise auszeichnen⁷. Zudem beeinflusst die Erkenntnis, dass einzelne Kommilitonen bevorzugt behandelt oder diskriminiert werden sowie eine vorherrschende verbalisierte und/oder stumme Willkür den Lernprozess aller Studenten⁷. Als Resultat sinken das Selbstbewusstsein der Studenten und infolge dessen auch deren Leistungsbereitschaft¹³.

Die folgenden Studien untersuchten zusätzlich weitere Aspekte bei der Beurteilung durch Fakultätsangestellte.

Einfluss Berufserfahrung der Beurteiler:

SHARAF et al. (2006) ließen in ihrer Untersuchung Präparationen von Studenten durch drei Fakultätsangestellte mit mehr als zehn Jahren klinischer und ausbildnerischer Erfahrung innerhalb von drei Tagen zweimal beurteilen¹³. Dabei stellten sie eine signifikante Variabilität in der Interrater-Beurteilerqualität fest¹³. Allerdings konnte keine signifikante Variabilität in der Intrarater-Beurteilungsqualität nachgewiesen werden¹³.

Eine weitere Studie stellt fest, dass – obwohl zwei der eingesetzten Fakultätsangestellten mehr als zehn Jahre Berufs- und Ausbildungserfahrung aufwiesen – dies keinen Hinweis auf eine erhöhte Reliabilität liefert¹⁴.

Einfluss einer Ermüdung der Beurteiler:

Die Studie von AL AMRI et al. (2016) zur Ermittlung des Effektes von Ermüdung und Erfahrung der betreuenden Personen auf die Beurteilungsqualität beschreibt, dass jüngere Beurteiler eine Ermüdung besser kompensieren können und eher zur Vergabe besserer Noten tendieren¹⁴.

Bestehen einer Prüfung:

In der Studie von JENKINS et al. (1998) wurden Präparationen, die hauptsächlich durch Studenten, zum Teil aber auch durch erfahrene Fakultätsangestellte ausgeführt wurden, von fünf Universitätsmitarbeitern unterschiedlicher Berufs- und Ausbildungserfahrung beurteilt³. Die Ergebnisse zeigen sowohl ein hohes Maß an Intra- als auch an Interrater-Variabilität³. In einigen Fällen erhielten Präparationen in einem Durchgang ein Urteil, welches zum Bestehen ausgereicht hätte, während in einem anderen Durchgang die erteilte Zensur ein Nichtbestehen bedeutete³. Zu vergleichbaren Ergebnissen kommt auch eine zweite Untersuchung bei der Benotung der Zähne nach einem einfachen zweistufigen Notensystem (akzeptabel/inakzeptabel)¹⁵. Auch hier wurden viele diskordante Urteile festgestellt¹⁵.

Schwierigstes Kriterium/Notenkategorie:

Nach MAYS et al. (2016) stellt die Beurteilung des Präparationswinkels einer Präparation eine besonders große Herausforderung für die Fakultätsangestellten dar¹⁶. Im Vergleich zwischen menschlichen Beurteilern und der E4D-Software fanden sie heraus, dass die CAD/CAM basierte Technologie der Analysesoftware den Präparationswinkel wesentlich präziser bestimmen kann¹⁶.

3.2.2. Lösungsansätze

Folgende Lösungsansätze zur Verbesserung der Beurteilungsqualität in der zahnärztlichen Ausbildung wurden bereits untersucht.

Kalibrierung der Beurteiler:

Die Nutzung von Kalibrierungseinheiten und/oder die Verwendung von schriftlich fixierten Leistungsanforderungen zur Verbesserung der Objektivität sind sowohl im pädagogischen Handlungsfeld als auch in der universitären zahnärztlichen Ausbildung ein möglicher Lösungsansatz.

So empfehlen zwei Studien generell mehr Personalschulungen unter den Beurteilern und die Einführung eines einheitlichen Beurteilungssystems^{3,17}. Weitere Autoren stellten in Untersuchungen fest, dass ein analytisches Vorgehen mit Checklisten die Beurteilerobjektivität

verbessert¹⁸⁻²³. Auch eine gezielte Kalibrierung der Fakultätsangestellten zur Optimierung der Reliabilität wird für angemessen gehalten²⁴.

HAJ-ALI und FEIL (2006) untersuchten zudem die Langzeiteffekte von Beurteilerkalibrierungen¹⁵. Dabei analysierten sie die Interrater-Reliabilität von Fakultätsangestellten vor einer durchgeführten Kalibrierung, direkt danach und zehn Wochen später¹⁵. Die Kalibrierung bestand aus einem Vergleich der zu beurteilenden Präparation mit einem Goldstandard¹⁵. Die Übereinstimmung zwischen den Beurteilern lag vor der Kalibrierung bei 54,5%, unmittelbar danach bei 66,9% und zehn Wochen später bei 64,6%¹⁵. Dies zeigt, dass ein Vergleich mit einem Goldstandard die Beurteilerreliabilität verbessert und dieser Effekt über einen Zeitraum von zehn Wochen nur geringfügig nachlässt¹⁵.

AHMED et al. (2016) ließen in ihrer Studie 32 Präparationen von acht Fakultätsangestellten beurteilen²⁵. Die Übereinstimmungsrate lag dabei bei unter 60%²⁵. Es folgte eine 30-minütige Kalibrierung unter Verwendung von 3D-Modellen und Beispielfotos verschiedener Notenabstufungen²⁵. Die Beurteilung wurde danach zweimal wiederholt; zunächst unmittelbar nach der Kalibrierung und anschließend nach einer Zeitspanne von sechs Monaten ohne Durchführung einer erneuten Kalibrierung²⁵. Die Ergebnisse zeigen, dass die interkollegiale Übereinstimmung nach der Kalibrierung höher ist und der Effekt nach Ablauf von sechs Monaten zwar nachlässt, aber immer noch höher liegt als die Übereinstimmung vor einer Kalibrierung²⁵.

AL AMRI et. al (2016) beschreiben in einer Studie den Effekt von Ermüdung und Erfahrung der Fakultätsangestellten auf die Beurteilungsqualität¹⁴. Die Benotungen wurden dabei sowohl unter Zuhilfenahme von schriftlich fixierten Leistungskriterien als auch ohne jegliche Hilfsmittel durchgeführt¹⁴. Die erste Beurteilung fand nach einem achtstündigen Arbeitstag statt, die zweite Beurteilung direkt zu Beginn des darauffolgenden Tages. Sie fanden heraus, dass die Beurteilung ohne Hilfsmittel stärker durch den Faktor Ermüdung beeinflusst wird als Beurteilungen, die unter Verwendung von Hilfsmitteln getroffen werden¹⁴. Ferner werden bessere Noten bei einer Verwendung von Hilfsmitteln vergeben¹⁴.

Im Gegensatz dazu können zwei weitere Studien keinen Unterschied zwischen der herkömmlichen Beurteilungsweise durch alleinige optische Betrachtung und dem Vorgehen mit einer Checkliste feststellen^{13, 26}.

Einsatz computerbasierter Ausbildungssysteme:

Auch der Einsatz computerbasierter Ausbildungssysteme wird in verschiedenen Studien thematisiert. Eine Untersuchung beschreibt, dass allein durch den Einsatz eines digitalen Messgerätes zur Ermittlung der Tiefe, Glättung und Ebenheit des Kavitätenbodens eine signifikante Verbesserung der Beurteilerreliabilität im Gegensatz zur subjektiven Beurteilung durch Fakultätsangestellte ermöglicht wird²⁷.

Der Einzug von CAD/CAM-Systemen in die zahntechnischen Labore und Praxen ebnete den Weg der digitalen Beurteilung von Präparationen in der zahnmedizinischen Ausbildung.

In Verbindung mit CAD-unterstützten Lehrprogrammen – welche es ermöglichen, präparierte Zähne zu beurteilen – können Teile der praktischen Ausbildung von Studenten digitalisiert werden. Die Unterteilung der Systeme gliedert sich in CAI (Computer-aided Instruction), CAL (Computer-assisted Learning) und CAS (Computer-assisted Simulation)²⁸. Die genannten computerbasierten Auswertungssysteme ermöglichen eine objektive Benotung präparierter Zähne, können betreuende Personen entlasten und führen somit langfristig zu Kosteneinsparungen^{29,30}. Zahlreiche Studien weisen den positiven Nutzen digitaler Ausbildungstechnik nach²⁹⁻³⁶. Ferner wurde die Einstellung von Zahnmedizinstudenten gegenüber dem Einsatz verschiedener technischer Neuerungen untersucht und mehrheitlich eine Akzeptanz der Studenten festgestellt^{34,37-43}. Es muss bei der Einführung dieser Systeme mit einem zeitlichen Mehraufwand in der Anfangsphase gerechnet werden, um sowohl den zahnärztlichen Betreuern als auch den Studenten die Möglichkeit zu geben, einen sicheren Umgang mit der zur Verfügung gestellten Technologie zu erlernen⁴²⁻⁴⁴. Außerdem sind die hohen Investitionskosten zu bedenken, welche sich durch die langfristig mögliche Reduktion des vorklinischen Ausbildungspersonals amortisieren beziehungsweise eine Fokussierung auf die direkte Betreuung der Studenten ermöglichen⁴¹.

3.2.3. Direkter Vergleich Analysesoftware und wissenschaftliche Mitarbeiter

Die folgenden drei Studien vergleichen die Beurteilungsqualität von Fakultätsangestellten mit computerbasierten Beurteilungssystemen hinsichtlich ihrer Reliabilität.

KATEEB et al. (2016) ließen Kronenpräparationen von Studenten durchführen und diese anschließend von vier Fakultätsangestellten und einer Analysesoftware (Nissin Fair Grader 100) beurteilen⁴. Zur Ermittlung der Reliabilität wurde die ICC (Inter-Class-Correlation) bestimmt⁴.

Die ermittelten Ergebnisse der Intrarater-Reliabilität zweier zahnärztlicher Betreuer liegen bei 0,73 und 0,78, die der Analysesoftware bei 0,99⁴. Die festgestellte Interrater-Reliabilität der

zahnärztlichen Betreuer beträgt $ICC = 0,76^4$. Die Übereinstimmungsrate zwischen den Noten der Fakultätsangestellten und den Noten der Analysesoftware wird mit einer mittleren Übereinstimmung im unteren Bereich angegeben⁴.

KWON et al. (2014) ermittelten ebenfalls die Intra- und Interrater-Reliabilität bei Beurteilungen durch Fakultätsangestellte und die Intrarater-Reliabilität einer digitalen Analysesoftware (E4D-Software)⁵. Untersucht wurden dabei Aufwuchsübungen⁵. Zwei Betreuer mit mehr als zehn Jahren Tätigkeit in der Ausbildung beurteilten alle Arbeiten unter Zuhilfenahme von Checklisten und Beurteilungsskalen⁵. Eine erneute Beurteilung fand nach sieben Tagen statt⁵. Zur Ermittlung der Reliabilität wurde ebenfalls die ICC verwendet⁵. Die ermittelte Intrarater-Reliabilität der E4D-Software liegt über einem ICC von $0,9^5$. Die Intrarater-Reliabilität der Fakultätsangestellten liegt bei $ICC = 0,71$ beziehungsweise $ICC = 0,69^5$. Die Interrater-Reliabilität liegt noch niedriger: im ersten Durchgang bei $ICC = 0,37$ und im zweiten Durchgang bei $ICC = 0,24^5$.

Auch RENNE et al. (2013) untersuchten das E4D-System und verglichen es mit zahnärztlichen Beurteilungsleistungen⁷. Vorab fand eine zweistündige Notenkalisierung der Fakultätsangestellten anhand von Beispielpräparationen im Vergleich zum Goldstandard statt⁷. Die Ergebnisse zeigen eine Intrarater-Reliabilität der zahnärztlichen Betreuer von $ICC = 0,62$ und der E4D-Software von $ICC = 0,98^7$.

Trotz der bekannten Problematik – insbesondere das Fehlen von Zuverlässigkeit, Nachvollziehbarkeit und Konstanz bei der Notenvergabe – stellt die hauptsächlich auf subjektiven Parametern beruhende und durch Fakultätsangestellte durchgeführte Ausbildung bis dato die noch immer am häufigsten verwendete Beurteilungsmethode im Rahmen der universitären zahnmedizinischen Ausbildung dar^{29, 45}.

3.2.4. Vergleich zwischen Schule und vorklinisch-universitärer Ausbildung

Im Vorfeld soll erwähnt werden, dass ein direkter Vergleich des schulischen Lehrsystems mit der universitären vorklinischen Ausbildung angehender Zahnärzte alleine nicht zielführend wäre. Die für die schulische Bildung prägende pädagogische Komponente kommt innerhalb des zahnmedizinischen Studiums nicht zum Tragen. Des Weiteren finden Beurteilungen mündlicher Leistungen als wichtiger Bestandteil der schulischen Notenfindung in der vorklinischen zahnmedizinischen Ausbildung ebenfalls keine Verwendung.

Allen Bildungseinrichtungen ist jedoch gemein, dass sie Beurteilungen zur Erlangung einer Berechtigung für einen jeweils weiterführenden Abschnitt verwenden⁴⁶. Aufgrund dessen ergibt sich eine indirekte Vergleichbarkeit der beiden Systeme, da beiden die Notwendigkeit eines Beurteilungsprozesses obliegt, der zum einen das gleiche diagnostische Mittel in Form eines mehrstufiges Notensystems beinhaltet, der zum anderen aber in beiden Systemen einen schon lange kritisierten Zustand darstellt^{10,47}. Darüber hinaus ist auch das Verhältnis Beurteiler zu Schüler/Student vergleichbar, da sowohl Lehrer als auch wissenschaftliche Mitarbeiter sich im Vorfeld einen Eindruck über bereits erbrachte Leistungen und über die Persönlichkeit der zu Beurteilenden machen können. Diese Beziehung ist in der Schule deutlich ausgeprägter. Lehrer kennen ihre Schüler, die sie teilweise jahrelang unterrichten, deutlich besser. Elternsprechtage bieten zudem die Möglichkeit des Einholens weiterführender Informationen.

Im weiteren Verlauf wird die Beurteilung schriftlicher Leistungen in der Schule durch Lehrer mit der Beurteilung praktischer Leistungen durch wissenschaftliche Mitarbeiter gleichgesetzt. Die verwendeten Begriffe Schüler/Lehrer stehen in der vorliegenden Arbeit stellvertretend für Student/Beurteiler.

3.2.5. Das Problem der gerechten Beurteilung

Der Umstand einer mangelhaften Leistungsbeurteilung ist eine in der Pädagogik schon lang bekannte Problematik. Bereits zu Beginn des 20. Jahrhunderts – kurz nach der Einführung des noch heute verwendeten Leistungsbeurteilungssystems – kam heftige Kritik auf, die bis heute anhält⁴⁷. PFISTER (1981) geht sogar so weit, zu sagen, dass Lehrer niemals objektiv beurteilen können, da schon im Alltag getroffene Urteile häufig Fehler aufweisen und sich diese Problematik in an hohe Ansprüche gekoppelten Beurteilungssituationen zusätzlich verschärft⁴⁸.

WINTER (2016) sieht in der Leistungsbeurteilung zusätzlich die Option einer Disziplinierungsmaßnahme, welche stets das Risiko eines Machtmissbrauchs beinhaltet⁴⁹.

Auswirkungen auf die zu Beurteilenden:

Der Beurteilungsprozess von Klassenarbeiten, welcher in der Regel allein durch die Lehrkraft durchgeführt wird und sich so jeglicher Kontrolle entzieht, birgt die erhöhte Gefahr einer Machtausübung des Beurteilers⁴⁹. Schüler kommen automatisch zu der Schlussfolgerung, dass Beurteilungen völlig willkürlich vergeben werden, da diese meist nicht im Zusammenhang mit den erbrachten Leistungen stehen⁵⁰. Daraus resultiert häufig eine reduzierte Leistungsbereitschaft der Schüler⁵⁰. Nach KORNADT (1975) trifft FEND (1970) die Aussage, dass diese unsicheren, aber unbeeinflussbaren Notenvergaben zu einem ausgeprägten Selektionsdruck führen⁵⁰. Dieser erzeugt bei den Schülern ein Ohnmachtsgefühl, welchem mit Widerstand, Aggressivität und/oder Resignation begegnet wird⁵⁰.

3.2.6. Gütekriterien für die schulische Leistungsmessung und -beurteilung

Zu den Gütekriterien zählen Objektivität, Reliabilität und Validität⁵¹. Je mehr Gütekriterien eingehalten werden, desto gerechter sind laut der Autoren die erteilten Beurteilungen⁵¹.

Objektivität:

Als objektives Urteil bezeichnet man Beurteilungen, die allein dem Objekt entsprechen und frei von jeglichem subjektiven Einfluss sind⁵². Objektivität lässt sich dabei in folgende Teilgebiete aufteilen:

- *Durchführungsobjektivität:* Bemisst den Einfluss einer Situation und das Verhalten des Beurteilers auf die Ergebnisse⁵². Um die Durchführungsobjektivität zu erhöhen, sollten für alle Beurteiler während des Benotungsprozesses die gleichen standardisierten Bedingungen herrschen⁵². Alle vorab festgelegten Instruktionen sollten identisch sein und den Beurteilern schriftlich vorliegen⁵².
- *Auswertungsobjektivität:* Damit ist die Objektivierung einer Auswertungssituation nach vorgegebenen Kriterien gemeint⁵². So weisen Prüfungen im Multiple-Choice-Verfahren die maximale Auswertungsobjektivität auf⁵². Bei der Auswertung anderer Leistungen (offene Fragen oder praktische Arbeiten) bieten Auswertungsschlüssel als schriftlich vorliegende Checklisten oder Vergleichsmaßstäbe im Sinne einer sich immer wieder wiederholenden Kalibrierung der Beurteiler eine Möglichkeit, die Auswertungsobjektivität zu erhöhen⁵².
- *Interpretationsobjektivität:* Diese bezeichnet das Maß der Unabhängigkeit einer Zensur von subjektiven Einflüssen eines Beurteilers⁵². Häufig führt eine ungenaue Formulierung der zu beurteilenden Leistung zu großen Diskrepanzen zwischen den Beurteilern⁵². Bei

verschiedenen Beurteilern verschärft sich diese Problematik zusätzlich, da diese – wenn überhaupt – lediglich partiell übereinstimmende Beurteilungsmaßstäbe verwenden⁵².

Reliabilität:

Eine Beurteilung ist reliabel, wenn das Ergebnis mit dem erbrachten Leistungsniveau übereinstimmt⁵¹.

Validität:

Die Validität ist gleichzusetzen mit dem Begriff Gültigkeit⁵³.

- *Inhaltsvalidität:* In einer Prüfung wird nur Wissen abgefragt, welches auch vorher im Unterricht behandelt wurde⁵⁴. Zudem sollte die Übertragungsleistung von Wissen bereits erprobt worden sein⁵⁴. Ebenso sollte den Schülern das gleiche Zeitvolumen zur Verfügung stehen, um Verzerrungen in der Leistungsmessung hinsichtlich der Schnelligkeit im Lösen einer Aufgabe zu vermeiden⁵³.
- *Prognosevalidität:* Sie gibt Auskunft darüber, ob und bis zu welchem Grad sich noch anstehende aus bereits erbrachten Leistungen ableiten lassen⁵¹.

3.2.7. Systematische Beeinflussungstendenzen der Beurteilung

Die Beurteilung von praktischen Arbeiten der Studenten im vorklinischen Kurs findet zum Großteil mittels optischer Betrachtung statt. Jede menschliche Wahrnehmung unterliegt jedoch einem Selektierungsprozess⁵². Verschiedenste Kriterien beeinflussen dabei die Filterung der menschlichen Wahrnehmung⁵². So konnte die Auswirkung von Hunger und anderen Zuständen des Mangels sowie die Auswirkung von gesteigertem Interesse auf die Wahrnehmung nachgewiesen werden^{55,56}. Ein zusätzlich beeinflussender Faktor ist die Persönlichkeit des jeweiligen Beurteilers selbst⁵². Demnach neigen beispielsweise unsichere Beurteiler stärker zu einer autoritären Haltung und infolgedessen zu wenig objektiven Beurteilungen⁵².

Stress und psychische Belastung können zu Verzerrungsprozessen wie Selektion (Ausblendung), Assimilation (Vermischung) oder Akzentuierung (Hervorhebung) während einer Beurteilung führen⁵². Das Ausmaß der Wahrnehmungsverzerrung korreliert dabei direkt mit dem Ausmaß der negativen Belastung für die beurteilende Person⁵².

Bei jeder Wahrnehmung wird eine Kompatibilität von Realität und eigener Wahrnehmung geprüft, da jeder Beurteiler schon mit einer Erwartungshaltung (Hypothesen) an eine Wahrnehmung herantritt⁵⁷. Dabei benötigen schon häufig bestätigte Hypothesen nur eine geringe

Anzahl entsprechender Wahrnehmungssignale, um bestätigt zu werden⁵². Um schwache Hypothesen zu bestätigen, braucht es entsprechend eine größere Anzahl⁵².

Kausalattribution:

Eine entscheidende weitere Beeinflussung bei der Wahrnehmung ist die Tendenz zur Attribuierung⁵². Dabei werden nicht vorhandene Eigenschaften hinzugezählt, wenn sie in die Erwartungshaltung eines Beurteilers hineinpassen⁵². Als Beispiel sei folgende Situation beschrieben: Ein Prüfling, welcher zuvor im Kurs vor allem durch schlechte Arbeiten aufgefallen ist, wird mit der Vorlage einer sehr guten Arbeit besondere Kontrollen und bestimmte Kausalzuschreibungen wie die Unterstellung eines Täuschungsversuches auslösen, da seine gute Leistung nicht im Einklang mit der erwarteten Haltung der Beurteiler steht. Es wird darüber hinaus sogar gezielt nach Hinweisen gesucht, die eine solche Hypothese bekräftigen können. Dieses Vorgehen beschreibt die Grundlage für die sogenannte Sich-selbst-erfüllende-Prophezeiung: einem Prozess, bei dem die Erwartungen der Beurteiler entsprechende Verhaltensweisen der Schüler hervorrufen⁵².

Erwartungsbedingte Urteilsreaktionen:

Auf Grundlage des oben erwähnten Prinzips der Sich-selbst-erfüllenden-Prophezeiung wurde ein Schema zur Entstehung des Pygmalion-Effekts entwickelt⁵⁸. Demnach wird ein Schüler in Folge des durch den Lehrer zuvor gebildeten vermeintlich leistungsschwachen Bildes entsprechend nachteilig behandelt⁵⁸. Dieser passt sich infolgedessen im Laufe der Zeit über eine Abwertung des eigenen Selbstbildes und den Zustand der Passivität an das vom Beurteiler erschaffene Bild an⁵⁸. Dieser Effekt kann auch ins Positive erfolgen⁵⁹. Dieses Schema stellt einen sich stetig wiederholenden Prozess zwischen Schüler und Lehrer dar⁵¹. Nur eine regelmäßige Überprüfung von Urteilen seitens der Beurteiler ermöglicht die Verhinderung der Entstehung eines Pygmalion-Effektes⁵².

Implizierte Persönlichkeitstheorie:

Bei der implizierten Persönlichkeitstheorie, welche auf persönlichen Erfahrungen basiert, sind zwei Ausprägungen zu unterscheiden: die naiv-psychologische Eigenschaftstheorie und die naiv-psychologische Prozesstheorie⁵¹. Die naiv-psychologische Eigenschaftstheorie stellt eine Vermutung dar, wie sich menschliche Eigenschaften gegenseitig beeinflussen⁵² (Beispiel: „Wer handwerklich begabt ist, ist auch ein guter Zahnarzt.“). Bei der naiv-psychologischen Prozesstheorie werden Relationen zwischen Verhaltensweisen/Eigenschaften und deren

Entstehung gestellt⁵² (Beispiel: „Eine strenge vorklinische Ausbildung dient der Abhärtung und bringt gute Zahnärzte hervor.“). Somit werden Leistungsbeurteilungen im Wesentlichen von der Persönlichkeit des Beurteilers und nicht von der des Schülers beeinflusst⁶⁰.

3.2.8. Beobachtungs- und Beurteilungsfehler

- *Interaktionsfehler I*: Legt ein Schüler eine unfreundliche oder sehr freundliche Art gegenüber dem Beurteiler an den Tag, beeinflusst dies die anschließende Beurteilung⁵².
- *Nachbarschaftsfehler/Nähefehler*: Die räumliche und/oder zeitliche Nähe von Ereignissen beeinflusst deren Beurteilung⁶¹. Sehr gute Leistungen haben auf nachfolgende Beurteilungen einen positiven Effekt, schlechte wirken sich negativ aus⁵¹.
- *Weitere Reihungs- und Schwankungseffekte*: BETZ (1974) wies bei mündlichen Prüfungen ein sich wiederholendes An- und Absteigen bei der Notengebung nach⁶². Zudem wird die erste Prüfung häufig als Referenz verwendet⁵³. Nach KLEBER (1992) trifft BAURMANN (1973) die Aussage, dass bei der Beurteilung schriftlicher Leistungen – unabhängig von welchem Schüler die Arbeit stammt – die Tendenz vorliegt, die ersten nach einem strengeren Maßstab zu beurteilen⁵². Lehrkräfte versuchen ferner, die aufeinanderfolgende wiederholte Vergabe sehr guter Noten zu vermeiden⁶³.
- *Mildeeffekt/Strengeneffekt*: Beim Mildeeffekt findet eine Akzentuierung guter Leistungen bei gleichzeitiger Ignoranz gegenüber schlechten Leistungen statt, weshalb insgesamt zu viele gute Noten erteilt werden^{51,52}. Mögliche Erklärungen für diesen Effekt sind folgende:
 - Eine schlechte Beurteilung stellt die Fähigkeiten des Beurteilers in Frage⁵².
 - Die Vermeidung einer möglicherweise notwendigen Rechtfertigung bei einer zu großen Anzahl schlechter Noten⁵².
 - Mangelnder Fleiß verhindert eine gründliche Auseinandersetzung mit der Leistungsbeurteilung⁵².

Grundsätzlich lassen sich die oben genannten Beweggründe auch für die Erklärung des Strengeneffektes anwenden⁵². Darüber hinaus können folgende Faktoren einen Strengeneffekt hervorrufen:

- Die Fachkompetenz des Beurteilers soll verdeutlicht werden⁵².
- Die Schüler sollen sich intensiv mit dem Lehrstoff befassen, um schlechte Noten zu vermeiden⁵².

Laut SACHER (1994) liegt ein generelles Vorkommen von Streng- und Mildbeurteilern vor⁶⁴.

KLEBER (1976) benennt den Ausprägungsgrad der Bekanntschaft zwischen Lehrern und

Schülern als Grund für Milde- und Strengeeffekt⁶⁵. So benoten Lehrer Schüler, die sie gut kennen, milder als andere Mitschüler und umgekehrt⁶⁵. ROSEMANN (1975) hält im schulischen Umfeld vor allem den Einfluss von Antipathie als verantwortlich für die Entstehung eines Strengeeffektes⁶⁶. Zusätzlich geht er davon aus, dass Beurteiler mit einer Spezialisierung Leistungen in diesem Bereich häufig strenger beurteilen⁶⁶.

- *Logischer Fehler*: Dies ist ein Fehler, der auf der Basis eines Vorurteils (implizierte Persönlichkeitstheorien) beruht^{52, 53, 67}.
- *Haloeffekt*: Lehrer tendieren dazu, Schüler auf Basis eines Merkmales hin zu beurteilen⁵¹. Personen mit einem positiven beziehungsweise negativen Gesamteindruck werden besser beziehungsweise schlechter bewertet als es ihren Leistungen entspricht⁶⁸. Für die Entstehung reicht bereits das Vorliegen von Gerüchten aus⁶⁸. Dieser Effekt beruht ebenfalls auf den implizierten Persönlichkeitstheorien⁵².
- *Zentrale Tendenz*: Beurteiler neigen dazu, die Extreme der Notenskala auszusparen und setzen ihren Notenschwerpunkt in der mittleren Kategorie^{51, 52}. BOHL (2006) sieht die Unsicherheit der Beurteiler als Ursache hierfür an⁶⁹.
- *Dichotomisierungstendenz*: Beurteiler verwenden im Besonderen die sehr guten und sehr schlechten Noten im Spektrum⁵¹. Für die Erteilung einer sehr guten Zensur muss ein gewisser Schwellenwert erreicht werden⁵¹. Ist das nicht der Fall, fällt die Leistung automatisch in eine sehr schlechte Kategorie⁵¹. SACHER (1994) vermutet als Ursache dafür eine simultan vorliegende niedrige Begeisterungs- und Enttäuschungsschwelle der Beurteiler⁶⁴.
- *Laienfehler*: Dies beschreibt Beurteilungen, die ohne die nötige Sachkompetenz erteilt wurden⁵².
- *Generalisierungsfehler*: Ein prägendes Ereignis wird in Bezug auf sein Auftreten überschätzt und es wird ihm eine zu hohe Bedeutung beigemessen⁵².
- *Referenzfehler*: Schüler werden gewissermaßen in falsche „Schubladen“ (Referenzen) eingeordnet und entsprechend der Charakteristika dieser Referenzgruppen bewertet⁵².
- *Wissen um die Folgen-Fehler*: Beurteiler lassen sich von den möglichen Konsequenzen ihrer Beurteilung beeinflussen. Dabei treten sowohl nach oben als auch nach unten korrigierte Ergebnisse auf⁵¹.
- *Pygmalion-Effekt*: Beurteiler provozieren durch spezifische Interaktionen, welche auf Erwartungshaltungen beruhen, bestimmte Reaktionen bei den zu Beurteilenden^{52, 58}. Diese nehmen Einfluss auf die Leistungsbereitschaft der Schüler⁶⁸.

3.2.9. Gegenmaßnahmen zur Verringerung systematischer Beurteilungsfehler

KLEBER (1992) sieht in der Objektivität der Leistungsbeurteilung ein nie zu erreichendes Ziel⁵². Der zwanghafte Versuch, eine vollkommene Objektivität als Maßstab anzustreben, führt zu überreaktiven Verhaltensweisen und infolgedessen zu kontraproduktiven Verzerrungsprozessen⁵². Wird der Anspruch einer objektiven Beurteilung jedoch als Triebfeder zur Einführung kritischen Hinterfragens und verbesserter Prüfungssituationen verstanden, ist sie ein durchaus sinnvolles Kriterium⁵². Ziel sollte demnach eine Leistungsbeurteilung im Sinne einer geprüften und hinterfragten Subjektivität sein, welche zu revidierbaren Urteilen führt^{52, 53}.

Entscheidende Voraussetzung zur Reduktion systematischer Beurteilungsfehler ist primär das Wissen und die Einsicht solcher Effekte⁵¹. Eine generelle Kooperation mit Kollegen bei unsicherer oder entscheidender Notengabe sollte bei allen Beurteilern vorhanden sein⁵¹. Die Annahme, durch die Übereinstimmung von Noten verschiedener Beurteiler objektivere Beurteilungen zu erhalten, ist als kritisch zu bezeichnen, da die Übereinstimmung auch aufgrund einer implizierten gruppenspezifischen Persönlichkeitstheorie als gemeinsame Fehleinschätzung zu Stande gekommen sein könnte^{52, 70}. Um die Objektivität dennoch durch den Einsatz mehrerer Beurteiler zu erhöhen und kollegiale Übereinstimmungen zu vermindern, ist es sinnvoll, Beurteiler verschiedener Schichten oder Gruppen heranzuziehen⁵². Darüber hinaus sollten alle Beurteiler mittels umfangreicher Schulungen auf Beobachtungs- und Beurteilungsprozesse sowie mögliche Fehlerquellen hingewiesen werden⁷¹. Die Notwendigkeit der Formulierung exakter, gemeinsamer Lernziele und die Möglichkeit, Beurteilungsprozesse praktisch zu erproben und mögliche Fehlerquellen selbst zu erfahren, sind weitere Faktoren zur Vermeidung leistungsverfälschender, subjektiver Einflusstendenzen⁵¹.

Folgende weitere präventive Vermeidungsstrategien von Beurteilungsfehlern lassen sich aufführen:

- Konsequente Trennung von Beobachtungs- und Beurteilungsprozess^{52, 64}
- Primäre Auseinandersetzung von vorher festgelegten Einzelaspekten der Leistung mit anschließender Verdichtung zu einer Gesamtbeurteilung⁶⁴
- Anonymisierte randomisierte Beurteilung⁶⁴
- Austausch mit Kollegen^{52, 64}
- Hinterfragen, ob in der eigenen Beurteilung eine fächer- oder kursspezifische Beurteilungstendenz vorliegt⁵¹.
- Überprüfen, ob beim Betreuen unterschiedlicher Kurse mit unterschiedlichen Anforderungen die jeweils gestellten Anforderungen angemessen sind⁵¹.

- Vermeidung der Vergabe von Extremurteilen durch nochmaliges Beurteilen sehr guter beziehungsweise sehr schlechter Zensuren⁵¹.
- Vermeidung eines falschen Gesamteindrucks einer Leistung durch die separate Beurteilung hinsichtlich eines Teilbereiches bei allen Arbeiten⁵¹.

Die Ausführung konträrer Handlung stellt eine weitere Möglichkeit zur Verringerung von Vorannahmen, implizierten Persönlichkeitstheorien und Halo-Effekten dar⁶⁹. Eine schriftliche Dokumentation des individuellen Beurteilungsverhaltens, welche für die erfolgreiche Umsetzung der genannten Kriterien obligat ist sowie eine ständig präsente schriftliche Erinnerung während des Beurteilungsprozesses, um den Beurteiler immer wieder an Schwierigkeiten und mögliche Vermeidungsstrategien zu erinnern, sind als weitere Optionen zu nennen⁵².

Eine mögliche Lösung der Beurteilungsproblematik durch computerbasierte Auswertungssysteme ist Gegenstand derzeitiger Überlegungen der folgenden Literatur.

Laut MAIER (2015) stellt eine erfahrene Lehrkraft oft die beste Diagnostik bei Leistungseinschätzungen dar, da sich diese durch eine intuitive Komponente, komplexe Situationen schnell zu Gesamteindrücken verdichten zu können, auszeichnet⁴⁶. Allerdings neigen menschliche Beurteiler schnell dazu, sich auf eine Facette der Leistungsvorlage zu fokussieren und den Rest nicht ausreichend zu berücksichtigen⁴⁶. Ferner ist die Verarbeitung von Informationen im Vergleich zu Computern deutlich langsamer⁴⁶. Digitale Beurteilungssysteme können auf größere Mengen von Vergleichsreferenzen zurückgreifen und ermöglichen die systematische Erfassung von Ergebnissen großer Lerngruppen⁴⁶. Allerdings sollte computerbasierten Systemen nicht vorbehaltlos vertraut werden: Als letzte Instanz sollte stets der Mensch die Kontrolle behalten, um ein eigenmächtiges Entscheiden der digitalisierten Beurteilung zu unterbinden⁴⁶.

3.3. Eigene Fragestellung

Vor dem Hintergrund der bereits beschriebenen Situation von Studenten der Zahnheilkunde war es das Ziel dieser Studie, den Faktor der mangelnden Beurteilungsqualität zu untersuchen. Dabei wurde die Reproduzierbarkeit in der Notengebung von fünf zahnärztlichen Beurteilern im Vergleich zu der prepCheck 1.1 Analysesoftware als wesentlicher Bestandteil einer fairen Notenvergabe untersucht. Die Reliabilität der einzelnen Noten wurde bestimmt, aber auch das Verhalten der Beurteilerreliabilität hinsichtlich des Bestehens oder Nichtbestehens einer Beurteilung. Ferner wurde die Reliabilität innerhalb der einzelnen Notenkategorien untersucht. Es wurden sowohl alle Beurteiler einzeln (Intrarater-Reliabilität) als auch in einem gemeinsamen Modell aller zahnärztlichen Beurteiler (Interrater-Reliabilität) betrachtet. Weiterhin war es Gegenstand der Arbeit, mit Hilfe von Fragebögen die Rahmenbedingungen in der vorklinischen Ausbildung an der Zahnklinik der Julius-Maximilians-Universität Würzburg sowohl aus Studenten- als auch aus Betreuersicht darzustellen.

3.3.1. Hypothesen

1. Die Reliabilität der zahnärztlichen Beurteiler (Intrarater-Reliabilität) ist signifikant niedriger als die der prepCheck 1.1 Analysesoftware.
2. Die Reliabilität aller zahnärztlichen Beurteiler (Interrater-Reliabilität) weist einen schwachen Reliabilitätskoeffizienten auf.
3. Die Reliabilität der zahnärztlichen Beurteiler weist ähnliche Werte auf und unterliegt nur geringfügigen individuellen Schwankungen.
4. Der Grad der Berufserfahrung wirkt sich auf die Reliabilität aus.
5. Die unterschiedlichen Beurteilungskategorien weisen ähnliche Reliabilitätskoeffizienten auf und unterliegen nur geringfügigen individuellen Schwankungen.
6. Die Studenten bewerten die gegenwärtige Situation als verbesserungswürdig.
7. Die Zahnärzte bewerten die gegenwärtige Situation als verbesserungswürdig.
8. Die zahnärztlichen Beurteiler weisen ähnliche Beobachtungs- und Beurteilungsfehler wie in der schulischen Leistungsdiagnostik auf.
9. Die computergestützte Beurteilung (CEREC und prepCheck 1.1) ist ein anwenderunabhängiges objektives Beurteilungssystem.

4. METHODIK

Die Versuche fanden zu Semesterbeginn am 25.04.2015 mit 25 Teilnehmern beziehungsweise am 09.05.2015 mit 18 Teilnehmern im vorklinischen Laboratorium im Zentrum für Zahn-, Mund- und Kiefergesundheit der Universitätsklinik Würzburg statt und wiesen eine zeitliche Dauer von acht beziehungsweise neun Stunden auf.

4.1. Versuchsteilnehmer

Die Studenten:

Insgesamt nahmen 43 Studenten der Zahnheilkunde an der Studie als verpflichtendem Bestandteil des dritten Fachsemesters (Durchschnitt: 3,05. Fachsemester) teil. Davon waren 24 weiblich und 19 männlich. Das Durchschnittsalter betrug 22,78 Jahre, wobei der älteste Teilnehmer 41 und der jüngste 19 Jahre alt waren. Die Studenten wurden auf zwei Versuchstage aufgeteilt, um die Abläufe möglichst effizient zu gestalten und lange Wartezeiten zu vermeiden.

Die Beurteiler:

Bei den fünf zahnärztlichen Beurteilern handelte es sich um vier approbierte Zahnärzte, unter denen sich ein Professor (im Folgenden mit „P“ abgekürzt), ein Oberarzt („B“ = Betreuer, fachlicher Leiter der vorklinischen Kurse), zwei als wissenschaftliche Mitarbeiter angestellte Zahnärzte („K“ = Kontrollperson mit sporadischer Tätigkeit im Kurs und „L“ = Langzeitkontrollperson mit mehrjähriger Tätigkeit im Kurs) und ein Zahnmedizinstudent im zehnten Fachsemester („A“ = Anfänger) befand, dessen fachlichen Kenntnisse denen eines Berufsanfängers gleichzusetzen waren. Im Folgenden wird für die oben genannten Personen der Begriff (zahnärztliche) Beurteiler verwendet.

Das Alter der Beurteiler lag zwischen 25 und 53 Jahren und betrug im Durchschnitt 35,4 Jahre.

Die Berufserfahrung und die Zeit der Anstellung der Beurteiler an der Zahnklinik in Würzburg zum Zeitpunkt der Untersuchung ist in Tabelle 1 dargestellt.

Tabelle 1: Berufserfahrung und Angestelltenzeit der Beurteiler

Erfahrung (Jahre)	P	A	L	K	B
Berufserfahrung insgesamt	28	0	4	3	7
Zeit an der Zahnklinik Würzburg	2	0	3	3	7

Die durchschnittliche Berufserfahrung betrug 8,4 Jahre, die durchschnittliche Zeit des Angestelltenverhältnisses an der Zahnklinik in Würzburg im Fachbereich zahnärztliche Prothetik 2,5 Jahre.

Die Analyseeinheit:



Abbildung 1: CEREC AC-Aufnahmeeinheit⁷² Wiedergabeerlaubnis mit freundlicher Genehmigung von Sirona

Verwendet wurde das Modell CEREC AC mit der Omnicam V2.2.5 (Sirona Dental Systems GmbH, Bensheim, Deutschland) als Erfassungs- und Beurteilungseinheit mit einem Intel Core i7-3930K CPU 3.20 GHz Prozessor mit 16 GB Arbeitsspeicher und einem 64 Bit-Betriebssystem (siehe Abbildung 1). Der Gerätetreiber (ACPI x64-based PC) wurde von Microsoft gestellt. Als Grafikkarte lag eine GeForce GTX 760 (NVIDIA) mit einer empfohlenen Bildschirmauflösung von 1280 x 1024 vor.

Bei der Analysesoftware handelte es sich um prepCheck Version 1.1 (Sirona Dental Systems GmbH 2014, Bensheim, Deutschland). Während der beiden Versuchstage standen jeweils zwei CEREC AC-Aufnahmeeinheiten mit der Analysesoftware prepCheck 1.1 zur Verfügung, welche von zwei Studenten des zehnten Fachsemesters der Zahnheilkunde bedient wurden.

Die prepCheck 1.1 Analysesoftware wird im Folgenden mit der Kennung „M“ (Maschine) abgekürzt.

4.2. Material

Zur Verfügung standen jedem Studenten folgende Materialien der Firma KaVo (KaVo Dental GmbH, Biberach, Deutschland):

- Simulationseinheit DSEPlus
- Studiomodell Basic mit über einen Rastermechanismus austauschbaren Modellzähnen (siehe Abbildung 2)
- Expertmatic E25L Rotringwinkelstück mit 1:5 Übersetzung.



Abbildung 2: Studiomodell Basic

Darüber hinaus wurde das zehnteilige Komet Präparationsset I (Satz TD2343) (Gebr. Brasseler GmbH & Co. KG, Lemgo, Deutschland) verwendet (siehe Abbildung 3).

Bei der Vorpräparation kamen diamantierte Schleifinstrumente mit mittlerer Körnung (70 – 80 μm), für die abschließende Finierung Instrumente mit feiner Körnung (30 – 40 μm , gekennzeichnet mit einem roten Ring) zum Einsatz. Zur Verfügung standen ein spitz zulaufender Schleifkörper mittlerer Körnung mit 1,0 mm Durchmesser (siehe Abbildung 3: 859 010), zwei eiförmige Schleifkörper mittlerer und feiner Körnung mit 2,1 mm Durchmesser (siehe Abbildung 3: 379 021, 8379 021), zwei walzenförmige Schleifkörper mittlerer und feiner Körnung mit 1,6 mm Durchmesser (siehe Abbildung 3: 837 KR 016, 8837 KR 016) und vier torpedoförmige Schleifkörper mittlerer und feiner Körnung mit jeweils 1,2 mm beziehungsweise 1,6 mm Durchmesser (siehe Abbildung 3: 878 012, 8878 012, 879 016, 8879 016). Das enthaltene Instrument H1 016 wurde für die Kronenpräparation nicht verwendet.

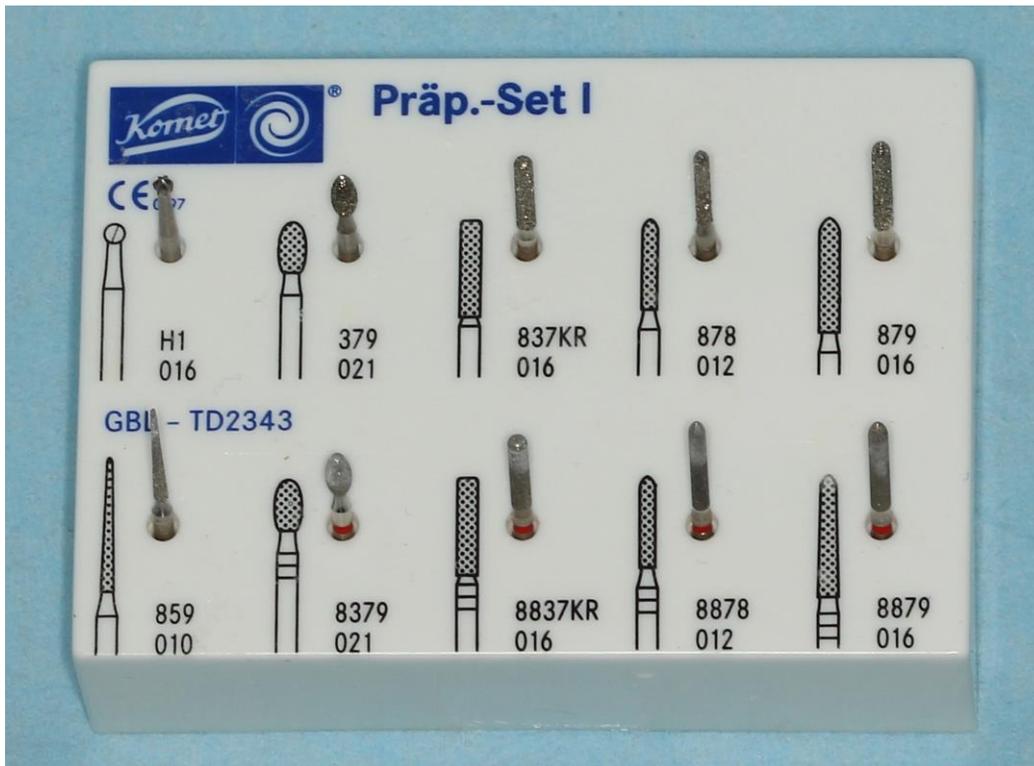


Abbildung 3: Präparationsinstrumente



Abbildung 4: Markierte Zähne

Für die zwei durchgeführten Versuchstestate wurden jedem Studenten zwei markierte herausnehmbare Kunststoffzähne zur Verfügung gestellt. Im Vorfeld wurden in die Wurzeln der Modellzähne Platz- und Versuchstestatnummer eingraviert und anschließend mit verschiedenfarbigen wasserfesten Filzstiften nachgezeichnet. Auch die Studiodelle waren mit Platznummern gekennzeichnet. Auf diese Weise konnte jeder Zahn sowohl dem dazugehörigen Studiomodell als auch dem Versuchstestat (1. Versuchstestat: Rot, 2. Versuchstestat: Schwarz) zugeordnet werden (siehe Abbildung 4). Die Markierung diente zusätzlich der Prophylaxe eines

Täuschungsversuches im Sinne des vorsätzlichen Austausches der Zähne. Nach Abschluss der Versuchsreihe wurden die Zähne einzeln in dazugehörigen Kunststoffhüllen eingesammelt.

4.3. Präparationsdurchführung

Präpariert wurde der Zahn 26 für die prothetische Versorgung einer Vollgusskrone. Dabei sollten folgende Vorgaben eingehalten werden (siehe Tabelle 2):

Tabelle 2: Vorgaben für die Vollgusspräparation

Okklusale Reduktion	Zirkuläre Reduktion	Präparationsgrenze	Lage der Präparationsgrenze
1 mm	0,8 mm	0,5 mm Hohlkehle	1 mm supragingival

Die Durchführung einer Vollgusspräparation war den Teilnehmern bereits bekannt, da im Rahmen des Technisch Propädeutischen Kurses bereits Vollgusspräparationen der Zähne 35 und 36 durchgeführt werden. Die Präparation eines Oberkiefermolars stellte für die Studenten eine Transferaufgabe dar.

Aus der Organisation des Studiums ergab sich zwischen der letzten praktischen Übung und dem Versuchstag für den Großteil der Studenten eine etwa neunmonatige Pause. Dieser Zeitraum gilt für die Studenten in Regelstudienzeit.



Abbildung 5: Fertiggestellte Präparation (Goldstandard)



Abbildung 6: Vervielfältigter Goldstandard

Goldstandard:

Im Vorfeld wurde eine ideale Vollgusspräparation des Zahnes 26 unter den gleichen Bedingungen wie im Versuchstestat hergestellt (siehe Abbildung 5), gescannt und in prepCheck 1.1 als Referenz für die Analyse des Substanzabtrages hinterlegt. Darüber hinaus stand den Studenten als Orientierung während der Präparationen der Goldstandard als Gipsmodell zur Verfügung. Dafür wurde dieser mit Hilfe einer Silikonform dreiundvierzigmal mit Blaugips dupliziert (siehe Abbildung 6).

4.4. Versuchsdurchführung

Die Studenten wurden zu Beginn randomisiert in zwei Gruppen eingeteilt und absolvierten insgesamt zwei Versuchstestate. Bei diesen führten sie jeweils in 60 Minuten die Vollgusspräparation des Zahnes 26 an der Simulationseinheit durch. Das gegebene Zeitfenster von 60 Minuten konnte von den Studenten nicht überschritten werden, da nach Ablauf der Zeit die Stromversorgung für die Simulationseinheiten unterbrochen wurde. Ein vorzeitiges Abschließen der Präparation war den Studenten möglich, allerdings wurde die Beurteilung erst nach vollständigem Ablauf der Zeit vorgenommen. Das erste Versuchstestat fand unmittelbar zu Beginn und ohne jegliche Übung statt, das zweite Testat schloss sich am Ende des Versuchstages an eine dazwischenliegende Übungsphase an. Die Studenten konnten im Sinne einer Randomisierung ihre Arbeitsplätze frei wählen; die Gruppeneinteilung nach Laborplätzen war ihnen zu diesem Zeitpunkt unbekannt. Ein Teil der Studenten befand sich in der Gruppe „prepCheck“ und erhielt die Instruktionen und Beurteilungen ausschließlich durch die prepCheck 1.1 Analysesoftware (Anzahl (n) = 21). Die zweite Gruppe wurde durch einen wissenschaftlichen Mitarbeiter betreut, welcher auch regulär Kursen tätig und dementsprechend mit der Situation vertraut ist (n = 22). Für die vorliegende Arbeit wurden ausschließlich die Beurteilungen der Präparationen aus dem zweiten Versuchstestat herangezogen. Die Beurteilungsergebnisse des ersten Versuchstestates sowie eine Analyse der Übungsphase sind nicht Gegenstand dieser Arbeit. Alle 43 Zahnstümpfe wurden dabei jeweils zweimal direkt nacheinander durch die zahnärztlichen Beurteiler und prepCheck 1.1 zensiert.

4.5. Fragebögen

Im Anschluss an die Versuchstage beantworteten sowohl die Studenten als auch die an der Ausbildung beteiligten approbierten Zahnärzte zum Ankreuzen konzipierte anonyme Fragebögen zum Thema „Qualität der vorklinischen Ausbildung“, welche ohne zeitliche Limitierung und als verpflichtender Bestandteil der Untersuchung zu beantworten waren (siehe Anhang A – C).

Die Fragebögen haben alle Beurteiler – außer „A“ als Student der Zahnheilkunde – beantwortet. An der Beantwortung beteiligte sich darüber hinaus ein weiterer wissenschaftlicher Mitarbeiter der Universität Würzburg.

Die Fragebögen der Studenten unterschieden sich lediglich dahingehend, dass die 21 Teilnehmer, welche mit prepCheck 1.1 gearbeitet hatten, zusätzliche Fragen zu diesem Thema erhielten. Es soll darauf hingewiesen werden, dass die Begriffe „Assistent“ und „Studierender“ in den Fragebögen verwendet wurden. Diese sind in dieser Arbeit mit den Begriffen „wissenschaftlicher Mitarbeiter“ und „Student“ gleichzusetzen.

Es wurden überwiegend jeweils vierstufige Ratingskalen mit verbaler Skalenbezeichnung („stimme voll zu“, „stimme eher zu“, „stimme eher nicht zu“, „stimme gar nicht zu“ beziehungsweise „ja“, „eher ja“, „eher nein“ und „nein“) verwendet. Zusätzliche kamen numerische Ratingskalen und ankreuzbare gebundene Antwortformate zum Einsatz; Mehrfachnennungen waren möglich. Bei den Fragebögen der Zahnärzte kamen zudem auch offene Antwortformate zum Einsatz. An dieser Stelle sei ebenfalls darauf hingewiesen, dass ein Teil der Fragen in den Originalfragebögen nicht verwendet wurde, da sich diese auf andere Themengebiete beziehen, welche nicht Gegenstand der vorliegenden Arbeit sind. Die nachfolgenden Aufstellungen zeigen alle in dieser Arbeit verwendeten Fragen.

4.5.1. Fragebögen der Studenten

Die Ergebnisse der Fragebogenauswertungen (25 Fragen) unter den Studenten wurden zu Gunsten einer besseren Übersicht in fünf Themengebiete unterteilt.

- *Allgemeiner Zustand des Betreuungsverhältnisses:*
 - Das Betreuungsverhältnis Assistenten zu Studierenden ist gut.
 - Die Assistenten wirken manchmal überfordert.
 - Die Wartezeit beim Vorzeigen einer Arbeit ist in der Regel zu lang.
 - Ein Assistent ist während der Kurszeit immer anwesend.
 - Das Bewertungssystem (Note 1 – 6) ist gut.
- *Einstellungen, Erwartungen und Wünsche der Studierenden:*
 - Objektivität ist mir bei Benotungen sehr wichtig.
 - Die Noten im Kurs haben entscheidenden Einfluss auf meine Motivation.
 - Die Notenvergabe empfinde ich als gerecht und nachvollziehbar.
 - Ich vertraue dem Urteil der Assistenten.
 - Bei Nichtbestehen eines praktischen Kurses /Exmatrikulation könnte ich mir vorstellen, rechtlich dagegen vorzugehen.
 - Objektivere Verhältnisse würden das Arbeitsklima verbessern.
- *Unterschiede in der Bewertungsqualität:*
 - Ein Unterschied in der Kritik/Benotung zwischen den einzelnen Assistenten ist vorhanden.
 - Die Kritik von unerfahrenen Assistenten hat weniger Qualität als die von Erfahrenen.
 - Die Kritik der Assistenten ist tagesformabhängig (z.B. morgens strenger als nachmittags).

- Dieselbe Arbeit wird von verschiedenen Assistenten unterschiedlich bewertet.
- Dieselbe Arbeit wird vom gleichen Assistenten unterschiedlich bewertet.
- *Persönliche Faktoren:*
 - Ich fühle mich manchmal unwohl beim Vorzeigen einer Arbeit.
 - Ich habe das Gefühl, dass einzelne Studierende in der Kritik/Bewertung bevorzugt werden.
 - Die Kritik der Assistenten ist manchmal herablassend.
 - Ich habe den Eindruck, dass der Assistent sich ein Bild von mir gemacht hat.
 - Ich habe das Gefühl, dass einzelne Studierende in der Kritik/Bewertung benachteiligt werden.
- *Umgang mit prepCheck 1.1 (n = 21):*
 - Die Bewertung durch einen Computer empfinde ich als gerecht.
 - Computerurteile sind generell vertrauenswürdiger als menschliche Bewertungen.
 - Welches System würden Sie im Laboralltag bevorzugen?
 - Welches System würden Sie bei Bewertungen von Präparationstestaten bevorzugen?

4.5.2. Fragebögen der Zahnärzte

Die Ergebnisse der Fragebogenauswertungen (25 Fragen) unter den zahnärztlichen Beurteilern wurden zu Gunsten einer besseren Übersicht in die folgenden sechs Themengebiete unterteilt:

- *Allgemeiner Zustand des Betreuungsverhältnisses:*
 - Das Betreuungsverhältnis Assistent-Studierende ist gut.
 - Die Betreuung der vorklinischen Kurse ist mit Stress verbunden.
 - Der Anspruch einer objektiven studentischen Betreuung überfordert Sie manchmal.
 - Meine Kritik während der Kurszeit beinhaltet Hinweise zur Verbesserung der Präparation (z.B. sich auf der Zahnreihe abstützen usw.).
 - Intensive Betreuung eines einzelnen Studierenden (z.B. in Form von Demonstrationen) ist zeitlich möglich.
 - Das Bewertungssystem (Note 1 – 6) ist sinnvoll.
- *Objektivität:*
 - Wenn ein präparierter Zahn von mehreren Assistenten mit der gleichen Note beurteilt wurde, ist er objektiv bewertet.
 - Lassen Sie sich bei Bewertungen von Kollegenmeinungen beeinflussen?
 - Die Qualität der objektiven Bewertung ist bei allen Assistenten gleich.

- *Anonymität:*
 - Eine Anonymisierung steigert die Objektivität.
 - Besonders auffällige Studierende prägen sich mit ihrer Platznummer ein.
 - Liegt bei manchen Platznummern während des Benotungsprozesses eine gewisse Erwartungshaltung vor?
- *Zu Beeinflussende Faktoren I:*
 - Würden Sie sagen, dass es „typische Anfängerfehler“ gibt, die nahezu immer vorkommen?
 - Würden Sie sagen, dass Sie die Studierenden eines Kurses nach einer gewissen Zeit einschätzen können (strebsam, faul, geschickt, ungeschickt, usw.)?
 - Glauben Sie, dass es mit Hilfe des „ersten Eindrucks“ möglich ist, einzuschätzen, wie gut ein Studierender im Kurs zurechtkommt?
 - Ein unsauberer Arbeitsplatz geht auch mit einer ungenauen Arbeitsweise einher.
 - Beeinflusst Sympathie ihre Notengebung positiv?
 - Beeinflusst Antipathie ihre Notengebung negativ?
 - Das plötzlich sehr gute Abschneiden eines schwachen Studierenden im Testat macht Sie misstrauisch.
 - Eine z.B. unzureichende Oberflächenglättung hat einen negativen Einfluss auf die Beurteilung der Präparationsgrenze.
 - Lassen Sie sich von zuvor erbrachten Leistungen (z.B. Aufwuchsübungen) eines Studierenden bei der Bewertung beeinflussen?
- *Beeinflussende Faktoren II:*

Faktoren, welche die objektive Beurteilung beeinflussen (Anzahl der Nennungen). Zu Gunsten einer besseren Übersicht wurden jeweils die Begriffe Stress und Frust sowie Ablenkung, Lärm und Lichtverhältnisse zusammengefasst:

- Ermüdung
- Stress, Frust
- „Betriebsblindheit“ (eine Beurteilungssituation, in der viele Beurteilungen in kurzer Zeit getroffen werden müssen)
- Ablenkung, Lärm, Lichtverhältnisse
- eigene Erfahrungswerte
- Sympathie/Antipathie

- *Eigene Meinung:*
 - Finden Sie, dass die jetzige vorklinische Ausbildung einen wichtigen, unverzichtbaren Bestandteil der zahnmedizinischen Ausbildung darstellt?
 - Denken Sie, dass die jetzige vorklinische Ausbildung der Abhärtung dienen sollte?
 - Ihre persönliche Meinung: Besteht Verbesserungsbedarf bei der vorklinischen Bewertung?

4.6. Beurteilungsbögen

Um die Präparationsleistung zu beurteilen und Vergleiche zwischen den zahnärztlichen Beurteilern und prepCheck 1.1 zu ermöglichen, wurden zwei analoge Benotungssysteme mit gleichen Beurteilungskategorien und Notenskalen entwickelt.

Beurteilt wurden die präparierten Zähne anhand von vier Kategorien:

- *Präparationsgrenze* (Verlauf und Beschaffenheit)
- *Oberfläche* (scharfe Kanten und Hinterschnitt)
- *Substanzabtrag* (okklusal und zirkulär)
- *Präparationswinkel* (steil oder konisch)

Dabei konnte in jeder Kategorie eine der vier möglichen Noten *sehr gut*, *akzeptabel*, *verbesserungswürdig* und *zerstört* erteilt werden.

- *Sehr gut* steht dabei für eine Leistung, welche dem Goldstandard sehr ähnlich ist oder nur geringfügige Abweichungen aufweist.
- *Akzeptabel* ist eine Leistung, wenn sie dem Goldstandard nicht entspricht, aber ohne weitere Korrekturmaßnahmen an einem potentiellen Patienten belassen werden kann.
- *Verbesserungswürdig* wird vergeben, wenn die Präparation in dieser Form nicht für die Aufnahme einer Kronenversorgung geeignet ist, sich aber durch Optimierungsmaßnahmen in den Zustand *akzeptabel* oder *sehr gut* überführen lässt (zu wenig *Substanzabtrag*, eine zu steile Präparation, eine ungenaue *Präparationsgrenze* oder eine unzureichende Glättung der *Oberfläche*).
- Als *zerstört* wird ein Zahn bezeichnet, wenn irreversible Präparationsfehler begangen wurden (zu hoher *Substanzabtrag*, zu konische Winkel, subgingivale *Präparationsgrenze*, ausgeprägte *Hinterschnitte*).

Beurteilungssystem für die zahnärztlichen Beurteiler:

Zur genaueren Definition der Beurteilungskategorien und der Noten enthalten die Beurteilungsbögen die unten dargestellten stichpunktartigen Kriterien sowie eine schriftlich vorliegende Ausfüllhilfe (siehe Abbildung 7). Ergänzend dazu erhielten alle Beurteiler eine mündliche Einweisung und ein Duplikat des Goldstandards.

Bewertungsbogen für Assistenten				
Platznummer: Versuch: <input type="radio"/> 1 <input type="radio"/> 2				
	sehr gut ≈ Goldstandard, minimale Abweichung	akzeptabel ≈ mittelmäßige Leistung, noch ok	verbesserungswürdig ≈ zu hoch, zu steil, scharfe Kanten	zerstört ≈ zu klein, zu konisch, Präpfgrenze falsch
Präparationsgrenze Verlauf u. Beschaffenheit				
Oberfläche scharfe Kanten, Hinterschmitte				
Substanzabtrag okklusal und zirkulär				
Präparationswinkel steil/konisch				

Hinweis/Ausfüllhilfe

sehr gut: Dem Goldstandard sehr ähnlich / entsprechend

akzeptabel: Nicht perfekt, aber am Patienten verwendbar

verbesserungswürdig: Alles was noch zu retten ist: zu wenig Substanzabtrag, unebene Oberfläche, Prägfgrenze >3mm über Gingiva

Zerstört: Prägfgrenze subgingival oder isogingival, zu viel Substanzabtrag, zu konisch, Prägfgrenze sehr uneben, große Rillen, ausgeprägte Hinterschmitte

Abbildung 7: Beurteilungsbogen für die zahnärztlichen Beurteiler

Beurteilungssystem prepCheck 1.1:

Das zweite System wurde für die Beurteilung der präparierten Zähne durch prepCheck 1.1 entwickelt (siehe Tabelle 3). Durch vorab durchgeführte Tests mit prepCheck 1.1 wurden für jede Kategorie Notenabstufungen definiert, um eine Beurteilung der Leistungen zu ermöglichen (siehe Anhang D). Die Notengrenzen im Beurteilungsbogen basieren dabei auf den von Prof. Dr. Pospiech verifizierten Benotungen von Beispielpräparationen. Die genaue Festlegung der Schwellenwerte erfolgte ebenfalls unter seiner Abstimmung. Aufgrund der Analysemethodik von prepCheck 1.1 ergeben sich sechs Beurteilungskategorien, die in folgende Kategorien gebündelt wurden, um eine Vergleichbarkeit zu ermöglichen:

Die Notenkategorie *Präparationsgrenze* setzt sich aus *Präparationsrand* und *Randabschlussstyp* zusammen. In der Notenkategorie *Oberfläche* werden die Ergebnisse *Oberflächenbeschaffenheit* und *Hinterschnitt* zusammengefasst. In der Subkategorie *Oberflächenbeschaffenheit* liefert prepCheck 1.1 keine prozentuale Angabe der Störstellen, sondern lediglich eine farbliche Darstellung auf dem präparierten Zahn. Somit unterliegt die *Oberflächenbeschaffenheit* der subjektiven Einschätzung des jeweiligen Programmanwenders. Um den subjektiven Einfluss der Anwender zu minimieren, wurde im Vorfeld eine Kalibrierungshilfe erstellt und konnte während des Beurteilungsprozesses genutzt werden (siehe Anhang E). Ähnlich stellt es sich mit der Subkategorie *Präparationsrand* dar. prepCheck 1.1 bietet zur Analyse des *Präparationsrandes* ebenfalls nur eine farbliche Markierung der Störstellen an. Die Längen der detektierten Fehlerstellen können im Gegensatz zur Subkategorie *Oberflächenbeschaffenheit* mit einem digitalen Lineal vermessen werden. Die Subkategorien *Randabschlussstyp* und *Hinterschnitt* weisen aufgrund ihrer besseren Messbarkeit (im Sinne prozentualer Angaben) innerhalb einer Notenkategorie größere Gewichtung auf. Dies bedeutet im Einzelnen: differieren die Noten um eine Stufe, wird die Note vergeben, welche die Subkategorie mit größerer Gewichtung aufweist. Liegt mehr als eine Notenstufe zwischen den Ergebnissen, wird die endgültige Note der Kategorie gemittelt – mit der Tendenz der stärker gewichteten Subkategorie. Von der Erhebung des Median wurde abgesehen, um der sicher zu beurteilenden Subkategorie mehr Gewichtung zu verleihen.

Tabelle 3: Beurteilungssystem prepCheck 1.1 Subkategorien Randabschlusstyp und Hinterschnitt mit stärkerer Gewichtung;
 Klasse I: dunkelblau, Klasse II: mittelblau, Klasse III: hellblau

Notenkategorie	Subkategorie	sehr gut	akzeptabel	verbesserungswürdig	zerstört
Präparationsgrenze	Präparationsrand	keine gelben Stellen (oder < 1 mm in der Summe, supra-gingival)	gelbe Stellen < 2 mm Breite (oder < 4 mm in der Summe, supra-gingival)	gelbe Stellen < 3 mm Breite (oder < 6 mm in der Summe, supra-gingival)	gelbe Stellen \geq 3 mm Breite (oder \geq 6 mm in der Summe oder > 1mm iso-/subgingival)
	Randabschlusstyp	\geq 50% Klasse I, \leq 2% rot, \leq 5% grün	< 50% Klasse I, \leq 5% rot, \leq 10% grün	\leq 10% rot, > 10% grün	> 10% rot
Oberfläche	Oberflächenbeschaffenheit	sehr wenige scharfe Kanten	wenige scharfe Kanten	mehrere scharfe Kanten	viele scharfe Kanten
	Hinterschnitt	\leq 2% Hinterschnitt	\leq 5% Hinterschnitt	\leq 10% Hinterschnitt	> 10% Hinterschnitt
Substanzabtrag	-	Klasse I \leq 50%, 0% rot, 0% grün	Klasse I+II+III, 0% rot, \leq 15% grün	\leq 10% rot, > 15% grün	> 10% rot
Präparationswinkel	-	Klasse I – III, \leq 5% rot, \leq 10% grün	Klasse I – III, \leq 10% rot, \leq 20% grün	\leq 15% rot, > 20% grün	> 15% rot

4.7. Beurteilungsdurchführung

Zahnärztliche Beurteiler:

Die Beurteilungen fanden schriftlich direkt im Anschluss an die Versuchstage unter Aufsicht statt. Zwischen der Beurteilung des ersten und des zweiten Durchganges lag eine maximale Zeitspanne von einer Stunde. Mit Hilfe der Beurteilungsbögen fand – in zufälliger Reihenfolge – die Beurteilung aller vorher eingesammelten Modelle an einem Beurteilungstisch statt. Die Kommunikation der Beurteiler untereinander war während der Notenvergabe nicht gestattet. Ausgefüllte Beurteilungszettel wurden umgehend eingesammelt.

prepCheck 1.1:

Die Beurteilung der Analysesoftware fand im Anschluss an die Beurteilung der zahnärztlichen Beurteiler statt. Die Präparationen wurden dabei randomisiert und gleichmäßig auf zwei unterschiedliche unabhängige Anwender (Studenten des zehnten Semesters der Zahnheilkunde) aufgeteilt. Nach erfolgter erster Beurteilung kam es zum Austausch der jeweils zugeteilten Präparationen und ein zweiter Beurteilungsdurchgang schloss sich an. So konnte eine beurteilerunabhängige Analyse gewährleistet werden. Die Beurteilung der Subkategorie *Oberflächenbeschaffenheit* erfolgte mittels Kalibrierungshilfe (siehe Anhang E). Eine nachträgliche Änderung der Noten war nicht gestattet.

4.8. Vorgehensweise/Bedienung der Analysesoftware

Die Vorgehensweise der dreidimensionalen Analyse von Präparationen gliedert sich in den Abschnitt der Datenerfassung mittels CEREC-Programm und die sich anschließende Datenanalyse mit prepCheck 1.1.

4.8.1. CEREC

Nach Auswahl der Platznummer, unter welcher der eingescannte Zahn eingespeichert werden soll, erfolgt die visuelle Erfassung des – nach Empfehlung der Gebrauchsanweisung – gesäuberten Pfeilerzahnes in einem 45 – 90° Winkel mit der Kamera Omnicam⁷². Sobald sich der präparierte Stumpf sowie die jeweils angrenzenden Nachbarzähne in ausreichender Qualität auf dem Kontrollbildschirm darstellen, wird der Scanvorgang beendet. Daraufhin berechnet das CEREC-Programm ein dreidimensionales Modell des betreffenden Oberkieferabschnittes. Im Anschluss daran erfolgt die manuelle Einstellung der Einschubachse, bei der die Kronenachse möglichst wenige untersichgehende Stellen aufweist. Die Übertragung der Daten in die Analysesoftware erfolgt anschließend über den Menüpunkt „Applikation ausführen“⁷³.

4.8.2. prepCheck 1.1

Zunächst ist eine Festlegung der nachfolgenden drei Parameter durch den Anwender erforderlich⁷³.

Kronenachse:

Primär erfolgt eine nochmalige Kontrolle der Einschubrichtung, um möglichst wenige untersichgehende Stellen einzustellen⁷³.

Okklusallinie:

Bei der Festlegung der Okklusallinie ist zu berücksichtigen, dass die okklusale Fläche eingeschlossen ist, da diese prepCheck 1.1 als festgelegte Abgrenzung zwischen okklusaler und zirkulärer Fläche dient. Die vestibuläre Ansträgung liegt außerhalb der Okklusallinie (siehe Abbildung 8).

Präparationsrand:

Sollte nach vorherigem Trimmen exakt angezeichnet werden⁷³ (siehe Abbildung 9).

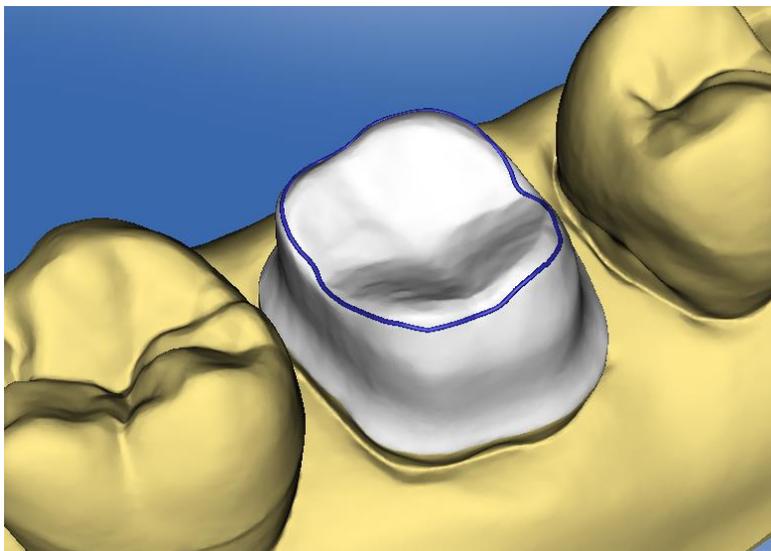


Abbildung 8: Okklusallinie

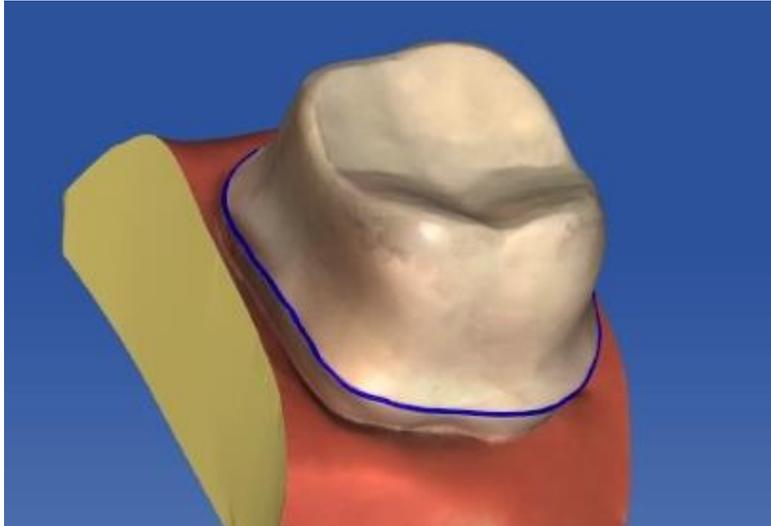


Abbildung 9: Präparationsrand

prepCheck 1.1 analysiert sechs verschiedene Parameter einer Präparation. Die Konfiguration dieser Parameter kann durch den Anwender definiert werden⁷³ (siehe Anhang D).

Die nachfolgenden Abbildungen zeigen den Goldstandard und entsprechend die positiven Analyseergebnisse der einzelnen Parameter.

Oberflächenbeschaffenheit:

Die Oberfläche wird auf den Glättungsgrad hin untersucht; dabei stellt das Programm scharfe Kanten als farblichen Kontrast in Orange dar, flache Bereiche in der Farbe Grau⁷³. Die Stufe des Grautons gibt Auskunft über den Ausprägungsgrad der Krümmung, allerdings nicht über deren Ausrichtung⁷³ (siehe Abbildung 10).

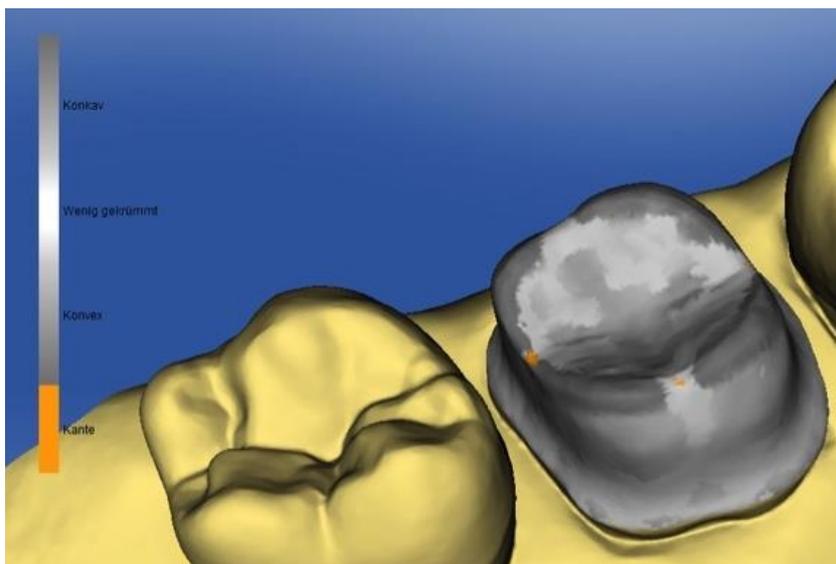


Abbildung 10: Oberflächenbeschaffenheit

Hinterschnitt:

Dieser Parameter visualisiert alle untersichgehenden Bereiche orange in Bezug auf die zuvor definierte Einschubachse, welche mit einem grünen Pfeil angezeigt wird⁷³. Zusätzlich dazu erfolgt eine prozentuale Darstellung der Flächenanteile in einem Kreisdiagramm (siehe Abbildung 11). Die Tiefe des ermittelten *Hinterschnittes* in Millimetern kann darüber hinaus mit dem Cursor abgefragt werden⁷³.

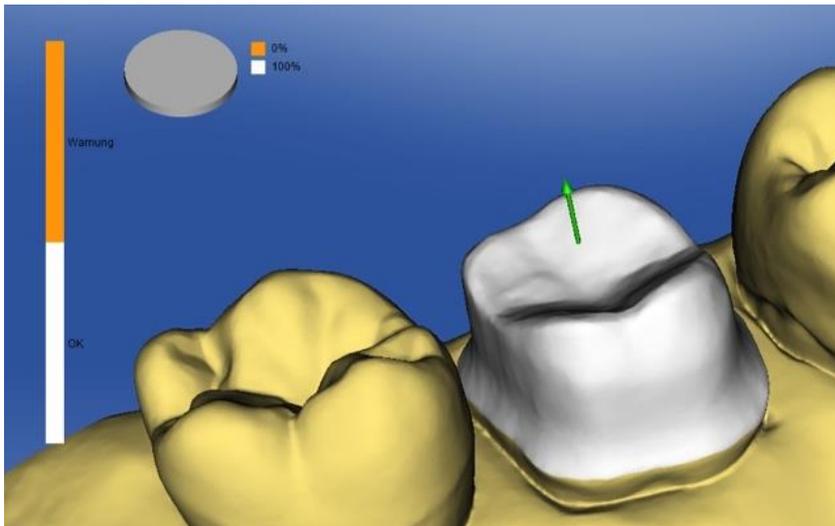


Abbildung 11: Hinterschnitt

Präparationsrand:

Lokalisiert prepCheck 1.1 stark konkave Flächen oder scharfe Kanten im Bereich der Präparationsgrenze, werden diese ebenfalls in Orange dargestellt⁷³. Bereiche, die sich innerhalb der vorher festgelegten Toleranz befinden, erscheinen als weiße Flächen⁷³ (siehe Abbildung 12).

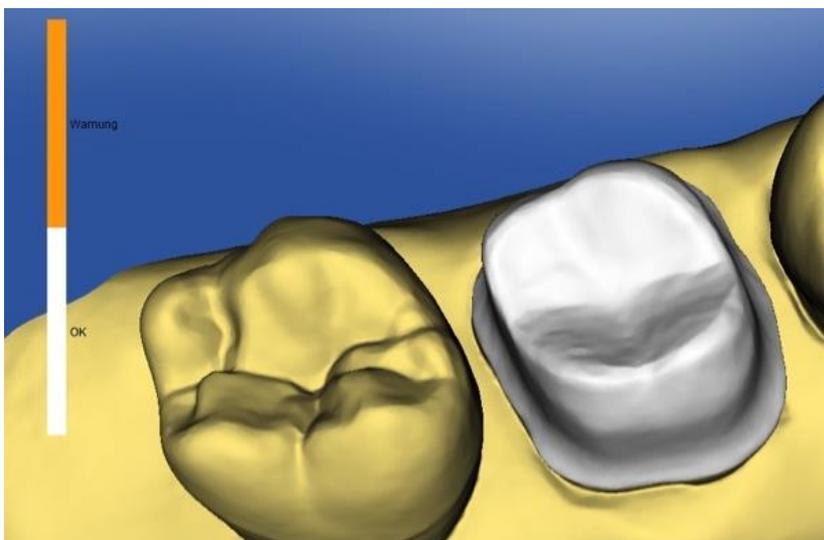


Abbildung 12: Präparationsrand

Randabschlusstyp:

Die Funktion *Randabschlusstyp* analysiert den Verlauf der Präparationsgrenze mit Hilfe vorab definierter Toleranzbereiche⁷³. Zur Darstellung der Ergebnisse verwendet prepCheck 1.1 ein dreistufiges Farbsystem, welches auch bei der Analyse des *Präparationswinkels* und des *Substanzabtrages* verwendet wird. Grün signalisiert dem Anwender, dass ein zu geringer Abtrag erfolgte, welcher aber eine nachfolgende Korrektur noch ermöglicht⁷³. Die Farbe Rot vermittelt einen zu hohen Abtrag, welcher nicht revidierbar ist⁷³. Blau zeigt, dass die Präparation sich im Toleranzbereich befindet⁷³. Zusätzlich wird zwischen Hellblau, Mittelblau und Dunkelblau unterschieden. Hellblau visualisiert, dass sich die Präparation am äußersten Rand des Toleranzbereiches befindet. Dunkelblau stellt den Bereich mit der geringsten Abweichung dar. Mittelblau fungiert als Übergangsfarbe zwischen den beiden anderen Bereichen. Die Darstellung der prozentualen Anteile der Farben werden durch prepCheck 1.1 in einem Kreisdiagramm zusammengefasst (siehe Abbildung 13).

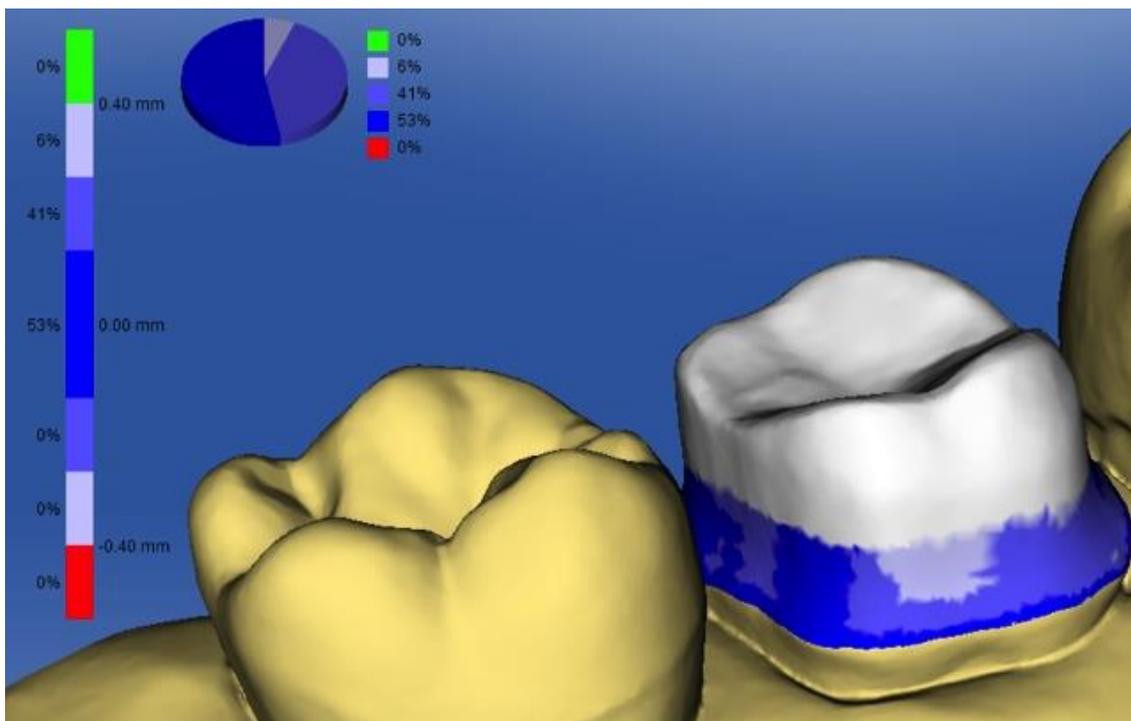


Abbildung 13: Randabschlusstyp

Substanzabtrag:

Den Substanzabtrag der vorliegenden Präparation errechnet prepCheck 1.1 anhand des Vergleichs zum Goldstandard. Unter Berücksichtigung des Versuchsablaufes und der möglichen Nachteile wurde auf die weitere Option des Vergleichs zum Gegenkiefer verzichtet.

Vergleich zum Goldstandard:

Hierbei werden die erfasste Präparation und der Goldstandard anhand identischer Referenzpunkte übereinandergelegt und verglichen. Dabei kann der Anteil des eingeblendeten Goldstandards manuell eingestellt werden⁷³. Zur Beurteilung kommen dieselben Farben wie beim *Randabschlussstyp* zum Einsatz (siehe Abbildung 14).

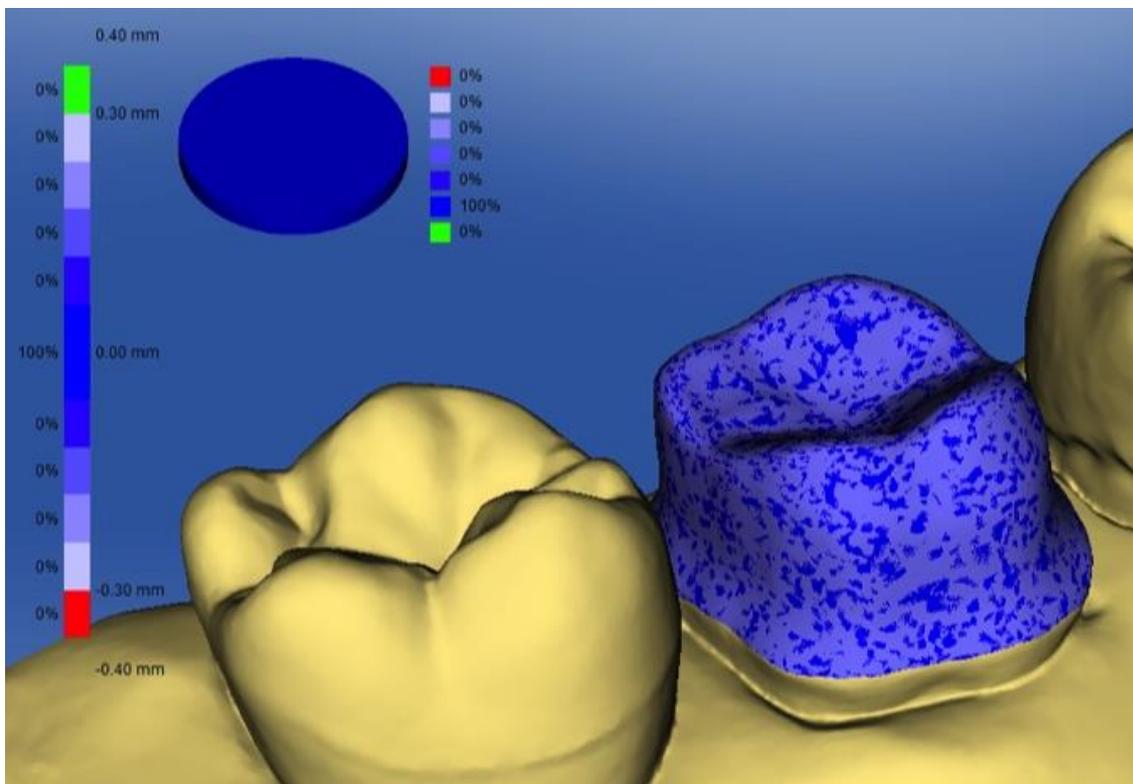


Abbildung 14: Substanzabtrag

Präparationswinkel:

Bei der Ermittlung des *Präparationswinkels* erfolgt die Analyse aller Außenflächen und Höckerabschrägungen; hierbei sind sowohl die vorab festgelegte Definition der Einschubrichtung als auch die Okklusallinie entscheidend⁷³ (siehe Abbildung 15).

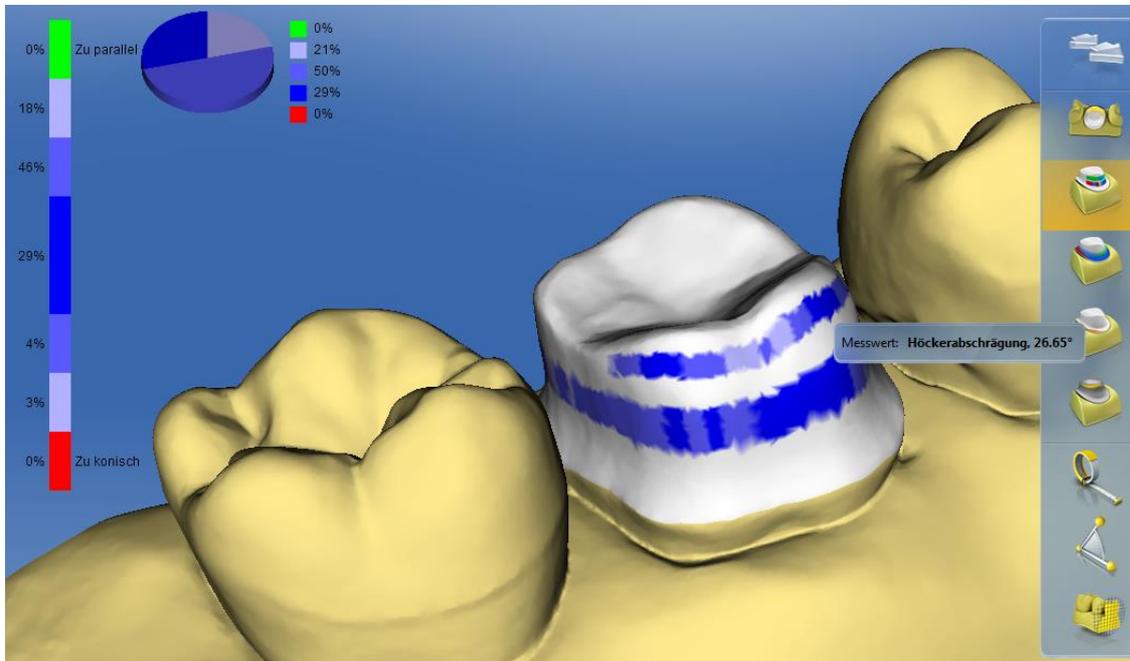


Abbildung 15: Präparationswinkel

4.9. Statistische Auswertung

Die statistischen Auswertungen wurden mit dem Statistikprogramm SPSS 23 für MacOS X durchgeführt. Weitergehende Analysen, die zur Berechnung der Interrater-Reliabilität (AC1-Koeffizienten) nötig waren, wurden mit Hilfe des Statistikprogramms SAS 9.3 für Windows durch eine wissenschaftliche Mitarbeiterin am Institut für Klinische Epidemiologie und Biometrie der Universität Würzburg erstellt.

Zur Bestimmung der Reliabilität der Beurteiler wurden Cohens Kappa-Koeffizienten (κ), (Cohens) Weighted Kappa-Koeffizienten (κ_w) und (Gwets) AC1-Koeffizienten (AC1) sowie die dazugehörigen 95%-Konfidenzintervalle berechnet.

- *Cohen's Kappa* ist ein statistisches Maß für die Übereinstimmung zwischen zwei Beurteilern (Interrater-Reliabilität) oder für die Übereinstimmung desselben Beurteilers zu zwei verschiedenen Zeitpunkten (Intrarater-Reliabilität)⁷⁴. Im Vergleich zur prozentualen Übereinstimmung bietet Cohens Kappa den Vorteil der zufallskorrigierten Übereinstimmungsmessung⁷⁴.

- (Cohen's) *Weighted Kappa* ist ebenfalls ein Maß für die Übereinstimmung zwischen zwei Beurteilern beziehungsweise zwei Urteilen desselben Beurteilers, berücksichtigt darüber hinaus aber auch den Ausprägungsgrad der Abweichungen mit differentiellen Gewichtungen⁷⁴.
- (Gwets) *AC1-Koeffizient* ist ein Maß für die Übereinstimmung zwischen mehreren Beurteilern und gilt als ein stabilerer Schätzer der Interrater-Reliabilität als Kappa, welcher weniger stark von den Randverteilungen der dazugehörigen Kontingenztafeln abhängt⁷⁵.

Zur Überprüfung der Nullhypothese ($H_0: \text{Kappa}/\text{AC1} = 0$) wurde ein Test zum Signifikanzniveau von 5% durchgeführt. Die Interpretation aller Reliabilitätskoeffizienten erfolgte in Anlehnung an die englischsprachige Einteilung nach LANDIS und KOCH (1977)^{76,77} (siehe Tabelle 4):

Tabelle 4: Interpretationseinteilung der Reliabilitäts-Koeffizienten

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

Der Aspekt der Beurteilungsergebnisse wurde deskriptiv statistisch mittels Tabellen und Säulendiagrammen ausgewertet. Die Darstellung der Ergebnisse der Fragebögen wurde ebenfalls mittels deskriptiver Statistik im Sinne von Absolutwerten (Anzahl) und relativer Verteilung (Prozent) der Antworten durchgeführt.

5. ERGEBNISSE

5.1. Reliabilität der Beurteiler in der Notengebung/Intrarater-Reliabilität

Beurteiler „P“:

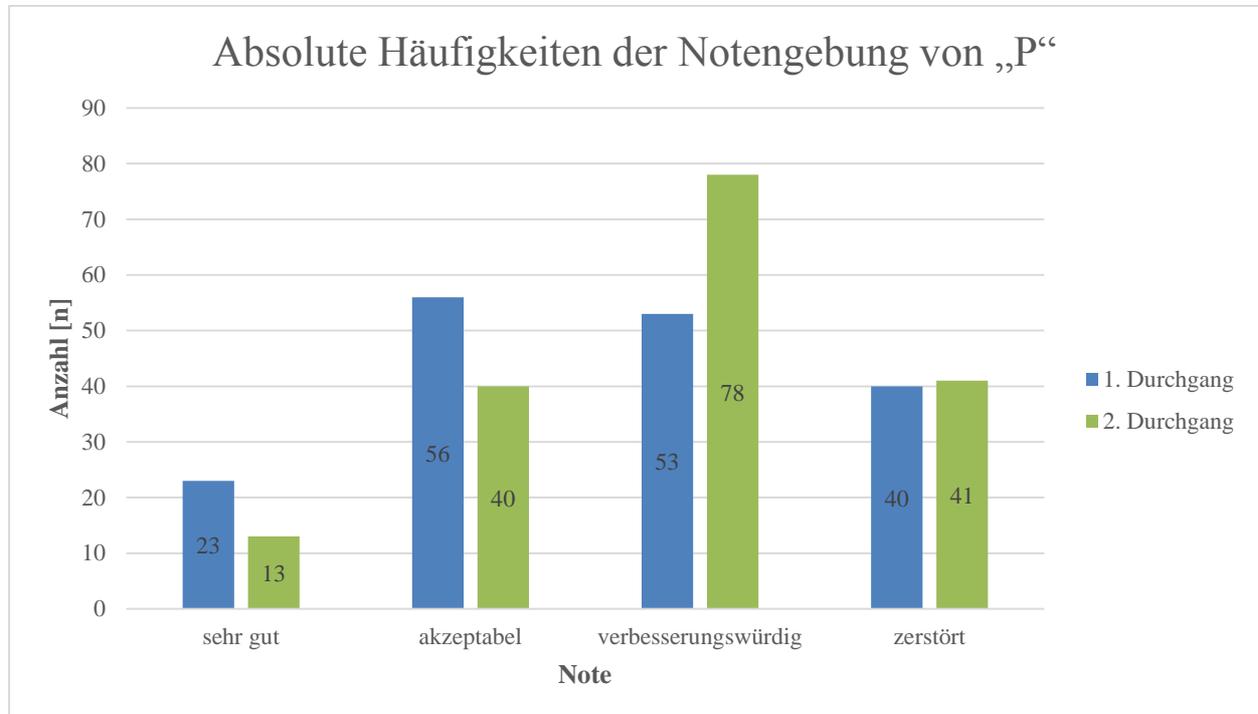


Abbildung 16: Absolute Häufigkeiten der Notengebung von „P“

Die Auswertung der Ergebnisse von „P“ ergibt, dass bei der Anzahl der vergebenen Noten im ersten Durchgang (n_1) das Maximum auf der Note *akzeptabel* ($n_1 = 56$) liegt. Im zweiten Durchgang (n_2) liegt das Maximum auf der Note *verbesserungswürdig* ($n_2 = 78$). Die Beurteilung *sehr gut* macht im ersten Durchgang knapp 14% ($n_1 = 23$), im zweiten Durchgang fast 8% ($n_2 = 13$) aller vergebenen Noten aus. Hinsichtlich der Verteilung der schlechtesten Beurteilung *zerstört* zeigt sich, dass sie mit $n_1 = 40$ beziehungsweise $n_2 = 41$ Urteilen annähernd stabil ist (siehe Abbildung 16).

Der Weighted Kappa-Koeffizient liegt bei $\kappa_w = 0,589$ mit 95%-Konfidenzintervall von $0,49 - 0,68$. Dies ist nach LANDIS und KOCH (1977) als mittlere Übereinstimmung zu beurteilen⁷⁶. Der Test of $H_0: \text{Kappa} = 0$ ergibt einen p-Wert von $< 0,001$, κ_w weicht signifikant von 0 ab.

Beurteiler „A“:

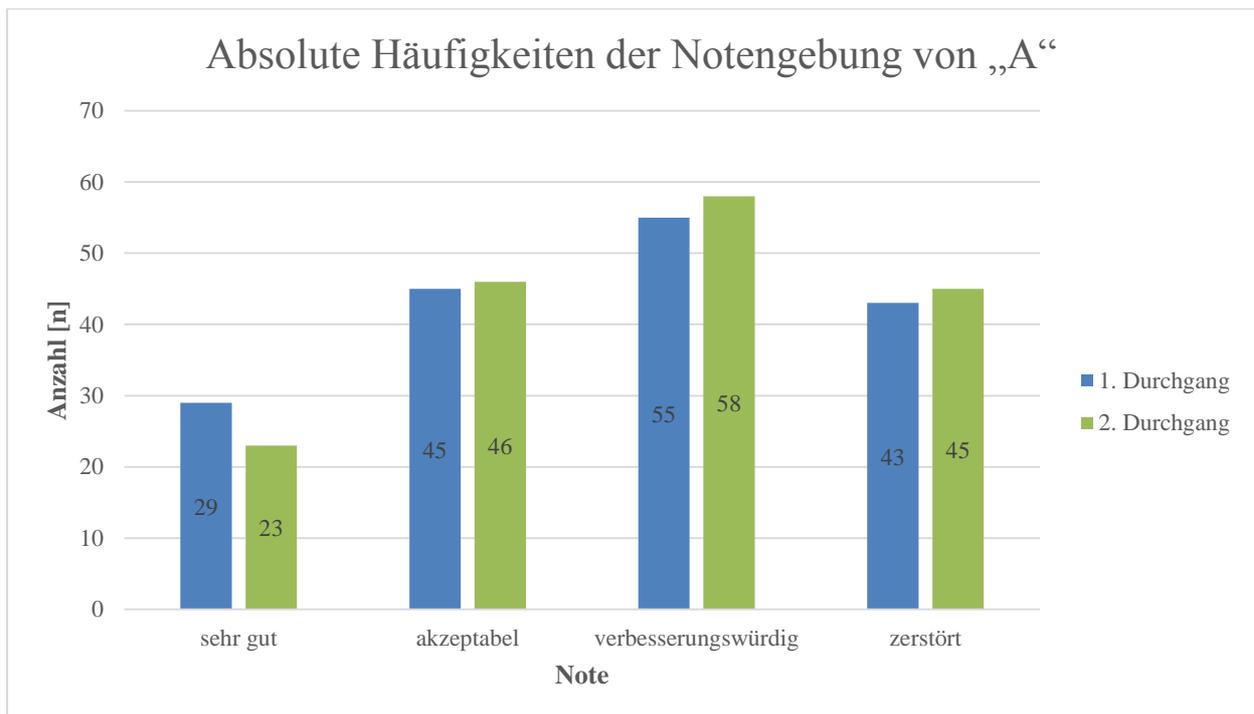


Abbildung 17: Absolute Häufigkeiten der Notengebung von „A“

Die Auswertungen der Notenvergabe von „A“ zeigen nur geringfügig abweichende absolute Häufigkeiten der einzelnen Noten (Note *sehr gut* von $n_1 = 29$ auf $n_2 = 23$ gesunken, Anzahl der Noten *akzeptabel*, *verbesserungswürdig* und *zerstört* gestiegen) (siehe Abbildung 17).

Der Weighted Kappa-Koeffizient liegt bei $\kappa_w = 0,571$ mit einem 95%-Konfidenzintervall von $0,49 - 0,66$. Dies ist nach LANDIS und KOCH (1977) eine mittlere Übereinstimmung⁷⁶. Der Test of $H_0: \text{Kappa} = 0$ ergibt einen p-Wert von $< 0,001$, κ_w weicht signifikant von 0 ab.

Beurteiler „L“:

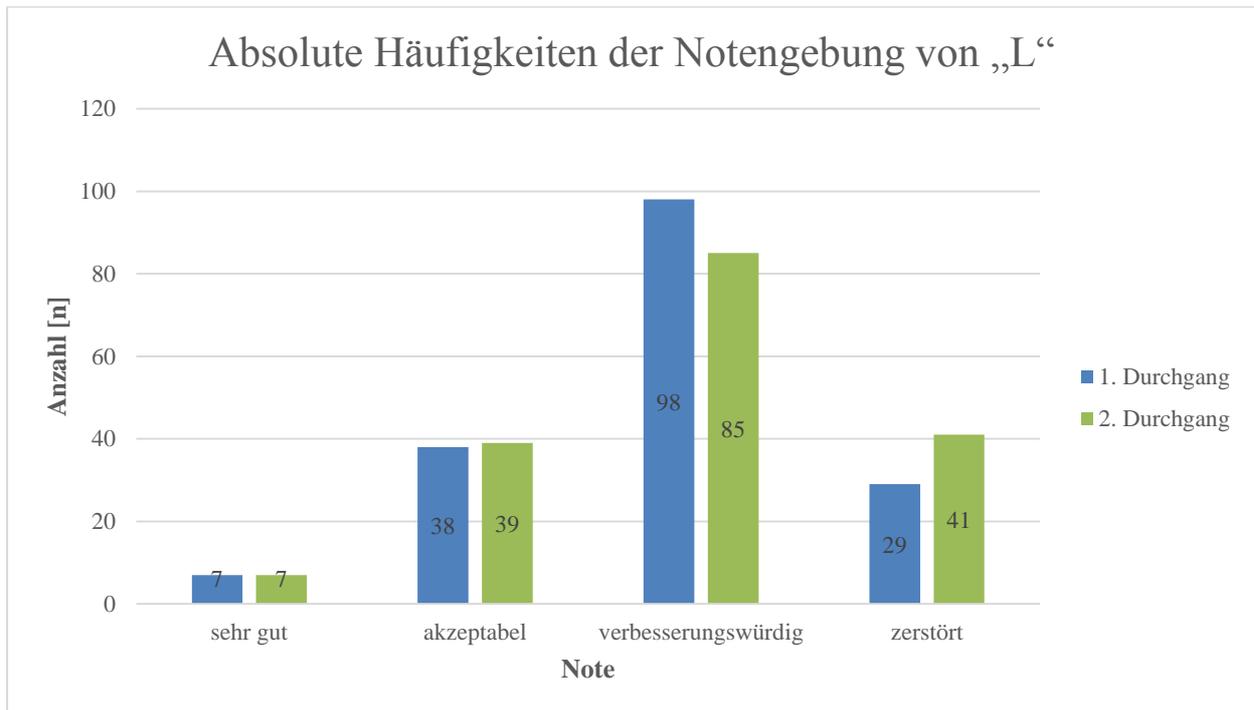


Abbildung 18: Absolute Häufigkeiten der Notengebung von „L“

Die Auswertung der Ergebnisse von „L“ ergibt, dass in beiden Durchgängen das Maximum der absoluten Häufigkeiten die Note *verbesserungswürdig* ($n_1 = 98$ und $n_2 = 85$) stellt. Die Note *sehr gut* erhält jeweils sieben Nennungen. Es kommt zu einer Steigerung der Note *zerstört* von $n_1 = 29$ auf $n_2 = 41$. Die Nennungen der Note *akzeptabel* bleiben nahezu konstant ($n_1 = 38$ und $n_2 = 39$) (siehe Abbildung 18).

Der Weighted Kappa-Koeffizient liegt bei $\kappa_w = 0,465$ mit einem 95%-Konfidenzintervall von $0,36 - 0,57$. Dies ist nach LANDIS und KOCH (1977) eine mittlere Übereinstimmung⁷⁶. Der Test of $H_0: \text{Kappa} = 0$ ergibt einen p-Wert von $< 0,001$, κ_w weicht signifikant von 0 ab.

Beurteiler „K“:

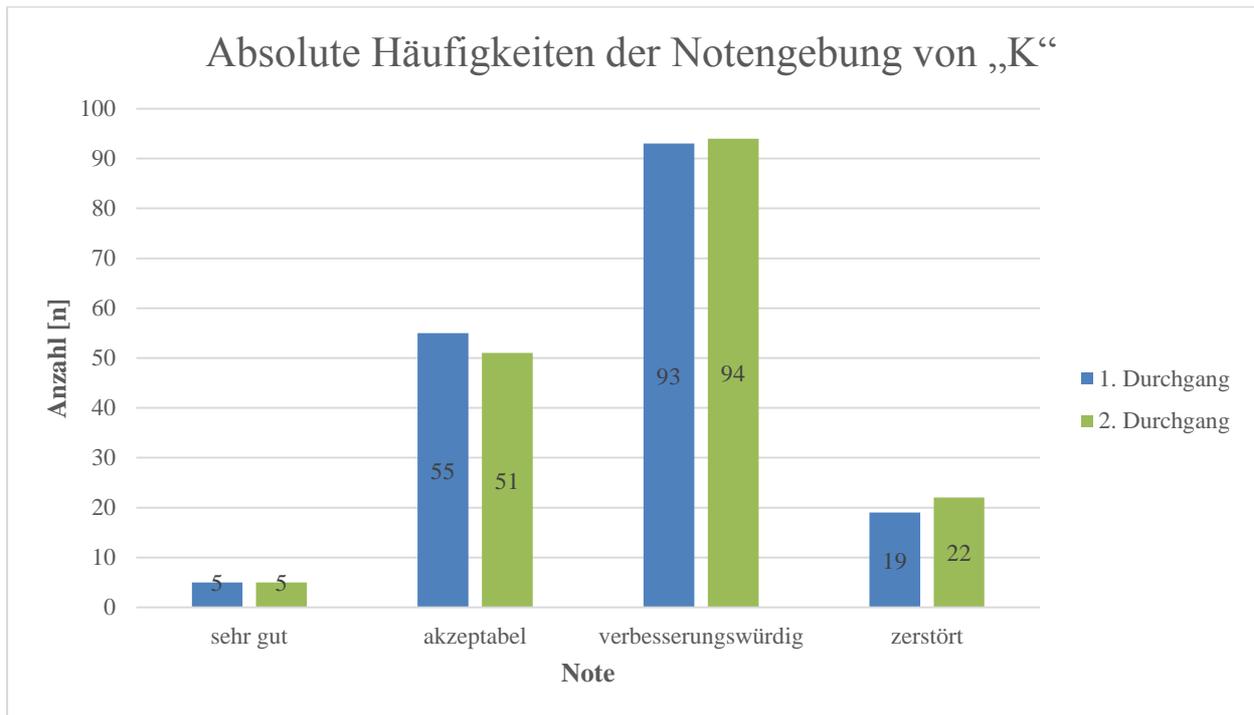


Abbildung 19: Absolute Häufigkeiten der Notengebung von „K“

Die Ergebnisse von Beurteiler „K“ zeigen ein Maximum der absoluten Häufigkeiten in beiden Durchgängen bei den Noten *verbesserungswürdig* ($n_1 = 93$ und $n_2 = 94$) und *akzeptabel* ($n_1 = 55$ und $n_2 = 51$) auf. Die Note *sehr gut* weist geringe Anzahlen auf ($n_1 = 5$ und $n_2 = 5$), *zerstört* erhält $n_1 = 19$ und $n_2 = 22$ Nennungen (siehe Abbildung 19).

Der Weighted Kappa-Koeffizient liegt bei $\kappa_w = 0,78$ mit einem 95%-Konfidenzintervall von $0,69 - 0,87$. Dies ist nach LANDIS und KOCH (1977) eine hohe Übereinstimmung⁷⁶. Der Test of $H_0: \text{Kappa} = 0$ ergab einen p-Wert von $< 0,001$, κ_w weicht signifikant von 0 ab.

Beurteiler „B“:

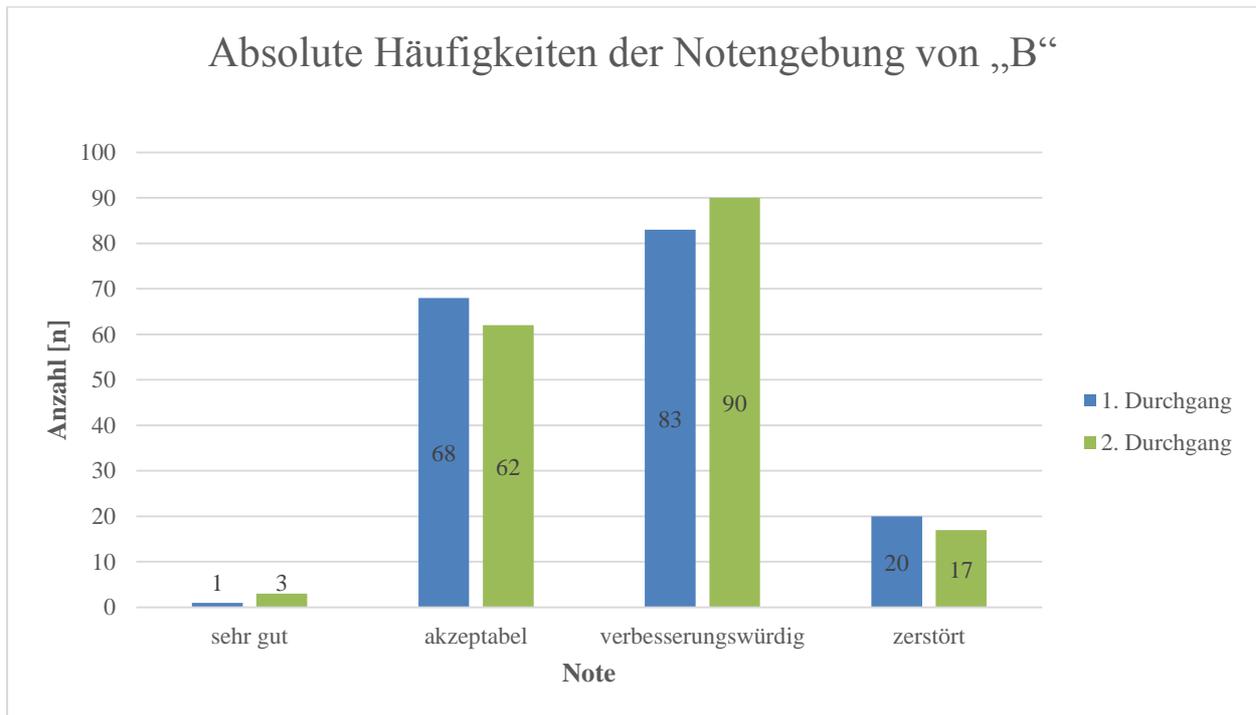


Abbildung 20: Absolute Häufigkeiten der Notengebung von „B“

Die Ergebnisse von Beurteiler „B“ zeigen ein Maximum der absoluten Häufigkeiten in beiden Durchgängen auf, bei den Noten *verbesserungswürdig* ($n_1 = 83$ und $n_2 = 90$) und *akzeptabel* ($n_1 = 68$ und $n_2 = 62$). Die Note *sehr gut* weist im ersten Durchgang von insgesamt 172 Urteilen nur eine Nennung auf. Die Note *zerstört* erhält $n_1 = 20$ und $n_2 = 17$ Nennungen (siehe Abbildung 20).

Der Weighted Kappa-Koeffizient liegt bei $\kappa_w = 0,42$ mit einem 95%-Konfidenzintervall von $0,31 - 0,53$. Dies ist nach LANDIS und KOCH (1977) eine mittlere Übereinstimmung⁷⁶. Der Test of $H_0: \text{Kappa} = 0$ ergibt einen p-Wert von $< 0,001$, κ_w weicht signifikant von 0 ab.

Beurteiler „M“:

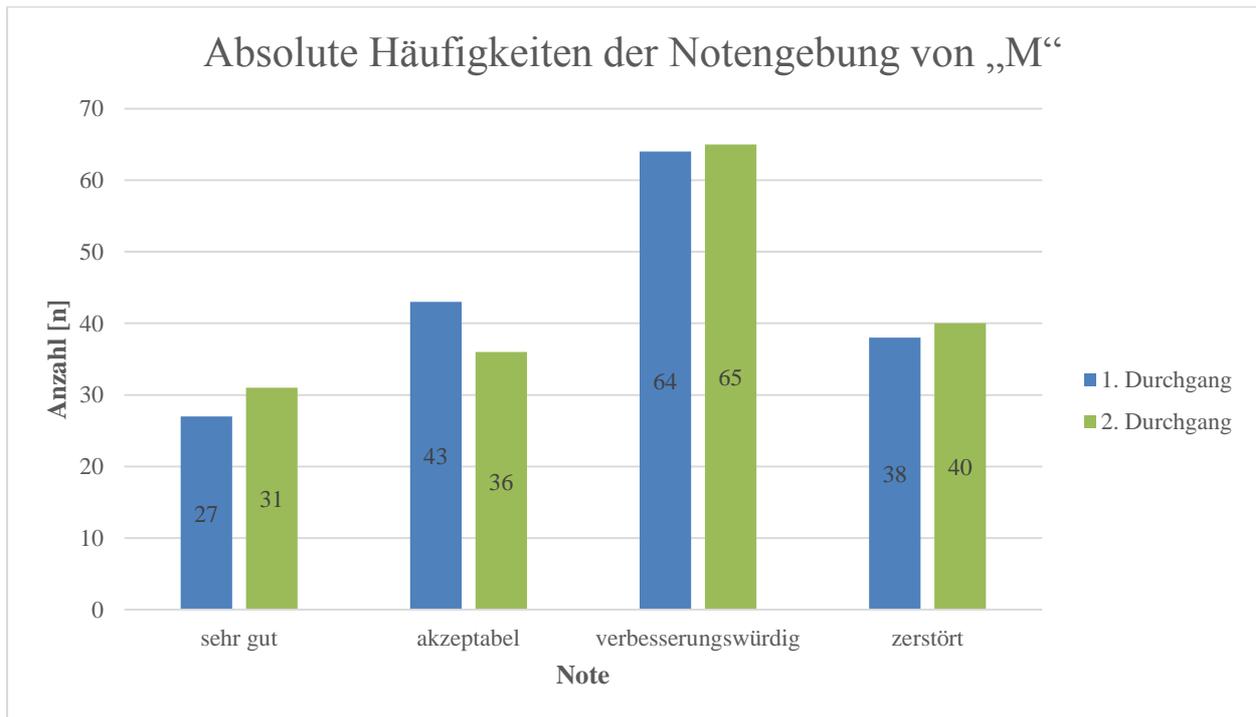


Abbildung 21: Absolute Häufigkeiten der Notengebung von „M“

Die Ergebnisse von „M“ zeigen ein Maximum der absoluten Häufigkeiten in beiden Durchgängen bei der Note *verbesserungswürdig* ($n_1 = 64$ und $n_2 = 65$). Die Note *sehr gut* weist $n_1 = 27$ und $n_2 = 31$ Nennungen auf; die Note *akzeptabel* $n_1 = 43$ und $n_2 = 36$ Nennungen; die Note *zerstört* $n_1 = 38$ und $n_2 = 40$ Nennungen (siehe Abbildung 21).

Der Weighted Kappa-Koeffizient liegt bei $\kappa_w = 0,817$ mit einem 95%-Konfidenzintervall von $0,76 - 0,87$. Dies ist gemäß LANDIS und KOCH (1977) als sehr hohe Übereinstimmung zu beurteilen⁷⁶. Der Test of $H_0: \text{Kappa} = 0$ ergibt einen p-Wert von $< 0,001$, κ_w weicht signifikant von 0 ab.

Vergleich:

Tabelle 5: Weighted Kappa-Koeffizienten aller Beurteiler

Weighted Kappa-Koeffizient	P	A	L	K	B	M
κ_w	0,589	0,571	0,465	0,78	0,42	0,817

Die Ergebnisse zeigen, dass „M“ den höchsten Weighted Kappa-Koeffizienten aufweist. In absteigender Reihenfolge folgen „K“, „P“, „A“, „L“ und „B“ (siehe Tabelle 5).

5.2. Reliabilität der zahnärztlichen Beurteiler in der Notengebung/Interrater-Reliabilität

Tabelle 6: AC1-Koeffizienten der Notengebung der zahnärztlichen Beurteiler

Durchgang	AC1	Standardfehler	p-Wert
1. Durchgang	0,354	0,022	0,000
2. Durchgang	0,371	0,023	0,000

Im ersten Durchgang ist $AC1 = 0,354$ mit einem Standardfehler von 0,022 und weicht signifikant von 0 ab, da $p < 0,001$. Im zweiten Durchgang ist $AC1 = 0,371$ mit einem Standardfehler von 0,023 und weicht signifikant von 0 ab, da $p < 0,001$ (siehe Tabelle 6).

Beide Durchgänge weisen einen AC1-Koeffizienten von 0,2 – 0,4 auf, was nach LANDIS und KOCH (1977) einer geringen Übereinstimmung entspricht⁷⁶.

5.3. Reliabilität der Beurteiler hinsichtlich des Bestehens/Intrarater-Reliabilität

Tabelle 7: Cohens Kappa-Koeffizienten aller Beurteiler hinsichtlich des Bestehens

Beurteiler	κ	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
P	0,822	0,721 – 0,923	< 0,001	sehr hoch
A	0,634	0,5 – 0,767	< 0,001	hoch
L	0,501	0,344 – 0,659	< 0,001	mittel
K	0,806	0,668 – 0,945	< 0,001	sehr hoch
B	0,486	0,276 – 0,695	< 0,001	mittel
M	0,9	0,823 – 0,979	< 0,001	sehr hoch

In Bezug auf die Reliabilität erreicht „M“ den höchsten Wert, gefolgt von „P“, „K“, „A“, „L“ und „B“ (siehe Tabelle 7).

5.4. Reliabilität der zahnärztlichen Beurteiler hinsichtlich des Bestehens/Interrater-Reliabilität

Tabelle 8: AC1-Koeffizienten aller Beurteiler hinsichtlich des Bestehens

Durchgang	AC1	Standardfehler	p-Wert
1. Durchgang	0,715	0,026	< 0,001
2. Durchgang	0,671	0,027	< 0,001

Im ersten Durchgang ist $AC1 = 0,715$ mit einem Standardfehler von 0,026 und weicht signifikant von 0 ab, da $p < 0,001$. Im zweiten Durchgang ist $AC1 = 0,671$ mit einem Standardfehler von 0,027 und weicht signifikant von 0 ab, da $p < 0,001$ (siehe Tabelle 8).

Beide Durchgänge weisen einen AC1-Koeffizienten von 0,6 – 0,8 auf, was nach LANDIS und KOCH (1977) einer hohen Übereinstimmung entspricht⁷⁶.

5.5. Reliabilität der Beurteiler in den Notenkategorien/Intrarater-Reliabilität

Beurteiler „P“:

Tabelle 9: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „P“

Beurteilungskategorie	κ_w	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
Präparationsgrenze	0,679	0,532 – 0,827	< 0,001	hoch
Oberfläche	0,406	0,146 – 0,666	< 0,001	mittel
Substanzabtrag	0,414	0,211 – 0,617	< 0,001	mittel
Präparationswinkel	0,506	0,304 – 0,707	< 0,001	mittel

Die in Bezug auf die Reliabilität am besten beurteilte Kategorie von „P“ ist die Notenkategorie *Präparationsgrenze* ($\kappa_w = 0,679$). Die verbliebenen drei Kategorien weisen eine Reliabilität zwischen $\kappa_w = 0,4$ und $\kappa_w = 0,6$ auf. Als schwächste Kategorie wurde bei „P“ die Notenkategorie *Oberfläche* ($\kappa_w = 0,406$) ermittelt (siehe Tabelle 9).

Beurteiler „A“:

Tabelle 10: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „A“

Beurteilungskategorie	κ_w	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
Präparationsgrenze	0,643	0,474 – 0,813	< 0,001	hoch
Oberfläche	0,455	0,238 – 0,672	< 0,001	mittel
Substanzabtrag	0,540	0,360 – 0,720	< 0,001	mittel
Präparationswinkel	0,513	0,354 – 0,672	< 0,001	mittel

Die in Bezug auf die Reliabilität am besten beurteilte Kategorie von Beurteiler „A“ ist die Notenkategorie *Präparationsgrenze* ($\kappa_w = 0,643$). Die verbliebenen drei Kategorien weisen eine Reliabilität zwischen $\kappa_w = 0,4$ und $\kappa_w = 0,6$ auf. Als schwächste Kategorie wurde bei „A“ die Notenkategorie *Oberfläche* ($\kappa_w = 0,455$) ermittelt (siehe Tabelle 10).

Beurteiler „L“:

Tabelle 11: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „L“

Beurteilungskategorie	κ_w	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
Präparationsgrenze	0,447	0,234 – 0,659	< 0,001	mittel
Oberfläche	0,438	0,168 – 0,709	< 0,001	mittel
Substanzabtrag	0,488	0,295 – 0,681	< 0,001	mittel
Präparationswinkel	0,369	0,163 – 0,576	< 0,001	gering

Die in Bezug auf die Reliabilität am besten beurteilte Kategorie von „L“ ist die Notenkategorie *Substanzabtrag* ($\kappa_w = 0,488$). Die verbliebenen drei Kategorien weisen eine Reliabilität zwischen $\kappa_w = 0,3$ und $\kappa_w = 0,5$ auf. Als schwächste Kategorie wurde bei „L“ die Notenkategorie *Oberfläche* ($\kappa_w = 0,438$) ermittelt (siehe Tabelle 11).

Beurteiler „K“:

Tabelle 12: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „K“

Beurteilungskategorie	κ_w	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
Präparationsgrenze	0,706	0,472 – 0,939	< 0,001	hoch
Oberfläche	0,873	0,739 – 1,0	< 0,001	sehr hoch
Substanzabtrag	0,684	0,473 – 0,895	< 0,001	hoch
Präparationswinkel	0,788	0,635 – 0,941	< 0,001	hoch

Die in Bezug auf die Reliabilität am besten beurteilte Kategorie von „K“ ist die Notenkategorie *Oberfläche* ($\kappa_w = 0,873$). Die verbliebenen drei Kategorien weisen eine Reliabilität zwischen $\kappa_w = 0,6$ und $\kappa_w = 0,8$ auf. Als schwächste Kategorie wurde bei „K“ die Notenkategorie *Substanzabtrag* ($\kappa_w = 0,684$) ermittelt (siehe Tabelle 12).

Beurteiler „B“:

Tabelle 13: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „B“

Beurteilungskategorie	κ_w	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
Präparationsgrenze	0,6	0,415 – 0,785	< 0,001	mittel
Oberfläche	0,237	-0,006 – 0,479	0,063	gering
Substanzabtrag	0,332	0,106 – 0,557	0,002	gering
Präparationswinkel	0,426	0,189 – 0,663	< 0,001	mittel

Die in Bezug auf die Reliabilität am besten beurteilte Kategorie von „B“ ist die Notenkategorie *Präparationsgrenze* ($\kappa_w = 0,6$). Die verbliebenen drei Kategorien weisen eine Reliabilität zwischen $\kappa_w = 0,2$ und $\kappa_w = 0,5$ auf. Als schwächste Kategorie wurde bei „B“ die Notenkategorie *Oberfläche* ($\kappa_w = 0,237$) mit einem p-Wert von 0,063 ermittelt. Es ist darauf hinzuweisen, dass dieser nur knapp oberhalb des zugrunde gelegten Signifikanzniveaus von 5% liegt (siehe Tabelle 13).

Beurteiler „M“:

Tabelle 14: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien „M“

Beurteilungskategorie	κ_w	95%-Konfidenzintervall	p-Wert	Interpretation nach LANDIS und KOCH (1977) ⁷⁶
Präparationsgrenze	0,688	0,524 – 0,852	< 0,001	hoch
Oberfläche	0,393	0,177 – 0,61	0,0015	gering
Substanzabtrag	0,907	0,781 – 1,0	< 0,001	sehr hoch
Präparationswinkel	0,857	0,714 – 0,999	< 0,001	sehr hoch

Die in Bezug auf die Reliabilität am besten beurteilte Kategorie von „M“ ist die Notenkategorie *Substanzabtrag* ($\kappa_w = 0,907$). Die verbliebenen drei Kategorien weisen eine Reliabilität zwischen $\kappa_w = 0,3$ und $\kappa_w = 0,9$ auf. Als schwächste Kategorien wurden bei „M“ die Notenkategorien *Präparationsgrenze* ($\kappa_w = 0,688$) und *Oberfläche* ($\kappa_w = 0,393$) ermittelt (siehe Tabelle 14).

Vergleich:

Tabelle 15: Vergleich Weighted Kappa-Koeffizienten der Notenkategorien aller Beurteiler

Beurteilungskategorie	P	A	L	K	B	M
Präparationsgrenze	0,679	0,643	0,447	0,706	0,600	0,688
Oberfläche	0,406	0,455	0,438	0,873	0,238	0,393
Substanzabtrag	0,414	0,540	0,488	0,684	0,332	0,907
Präparationswinkel	0,506	0,513	0,369	0,788	0,426	0,857

Bei „P“, „A“ und „B“ stellt sich die *Präparationsgrenze* als beste Kategorie hinsichtlich der Reliabilität dar. „P“ und „A“ weisen in den verbliebenen drei Kategorien Reliabilitätswerte zwischen $\kappa_w = 0,406$ und $\kappa_w = 0,540$ auf, während die Werte von „B“ zwischen $\kappa_w = 0,238$ bei *Oberfläche* und $\kappa_w = 0,426$ bei *Präparationswinkel* liegen. Die Ergebnisse von „L“ zeigen in allen Reliabilitätswerte zwischen $\kappa_w = 0,369$ und $\kappa_w = 0,488$, wobei *Substanzabtrag* mit einem Weighted Kappa-Koeffizienten von $\kappa_w = 0,488$ knapp den höchsten Wert erzielt. „K“ erzielt den höchsten Wert in Kategorie *Oberfläche* $\kappa_w = 0,873$. Die Analysesoftware erreicht den deutlich besten Wert in der Kategorie *Substanzabtrag* ($\kappa_w = 0,907$). *Oberfläche* fällt deutlich mit $\kappa_w = 0,393$ hinter *Präparationsgrenze* ($\kappa_w = 0,688$) zurück (siehe Tabelle 15).

5.6. Reliabilität der zahnärztlichen Beurteiler in den Notenkategorien/Interrater-Reliabilität

1. Durchgang:

Tabelle 16: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 1. Durchgang

Beurteilungskategorie	AC1	Standardfehler	p-Wert
Präparationsgrenze	0,339	0,041	< 0,001
Oberfläche	0,516	0,043	< 0,001
Substanzabtrag	0,324	0,049	< 0,001
Präparationswinkel	0,241	0,034	< 0,001

2. Durchgang:

Tabelle 17: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 2. Durchgang

Beurteilungskategorie	AC1	Standardfehler	p-Wert
Präparationsgrenze	0,317	0,037	< 0,001
Oberfläche	0,546	0,048	< 0,001
Substanzabtrag	0,382	0,046	< 0,001
Präparationswinkel	0,248	0,039	< 0,001

In beiden Durchgängen ist die Interreliabilität unter den zahnärztlichen Beurteilern in der Notenkategorie *Oberfläche* am höchsten.

Der AC1-Koeffizient beträgt im ersten Durchgang 0,516 (Standardfehler: 0,043, $p < 0,001$) und im zweiten Durchgang 0,546 (Standardfehler: 0,048, $p < 0,001$). Beides ist nach LANDIS und KOCH (1977) als mittlere Übereinstimmung einzuordnen⁷⁶. Alle weiteren Notenkategorien (AC1 = 0,248 – 0,382) fallen in die Kategorie einer geringen Übereinstimmungsrate (siehe Tabelle 16 und 17).

5.7. Reliabilität der zahnärztlichen Beurteiler in den Notenkategorien hinsichtlich des Bestehens/Interrater-Reliabilität

1. Durchgang:

Tabelle 18: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 1. Durchgang hinsichtlich des Bestehens

Beurteilungskategorie	AC1	Standardfehler	p-Wert
Präparationsgrenze	0,671	0,052	< 0,001
Oberfläche	0,951	0,021	< 0,001
Substanzabtrag	0,451	0,066	< 0,001
Präparationswinkel	0,688	0,056	< 0,001

2. Durchgang:

Tabelle 19: AC1-Koeffizienten der Notenkategorien der zahnärztlichen Beurteiler im 2. Durchgang hinsichtlich des Bestehens

Beurteilungskategorie	AC1	Standardfehler	p-Wert
Präparationsgrenze	0,642	0,046	< 0,001
Oberfläche	0,924	0,027	< 0,001
Substanzabtrag	0,423	0,078	< 0,001
Präparationswinkel	0,601	0,057	< 0,001

In beiden Durchgängen ist die Interreliabilität unter den zahnärztlichen Beurteilern in der Notenkategorie *Oberfläche* am höchsten.

Der AC1-Koeffizient beträgt im ersten Durchgang 0,951 (Standardfehler: 0,021, $p < 0,001$) und im zweiten Durchgang 0,924 (Standardfehler: 0,027, $p < 0,001$). Beides ist nach LANDIS und KOCH (1977) als sehr hohe Übereinstimmung einzuordnen⁷⁶. Alle weiteren Notenkategorien ($AC1 = 0,423 - 0,642$) fallen in die Kategorie einer mittleren beziehungsweise hohen Übereinstimmungsrate (siehe Tabelle 18 und 19).

5.8. Fragebögen der Studenten

Allgemeiner Zustand des Betreuungsverhältnisses:

79,1% (n = 34) der Befragten bewerten das gegenwärtige Betreuungsverhältnis positiv. Neun von 43 (20,9%) Befragten verneinen dies eingeschränkt. Ferner stimmen 90,7% (n = 39) der Aussage, dass die wissenschaftlichen Mitarbeiter manchmal überfordert wirken, eher nicht beziehungsweise gar nicht zu. Der Aussage, dass die Wartezeit, um einen wissenschaftlichen Mitarbeiter konsultieren zu können, in der Regel zu lang sei, verneinen 54,8% (n = 23) voll beziehungsweise eingeschränkt. 64,3% (n = 27) der Studenten stimmen der Aussage voll beziehungsweise eingeschränkt zu, dass ein wissenschaftlicher Mitarbeiter während der Kurszeit immer anwesend ist. 90,7% (n = 39) der Studenten stimmen der Aussage, dass das aktuell verwendete Notensystem sinnvoll ist, eingeschränkt beziehungsweise voll zu.

Einstellungen, Erwartungen und Wünsche der Studenten:

Alle Befragten geben an, dass ihnen Objektivität bei der Bewertung wichtig ist. 34 (79,1%) stimmen der Aussage dabei voll zu. Der Aussage, dass die Notengebung einen entscheidenden Einfluss auf die Motivation hat, stimmen 83,7% (n = 36) der Studenten voll beziehungsweise eingeschränkt zu. Die volle Zustimmung zur Aussage, dass eine gerechte und nachvollziehbare Notenvergabe vorliegt, trifft kein Student. 57,2% (n = 24) stehen dieser Aussage negativ gegenüber, 42,9% (n = 18) äußerten sich eingeschränkt positiv. 38 der 43 (88,4%) teilnehmenden Studenten vertrauen dem Urteil der wissenschaftlichen Mitarbeiter, wobei 30 (69,8%) dieser Aussage eingeschränkt zustimmen. 81% (n = 34) der Befragten lehnen die Eröffnung eines Klageverfahrens bei Nichtbestehen/Exmatrikulation voll beziehungsweise eingeschränkt ab. 86% (n = 37) sprechen sich voll beziehungsweise eingeschränkt für objektivere Verhältnisse in den vorklinischen Kursen aus.

Unterschiede in der Bewertungsqualität:

Eine deutliche Mehrheit der Studenten (97,7%, n = 42) stimmt der Aussage zu, einen Unterschied in der Kritik zu bemerken; 55,8% (n = 24) davon stimmen der Aussage uneingeschränkt zu. Eine geringere Qualität in der Kritik von unerfahrenen wissenschaftlichen Mitarbeitern im Vergleich zu erfahreneren Kollegen bestätigen voll beziehungsweise eingeschränkt 59,9% (n = 25) der Befragten. 53,5% (n = 23) der Studenten stimmen der Aussage zu, dass die Kritik der wissenschaftlichen Mitarbeiter von der Tagesform abhängig ist, wobei 14% (n = 6) dieser Aussage voll zustimmen. 97,7% (n = 42) stimmen der Aussage zu, dass dieselbe Arbeit von unterschiedlichen wissenschaftlichen Mitarbeitern divergierend bewertet

wird. 28 Studenten (65,1%) stimmen dieser Aussage voll zu. 46,5% (n = 20) der Befragten bemerken einen Unterschied in der Beurteilung einer Arbeit bei demselben wissenschaftlichen Mitarbeiter. Sechs Befragte (14%) stimmen der Aussage voll zu.

Persönliche Faktoren:

62,8% (n = 27) der Studenten geben in der Befragung an, dass sie sich manchmal unwohl beim Vorzeigen einer Arbeit fühlen. Sie stimmen dieser Aussage dabei voll beziehungsweise eingeschränkt zu. 28 von 43 Befragten (65,1%) stimmen eher nicht beziehungsweise gar nicht mit der Aussage überein, dass einzelne Studenten in der Kritik/Benotung bevorzugt werden.

41,9% (n = 18) der Studenten geben an, dass die Kritik manchmal herablassend ist, zehn Befragte (23,3%) stimmen ihr voll zu. 74,4% (n = 32) der Befragten haben den Eindruck, dass sich der wissenschaftliche Mitarbeiter ein Bild von ihnen gemacht hat. Dieser Aussage stimmen 24 (55,8%) eingeschränkt zu. Bei der Auswertung der Aussage „Ich habe das Gefühl, dass einzelne Studierende in der Kritik/Bewertung benachteiligt werden“ verteilen sich die Stimmen jeweils nahezu gleichmäßig auf die zustimmenden beziehungsweise ablehnenden Antwortkategorien (48,8% (n = 21) „stimme voll zu“/„stimme eher zu“ beziehungsweise 51,2% (n = 22) „stimme gar nicht zu“/„stimme eher nicht zu“).

Umgang mit prepCheck 1.1:

95,2% (n = 20) empfinden die Bewertung durch einen Computer als gerecht. 52,4% (n = 9) stimmen der Aussage voll zu. 57,1% (n = 12) der Studenten stimmen der Aussage, dass Computerurteile generell vertrauenswürdiger als menschlichen Bewertungen sind, gar nicht beziehungsweise eher nicht zu. Zehn von 17 (47,6%) Befragten bevorzugen eine Betreuung durch einen wissenschaftlichen Mitarbeiter im Laboralltag. 14 von 18 Studenten (77,8%) favorisieren eine Bewertung im Testat durch eine Analysesoftware.

5.9. Fragebögen der Zahnärzte

Allgemeiner Zustand des Betreuungsverhältnisses:

Drei der fünf (60%) Zahnärzte stimmen der Aussage, dass das Betreuungsverhältnis ausreichend ist, gar nicht beziehungsweise eher nicht zu. Drei Zahnärzte (60%) stimmen der Aussage, dass die Betreuung der Kurse mit Stress verbunden ist, voll beziehungsweise eingeschränkt zu. Einer Überforderung bezüglich einer objektiven Beurteilung stimmen zwei (40%) Befragte voll beziehungsweise eingeschränkt zu. Alle fünf Befragten geben darüber hinaus an, den Studenten Verbesserungsvorschläge während der Beurteilung einer Arbeit zu erteilen. Vier (80%) stimmen der Aussage voll zu. Die Aussage, dass eine intensive Betreuung einzelner Studenten zeitlich möglich ist, verneinen vier (80%) der Befragten. Jeweils zwei (40%) der Befragten stimmen dabei gar nicht beziehungsweise eher nicht zu. Alle fünf Befragten stimmen der Aussage, dass das verwendete sechstufige Notensystem sinnvoll ist, voll beziehungsweise eingeschränkt zu.

Objektivität:

Alle Zahnärzte bejahen die Aussage, dass eine Präparation, die von mehreren Kollegen mit der gleichen Note beurteilt wurde, eine objektive Bewertung erhalten hat. Drei (60%) stimmen dieser Aussage dabei voll zu. Alle Befragte geben an, sich bei Bewertungen nicht von Kollegenmeinungen beeinflussen zu lassen. Vier (80%) stehen dieser Aussage dabei eingeschränkt ablehnend gegenüber. Keiner der fünf Befragten vertritt die Meinung, dass die Qualität der objektiven Bewertung unter allen Kollegen das gleiche Niveau aufweist. Drei (60%) stimmen der Aussage dabei gar nicht zu.

Anonymität:

Vier von fünf (80%) Befragten stimmen der Aussage voll beziehungsweise eingeschränkt zu, dass eine Anonymisierung die Objektivität von Bewertungsprozessen steigert. Vier (80%) erteilen der Aussage, dass sich auffällige Studenten im Kursverlauf mit ihrer Platznummer einprägen, ihre volle beziehungsweise eingeschränkte Zustimmung. Drei (60%) stimmen der Aussage, dass Erwartungshaltungen gegenüber einigen Platznummern vorliegen, vollkommen beziehungsweise eingeschränkt zu.

Beeinflussende Faktoren I:

Alle fünf stimmen der Aussage zu, dass es typische Anfängerfehler gibt. Vier (80%) stimmen der Aussage dabei uneingeschränkt zu. Ferner stimmen alle Befragten der Aussage, dass eine Einschätzung der Studenten nach einer gewissen Zeitspanne möglich ist, voll beziehungsweise eingeschränkt zu. Zwei Befragte (40%) bejahen die Aussage eingeschränkt, dass mit Hilfe des Ersteindrucks auf ein Zurechtkommen eines Studenten im Kurs zu schließen ist. Vier (80%) stimmen der Aussage eingeschränkt zu, dass ein unsauberer Arbeitsplatz mit einer ungenauen Arbeitsweise assoziiert ist. Eine Einflussnahme von Sympathie oder Antipathie gegenüber Studenten auf die Notengebung verneinen 80% (n = 4) der Befragten vollkommen beziehungsweise eingeschränkt. Ein Befragter (20%) stimmt der Aussage voll zu, dass Sympathie die Notengebung positiv beeinflusst. Alle stimmen der Aussage zu, dass sie einem plötzlich sehr guten Abschneiden eines vorher leistungsschwachen Studenten misstrauisch gegenüberstehen. Drei (60%) stimmen dieser Aussage dabei voll zu. Der Aussage, dass es zu einer negativen Nachwirkung einer unzureichend beurteilten Kategorie auf die Benotung weiterer Beurteilungskategorien innerhalb derselben Leistungsermittlung kommt, stimmt keiner zu. Drei (60%) geben dabei „stimme gar nicht zu“ an. Drei (60%) stimmen der Aussage, dass zuvor erbrachte Leistungen gegenwärtige Benotungen beeinflussen, eingeschränkt zu.

Beeinflussende Faktoren II:

Die meisten Nennungen für die Beeinflussung einer objektiven Beurteilung erhalten die Faktoren „Betriebsblindheit“, Stress/Frust und Umgebungsverhältnisse (Ablenkungen, Lärm, Lichtverhältnisse) (jeweils n = 4). In absteigender Reihenfolge folgen Ermüdung (n = 3), eigene Erfahrungswerte (n = 3) und Sympathie/Antipathie (n = 2).

Eigene Meinung:

Vier (80%) Zahnärzte stimmen der Aussage voll zu, dass die jetzige vorklinische Ausbildung einen wichtigen, unverzichtbaren Bestandteil der zahnmedizinischen Ausbildung darstellt. Einer (20%) stimmt dieser Aussage eingeschränkt zu. Zwei (40%) der Befragten stimmen der Aussage, dass die vorklinische Ausbildung eine Abhärtungsmaßnahme darstellt, voll zu; drei (60%) Befragte stimmen dieser Aussage eher nicht zu. Drei (60%) Zahnärzte stimmen der Aussage, dass ein Verbesserungsbedarf besteht, voll zu; zwei (40%) sehen diese Aussage eingeschränkt kritisch.

6. DISKUSSION

6.1. Material und Auswahl der Präparationsform

6.1.1. Zahnmodelle

Für die Durchführung der Untersuchung war der Einsatz industriell gefertigter Zahnmodelle obligat, um die Ermittlung des *Substanzabtrags* via Goldstandard in der Analysesoftware zu ermöglichen. Verwendet wurden Modelle der Firma KaVo (KaVo Dental GmbH, Biberach, Deutschland), die eine eugnathe Bissituation darstellen und somit einen standardisierten Ablauf der Präparation ermöglichen. Allerdings unterliegen die am Modell verwendeten Gingivamanschetten zum Teil erheblichen Ausführungsschwankungen. Die Gründe dafür liegen zum einen in fortschreitenden Alterungsprozessen der Manschetten und zum anderen in den Fertigungsunterschieden der zur Herstellung verwendeten Gießformen, welche regelmäßig ausgetauscht werden müssen^b. Dies führte dazu, dass alle präparierten Zahnstümpfe in ihren jeweils zugehörigen Modellen beurteilt werden mussten. Ein Austausch der Zähne in andere Modelle führte zu unerwünschten Verschiebungen der Präparationsgrenze von bis zu drei Millimetern und hätte fehlerhafte Beurteilungen in dieser Kategorie aufgrund des zum Teil subgingivalen Verlaufs zu Folge gehabt.

Eine Studie untersuchte den Einfluss verschiedener Gingivamanschetten auf die Ergebnisse der Analysesoftware ED4 Compare (Version 2.0)⁷⁸. Dort wurde ein vorab präparierter und eingescannter Goldstandard in zehn unterschiedliche Zahnmodelle – der gleichen Firma und Ausführung – eingesetzt und erneut eingescannt⁷⁸. Fünf Modelle waren neu, fünf weitere mehrere Jahre alt⁷⁸. Die Ergebnisse zeigen, dass es bei einem Toleranzwert von 0,173 mm zu einer vollständigen Korrelation zwischen den verglichenen Modellen kommt⁷⁸. Zu ähnlichen Ergebnissen kommt auch eine weitere Untersuchung⁷⁹.

Abgesehen von der Lage der Präparationsgrenze führt ein Vergleich der Präparationen zum Goldstandard in prepCheck 1.1 nicht zu nennenswerten Verschiebungen und infolge dessen auch nicht zu unzutreffenden Beurteilungen. Allerdings werden im Rahmen der Nachbarzahnschädigungen erhebliche fehlerhafte Überlagerungen registriert.

^b Information auf schriftliche Nachfrage bei KaVo Dental GmbH, Biberach, Deutschland am 15.05.2017

6.1.2. Präparation

Um die Inhaltsvalidität während der Versuchstestate zu erhöhen, wurde in dieser Studie eine Vollgusspräparation ausgewählt, da die Vorgehensweise den Studenten bereits bekannt war und den gleichen Schwierigkeitsgrad aufwies wie die bereits geübte Präparation des Zahnes 36. Erforderlich war lediglich eine Transferleistung auf den Zahn 26. Zudem lagen die gleichen Rahmenbedingungen (Zeitvolumen) vor wie im Technisch Propädeutischen Kurs.

6.1.3. Verwendung der CEREC-Aufnahmeeinheiten

Aus der Literatur ist bekannt, dass ein sicherer Umgang mit dem computerbasierten Beurteilungssystem eine gewisse Zeitspanne benötigt, bis sich die Abläufe effizient gestalten^{40, 43}. Zudem können auftretende technische Probleme den Anwender frustrieren³¹. Um Verzerrungseffekten dieser Art entgegen zu wirken, fand die Bedienung der Auswertungseinheiten durch geübte Anwender statt. Im Speziellen wurde auf eine selbstständige Verwendung der Omnicam sowie die Festlegung des Präparationsrandes, der Einschubachse und der Okklusallinie durch die Studenten verzichtet, um einen fehlerhaften Einfluss zu reduzieren. Die Beurteilung des ersten und zweiten Durchganges durch zwei verschiedene Anwender ermöglichte darüber hinaus eine anwenderunabhängige Analyse.

6.2. Beurteilungssystem

Um eine erhöhte Anzahl an Beurteilungen zur verbesserten Messung der Reliabilität zu erhalten, wurden die Zähne in vier Kategorien beurteilt. Die Auswahl der Kategorien *Oberfläche*, *Präparationsgrenze*, *Substanzabtrag* und *Präparationswinkel* gibt zum einen prepCheck 1.1 als verfügbare Parameter vor, zum anderen stellen diese darüber hinaus maßgebliche Kriterien für die Qualität einer Präparation dar. Die Verwendung eines limitierten vierstufigen Notensystems (*sehr gut*, *akzeptabel*, *verbesserungswürdig* und *zerstört*) als Abweichung zu der sonst verwendeten sechsstufigen Notenskala begründete sich in der vereinfachten und schnelleren Ausführbarkeit. Die fehlende, aber medizinisch relevante Beurteilungskategorie der Nachbarzahnschädigung wurde durch einen zu geringen *Substanzabtrag* in beiden Beurteilungsbögen erfasst.

6.2.1. Beurteilungsbogen prepCheck 1.1

Da prepCheck 1.1 eine optische Darstellung von Fehlerstellen bietet, eine Beurteilung jedoch dem Anwender obliegt und für diese Untersuchung essentiell war, wurden im Vorfeld Notenabstufungen und Schwellenwerte festgelegt. Es soll an dieser Stelle erwähnt werden, dass

diese Einteilung eine subjektive Einschätzung darstellte. Diese hatte jedoch auf die Ergebnisse der Untersuchung keinen Einfluss, da lediglich die Reliabilität der Analysesoftware überprüft wurde. Der Einfluss subjektiver Einflusstendenzen der zahnärztlichen Beurteiler wurde in der Analyse der Interrater-Reliabilität untersucht. Die Benotung der Subkategorien *Präparationsrand* und *Oberflächenbeschaffenheit* unterlagen allerdings dem Einfluss subjektiver Tendenzen, da diese Auswertung durch den Programmanwender erfolgte. Da es sich dabei um ein neues Beurteilungssystem handelte, welches noch nicht regulär in einem vorklinischen Kurs angewendet wird, erfolgte die Bestimmung der Interrater-Reliabilität ohne die Einbeziehung der Ergebnisse der Analysesoftware.

6.2.2. Beurteilungsbogen zahnärztliche Beurteiler

Die Note *sehr gut* wurde als Ergänzung zum dreistufigen Beurteilungssystem (gut-mittelmäßig-schlecht) hinzugefügt, um Tendenzen in der Notenvergabe eindeutig zu identifizieren und mögliche Hemmschwellen zur Vergabe sehr guter Noten zu lokalisieren. Zusätzlich sollte die Auswahl einer mittleren Beurteilungskategorie vermieden werden, um die Beurteiler zu einer Notentendenz zu veranlassen. Dieses Vorgehen basiert auf der von der Literatur empfohlenen Gestaltung von Ratingskalen bei Fragebögen⁸⁰. Außerdem erforderte das unbekanntes vierstufiges System eine bewusste Auseinandersetzung der Beurteiler basierend auf der mündlichen Einweisung und der Ausfüllhilfe. Dadurch wurde der Zugriff auf bereits festgelegte subjektive Notenkriterien verhindert. Auf Grundlage einer Untersuchung wurde zusätzlich allen Beurteilern ein Gipsmodell des Goldstandards zur Verfügung gestellt¹⁵. Durch die Verwendung der oben genannten Maßnahmen sollte die Interpretationsobjektivität erhöht werden. Trotz Kalibrierung ist jedoch zu erwähnen, dass sich eine exakt definierte Notenschwelle, wie sie zum Teil im Beurteilungsbogen der Software vorhanden ist, nicht erreichen lässt. Durch die Kalibrierungsvorgänge sollte jedoch der Einfluss subjektiver Tendenzen minimiert und eine Vergleichbarkeit der Beurteilungen ermöglicht werden.

6.2.3. Beurteilungsdurchführung

Die Beurteilungssituation wies eine hohe Durchführungsobjektivität auf, da die Bedingungen der unter Aufsicht durchgeführten Benotung für alle Beurteiler identisch waren. Respektive bedeutete dies die gleiche Anzahl an zu beurteilenden Zähnen, die Verwendung gleicher Beurteilungsbögen und ein identisches Zeitvolumen. Darüber hinaus wurde die Auswertungsobjektivität durch die Verwendung von Beurteilungsbögen erhöht. Die Beurteilungsdurchgänge folgten direkt aufeinander, um eine Situation zu schaffen, wie sie

üblicherweise auch während der regulären Kurstestate herrscht, bei der in kurzer Zeit viele Entscheidungen getroffen werden müssen. Dies ermöglicht Rückschlüsse auf die Beurteilungsqualität in Testaten im Rahmen der vorklinischen Kurse. Die Ergebnisse unterliegen damit der Beeinflussung von Ermüdung und Stress. Der Einfluss von Stress, Ermüdung und Mangel wird in der Literatur beschrieben^{55, 81}.

6.3. Statistische Auswertung

Im Rahmen dieser Arbeit wurden zur Herausarbeitung der Reliabilität drei verschiedene Koeffizienten ermittelt. Für dichotome Benotungen, wie sie in der Beurteilungssituation Bestehen/Nichtbestehen vorlagen, wurde – um das Maß der Nichtübereinstimmung zu messen – der Cohens Kappa-Koeffizient verwendet. Bei den Ergebnissen der Reliabilitätsuntersuchung in der Notengebung sollte dem Ausprägungsgrad einer Abweichung innerhalb der Notenskala Rechnung getragen werden, weshalb der Weighted Kappa-Koeffizient ermittelt wurde. Die Interrater-Reliabilität wurde durch den AC1-Koeffizient ermittelt. Um der Entstehung von Bias durch eventuelle Lerneffekte seitens der wissenschaftlichen Mitarbeiter vorzubeugen, wurden die beiden Beurteilungsdurchgänge getrennt voneinander analysiert.

6.4. Diskussion der Ergebnisse

6.4.1. Vergleich der Intrarater-Reliabilität

Im Vergleich der Weighted Kappa-Koeffizienten unter den Untersuchungsergebnissen zur Reliabilität in der Notengebung der einzelnen Beurteiler wird deutlich, dass prepCheck 1.1 den höchsten Kappa-Koeffizienten erreicht. Dennoch bleibt der in der vorliegenden Arbeit ermittelte Reliabilitätskoeffizient hinter den in der Literatur beschriebenen Intrarater-Reliabilitätskoeffizienten zurück, auch wenn in allen Untersuchungen die Analysesoftware stets den höchsten Reliabilitätskoeffizienten erreichte^{4, 5, 7}. Als mögliche Gründe lassen sich unterschiedliche Softwareprogramme, aber auch Unterschiede im Versuchsaufbau aufführen.

Die zahnärztlichen Beurteiler in der durchgeführten Studie weisen sehr unterschiedliche Reliabilitätskoeffizienten auf. Die Werte von „K“, „P“ und „A“ halten dem Vergleich mit der Literatur stand^{4, 5, 7}. Auffallend weit zurück fallen jedoch die Reliabilitätskoeffizienten der Beurteiler „L“ und „B“.

6.4.2. Vergleich der Interrater-Reliabilität der zahnärztlichen Beurteiler

Die AC1-Koeffizienten in der Notengebung der vorliegenden Untersuchung sind mit der Literatur vergleichbar⁵. Der von KATEEB et al. (2016) festgestellte Wert konnte in der vorliegenden Studie nicht erreicht werden, dies deutet auf einen nur geringfügig ausgeprägten interkollegial definierten Beurteilungsmaßstab hin⁴.

Viele verschiedene Studien belegen den positiven Effekt von Kalibrierungen, Verwendung von Vergleichen mit Goldstandards und schriftlich festgelegten Leistungsmaßstäben^{3, 14, 15, 17-23, 25}. Diesen stehen deutlich weniger Untersuchungen, die keine Verbesserungen durch Checklisten nachweisen konnten, gegenüber^{13, 26}. Deswegen sollten zur Identifizierung der effektivsten Kalibrierungsmaßnahmen bezüglich der Reduktion der Interrater-Reliabilität weiterführende Studien durchgeführt werden.

Zusammenfassend lässt sich feststellen, dass die vorliegenden Ergebnisse trotz Kalibrierung und der Verwendung von Ausfüllhilfen die Schlussfolgerung zulassen, dass die Beurteilungsqualität durch zahnärztliche Betreuer erheblichen Schwankungen unterliegt und dies im wesentlichen Maße von der individuellen Beurteilungsqualität der Beurteiler und nicht von der Qualität der Leistung abhängt. Zu diesem Ergebnis kam auch eine weitere Studie⁸².

Die Ergebnisse hinsichtlich einer insgesamt defizitären Intrarater- und/oder Interrater-Reliabilität der zahnärztlichen Beurteiler stimmen mit den entsprechenden Resultaten der Literatur überein²⁻⁹.

6.4.3. Bestehen/Nichtbestehen

Der Vergleich der Cohens Kappa-Koeffizienten zeigt, dass erneut die Analysesoftware „M“ den besten Wert aufweist. Die weitere Reihenfolge ähnelt dem Ergebnis der Weighted Kappa-Koeffizienten der Intrarater-Reliabilität in der Notengebung. Allerdings zeigt der erzielte Wert von „P“ in dieser Analyse ein geringfügig besseres Ergebnis als „K“ und damit den zweitbesten Wert auf. Die ohnehin schon hohen Reliabilitätskoeffizienten von „M“ und „K“ in der Notengebung verbessern sich bei der Analyse hinsichtlich des Bestehens jeweils geringfügig. In absteigender Reihenfolge folgen „K“, „A“, „L“ und „B“. Auffällig ist, dass sich die Ergebnisse der Beurteiler „P“ und „A“ deutlich gegenüber denen zur Intrarater-Reliabilität in der Notengebung verbessern, während die Werte der Hauptbetreuer des Kurses – „B“ und „L“ – weiterhin zwischen $\kappa = 0,4$ und $\kappa = 0,5$ liegen und damit eine inkonstante Beurteilungsweise andeuten. Diese Ergebnisse lassen die Schlussfolgerungen zu, dass es bei „P“ und „A“ zu weniger bestehensrelevanten Urteilsänderungen kam.

Die AC1-Koeffizienten zur Identifizierung der Interrater-Reliabilität hinsichtlich des Bestehens sind im Vergleich zur Interrater-Reliabilität der Notengebung aufgrund der verringerten Variablenauswahl höher. Insgesamt lassen sich, wie in der Literatur beschrieben, auch in dieser Untersuchung diskordante Urteile nachweisen^{3, 15}.

Weitergehende Studien mit einer größeren Anzahl an Beurteilern zur Untersuchung der starken Unterschiede zwischen den Beurteilern sollten folgen.

6.4.4. Berufserfahrung

In der vorliegenden Studie fällt auf, dass „L“ und „B“, trotz langjähriger Erfahrung in Beruf und Lehre, besonders niedrige Reliabilitätskoeffizienten sowohl in der Notengebung als auch hinsichtlich des Bestehens aufweisen. Beurteiler „B“ fungierte zum Zeitpunkt der Untersuchung zudem als fachlicher Leiter der vorklinisch-prothetischen Kurse. „P“ wies zum Zeitpunkt der Untersuchung mit 28 Jahren die längste Berufserfahrung auf. Die Ergebnisse zeigen, dass „P“ der Beurteiler mit der höchsten (Bestehen) beziehungsweise zweithöchsten Reliabilität (Notengebung) unter den zahnärztlichen Beurteilern ist. Beachtenswerterweise erzielt „K“ mit deutlich weniger Berufserfahrung und nur sporadischer Tätigkeit im Kurs ähnlich hohe Werte, welche die Werte von „P“ in der Notengebung hinsichtlich der Reliabilität weit übertreffen. Beurteiler „A“ als Berufsanfänger und ohne Berufserfahrung erzielt höhere Reliabilitätskoeffizienten als „B“ und „L“. In der vorliegenden Untersuchung kann demnach – entsprechend der Literatur – kein eindeutiger Zusammenhang zwischen Berufs- beziehungsweise Ausbildungserfahrung und der Reliabilität eines Beurteilers nachgewiesen werden^{13, 14}. Dementsprechend ist auch das Vorliegen eines Laienfehlers nicht nachweisbar.

6.4.5. Schwierigstes Kriterium/Notenkategorie

Mit den Ergebnissen von MAYS et al. (2016) übereinstimmend, weisen die Beurteiler der vorliegenden Studie den schwächsten AC1-Koeffizienten (Interrater-Reliabilität) in der Kategorie *Präparationswinkel* sowie den höchsten in der Kategorie *Oberfläche* auf¹⁶. Die interkollegiale Übereinstimmungsrate ist demnach in der Kategorie *Präparationswinkel* auffallend gering, während in der Kategorie *Oberfläche* eine höhere Übereinstimmungsrate vorliegt. Die AC1-Koeffizienten zur Identifizierung der Interrater-Reliabilität hinsichtlich des Bestehens in den Notenkategorien sind im Vergleich zur Interrater-Reliabilität der Notengebung aufgrund der verringerten Variablenauswahl insgesamt höher. Auch hier ist *Oberfläche* die Beurteilungskategorie mit dem höchsten AC1-Koeffizienten. Anders als bei der Beurteilung hinsichtlich der Notengebung zeigt *Substanzabtrag* den geringsten AC1-Koeffizienten. Es kam

demnach insbesondere in der Beurteilung der Kategorie *Substanzabtrag* zu vermehrten Beurteilungen, die Einfluss auf das Bestehen einer Prüfung hatten.

Ergänzend dazu stellt der *Präparationswinkel* auch in der Intrarater-Reliabilität bei keinem der Beurteiler die am verlässlichsten beurteilte Kategorie dar. Insgesamt betrachtet lässt sich sagen, dass „P“, „A“ und „B“ die Notenkategorie *Präparationsgrenze* am reliabelsten beurteilt haben. Bei „L“ wurde *Substanzabtrag* am reliabelsten beurteilt, nur unwesentlich schlechter wurde ebenfalls die Kategorie *Oberfläche* beurteilt. Einzig bei Beurteiler „K“ weist die Kategorie *Oberfläche* die höchste Übereinstimmungsrate auf. Dies deutet darauf hin, dass die *Präparationsgrenze* einfacher, sicherer und somit reliabler zu beurteilen ist. Eine Studie beschreibt darüber hinaus die Korrelation zwischen visueller und digitaler Beurteilung⁸³. Die Kategorien *Präparationsgrenze*, *okklusale Reduktion* und *Präparationswinkel* erzielten dabei zufriedenstellende Korrelationswerte⁸³. Eine weitere Untersuchung vergleicht einen Gesamt-Übereinstimmungswert, welcher keine spezifischen Kriterien enthält⁸⁴. Nur in zwei Fällen liegt eine signifikante moderate Übereinstimmung vor⁸⁴.

Weiterhin zeigen die Ergebnisse, dass noch ein erheblicher Verbesserungsbedarf in allen zu beurteilenden Kategorien hinsichtlich der Interrater-Reliabilität besteht.

Bei der Analysesoftware „M“ fallen die Kategorien *Präparationsgrenze* und *Oberfläche* mit schwachen Weighted Kappa-Koeffizienten auf, wobei der Reliabilitätswert der *Oberfläche* noch einmal deutlich niedriger liegt im Vergleich zur *Präparationsgrenze*. Dabei handelt es sich bei beiden Kategorien um diejenigen, welche anwenderabhängige Beurteilungen in den Subkategorien beinhalten. Dies lässt die Annahme zu, dass die anwenderbezogene Beurteilung des *Präparationsrandes* mit dem digitalen Lineal zu reliableren Ergebnissen führt als die Beurteilung der *Oberflächenbeschaffenheit* durch die Kalibrierungshilfen. Dies lässt wiederum die Vermutung zu, dass der Einfluss von anwenderbezogenen subjektiven Beurteilungen mit einer Verringerung des Reliabilitätskoeffizienten korreliert.

6.4.6. Ermüdung

Die Ermittlung des Einflusses von Ermüdung erfolgt auf Basis der Reliabilitätskoeffizienten sowie einer nachweisbaren Benotungstendenz beim Vergleich des ersten und zweiten Durchganges (strenger oder milder). Zusätzlich werden die Ergebnisse der Fragebögen der Zahnärzte berücksichtigt. Die Anzahl der absoluten Häufigkeiten der Variablen der Notengebung liefert keinen eindeutigen Hinweis einer Benotungstendenz. Entgegen der in der Literatur gemachten Beobachtung lässt sich bei der Benotung von Präparationen in dieser Studie demnach keine eindeutige Tendenz zur mildereren Beurteilung der ersten Arbeiten feststellen⁵².

Anders als bei der Untersuchung von AL AMRI et al. (2016) kann darüber hinaus in der vorliegenden Studie kein Zusammenhang zwischen dem Lebensalter eines Beurteilers und der Kompensationsfähigkeit einer Ermüdung in Kombination mit der Tendenz zur Vergabe guter Noten hergestellt werden, da sich bei „A“ als jüngstem Beurteiler keine eindeutige Benotungstendenz ausmachen lässt und der Vergleich zu den anderen Beurteilern einen mittleren Reliabilitätskoeffizienten aufweist¹⁴. In den Ergebnissen der Fragebögen der approbierten Zahnärzte werden mit großer Mehrheit Ermüdung, Stress und „Betriebsblindheit“ respektive zu viele Beurteilungen in zu kurzer Zeit als die am stärksten beeinflussenden Faktoren für eine objektive Beurteilung genannt. Als weitere mögliche Ursache für unterschiedliche Reliabilitätskoeffizienten können Reihungs- beziehungsweise Schwankungseffekte in Betracht gezogen werden.

6.4.7. Beurteilungsmuster

Die Ergebnisse zur Reliabilität der Notengebung zeigen, dass die Beurteiler durch das Auftreten von Beurteilungsmustern auffallen. Von den bereits dargestellten Fehlern lassen sich jedoch nur die Folgenden anhand der absoluten Anzahl an Notenuurteilen ableiten.

- Mildeeffekt/Strengeneffekt
- Zentrale Tendenz
- Dichotomisierungstendenz

Die Note *verbesserungswürdig* etabliert sich in fast allen Durchgängen als die Notenkategorie mit den meisten Nennungen (einschließlich bei „M“). „K“ und „B“ weisen darüber hinaus jedoch eine ausgeprägte zentrale Tendenz mit einem Notenschwerpunkt auf den mittleren Urteilen *akzeptabel* und *verbesserungswürdig* auf. Nach BOHL (2006) handelt es sich dabei um einen Beurteilungsbereich, der sicher und wenig anfechtbar ist und vor allem für unerfahrene Beurteiler attraktiv ist⁶⁹. Diese Beobachtung kann in der vorliegenden Arbeit jedoch nicht eindeutig nachgewiesen werden, da es sich bei „B“ und „K“ um Beurteiler mittlerer Berufserfahrung handelt. Im Gegensatz dazu zeigt „L“ eine deutliche Häufung schlechter Noten unter fast vollständigem Ausschluss *sehr guter* Noten. Als mögliche Gründe nennt die Literatur die Herausstellung einer hohen Fachkompetenz, eine Überforderung in Verbindung mit einer unsicheren Persönlichkeit des Beurteilers oder eine generelle Tendenz zu strengerer Benotung^{51, 52, 64}. Bei den Beurteilern „P“ und „A“ lassen sich keine eindeutigen Beurteilungsmuster auffinden. Vielmehr erscheint insgesamt die Notenvergabe homogener. Dieses homogene Bild entspricht am ehesten dem Beurteilungsmuster der Analysesoftware.

Als konkrete Gegenmaßnahme zur Verringerung von Milde- beziehungsweise Strengeneffekten und der Zentralen Tendenz nennt die Literatur die Offenlegung einer Beurteilungsneigung infolge einer kritischen Selbstreflexion, welche nachfolgend durch eine bewusste Gegensteuerung kompensiert werden kann⁵¹. Zusätzlich sollte vor der Vergabe sehr schlechter oder sehr guter Noten eine Hinterfragung der Berechtigung stattfinden⁵¹. Eine strikte Trennung nach Stärken und Schwächen einer Leistung ist ebenfalls eine sinnvolle Strategie zur Verringerung beeinflussender Tendenzen⁵¹. Bei der Betreuung unterschiedlicher Kurse mit unterschiedlichen Anforderungen sollte darüber hinaus eine Überprüfung des Beurteilungsmaßstabes hinsichtlich der Angemessenheit erfolgen⁵¹. Zusätzlich kann eine schriftliche Erinnerungshilfe während des Benotungsvorganges den Beurteiler intermittierend an die vorhandene Beurteilungstendenz erinnern und ein Gegensteuern initiieren⁵². Eine Dichotomisierungstendenz kann bei keinem der zahnärztlichen Beurteiler festgestellt werden.

6.5. Diskussion der Fragebögen der Studenten

Auf Basis der genannten Studien soll die Auswertung der Fragebögen in der vorliegenden Arbeit weiterführende Eindrücke der Studenten aufführen.

Wenn man davon ausgeht, dass die Daten der Umfragen der Studenten repräsentativ sind, dann lassen die folgenden Ergebnisse den jeweils dargestellten Schluss zu.

Allgemeiner Zustand des Betreuungsverhältnisses:

Die allgemeinen Rahmenbedingungen im Kurs und bei Präparationstestaten schätzen die Studenten mehrheitlich positiv ein. Einzig bei der Bewertung der Wartezeit zeigt sich ein anderes Ergebnis. Feststellen lässt sich jedoch, dass eine Aufstockung der Betreuung – beispielsweise durch den Einsatz computerbasierter Beurteilungstechniken – zu Stoßzeiten die Wartezeiten verkürzen und die Effektivität der Betreuung verbessern kann.

Einstellungen, Erwartungen und Wünsche der Studenten:

Wie schon einige Autoren fordern alle Studenten der vorliegenden Befragung Objektivität bei der Benotung^{7, 10, 13}. Gleichzeitig wünschen sich die Studenten objektivere Verhältnisse und dadurch einen positiven Effekt auf das Arbeitsklima. Der Einsatz einer Analysesoftware zur Benotung von praktischen Arbeiten ist ein möglicher Lösungsansatz zur Objektivierung der Beurteilungsverhältnisse. Zusätzlich wäre es sinnvoll, Demonstrationen und Gespräche zwischen Beurteilern und Studenten bezüglich der Notenfindung einzurichten, um den Beurteilungsprozess für die Studenten transparenter und nachvollziehbarer zu gestalten.

Über 80% der Studenten geben darüber hinaus an, dass die Noten entscheidenden Einfluss auf ihre Motivation haben. Fast 60% der Befragten verneinen jedoch die Aussage, dass die Notenvergabe gerecht und nachvollziehbar ist; beide Aussagen sind mit den Aussagen in der Literatur vergleichbar¹¹. Die Voraussetzung, dass die Notenvergabe einen entscheidenden Einfluss auf die Motivation der Studenten hat, diese aber als nicht nachvollziehbar und ungerecht empfunden wird, lässt nach SHARAF et al. (2006) Rückschlüsse auf erhöhte Stresslevel der Studenten zu¹³. Nach KORNADT (1975) führt Stress zu einem signifikanten Sinken der Motivation⁵⁰. Allerdings ziehen weniger als 20% der befragten Studenten ein rechtliches Vorgehen bei mehrmaligem Nichtbestehen in Erwägung. Vergleichbare Literatur dazu ist nicht bekannt.

Unterschiede in der Bewertungsqualität:

Die Ergebnisse zeigen deutlich, dass sich die Studenten mit einer variablen Beurteilungsqualität konfrontiert sehen, welche – auf die Beurteilung desselben Werkstücks/Arbeit bezogen – zu gegensätzlichen Aussagen der zahnärztlichen Beurteiler führt (Interrater-Beurteilungsqualität).

Die Intrarater-Beurteilungsqualität stellen allerdings nur die Hälfte der Befragten infrage. Es bleibt durch die gewählte Methodik offen, weshalb das Ergebnis nicht so eindeutig ausgefallen ist wie bei der Auswertung der Interrater-Beurteilerqualität. Vorstellbar ist, dass zum einen nicht alle Studenten mit dieser Situation konfrontiert waren respektive nicht innerhalb eines Tages mehrmals vom gleichen wissenschaftlichen Mitarbeiter betreut worden sind. Zum anderen bleibt zu klären, ob es sich dabei um subjektiv empfundene Qualitätsunterschiede handelt.

Es lässt sich zusammenfassend feststellen, dass die Studenten sowohl von einer mangelhaften Interrater- als auch Intrarater-Beurteilungsqualität ausgehen. An dieser Stelle soll ausdrücklich betont werden, dass es sich bei der Auswertung der Fragebögen um eine vermutete Subjektivität der wissenschaftlichen Mitarbeiter handelt. Verzerrungen durch subjektive Einflusstendenzen der Studenten sind deshalb nicht auszuschließen.

Persönliche Faktoren:

Bei den Ergebnissen hinsichtlich wahrgenommener Bevorzugung beziehungsweise Benachteiligung ist fraglich, ob es sich primär um eigene oder um beobachtete Erfahrungen handelt. Die Mehrheit der Studenten stimmt der Aussage zu, sie haben den Eindruck, der wissenschaftliche Mitarbeiter habe sich ein Bild von ihnen gemacht. Eine Aussage, die auf die Existenz von Referenz- und logischen Fehlern hindeutet. Solche Erwartungshaltungen können nach MIETZEL (1982) im Sinne einer Sich-selbst-erfüllenden-Prophezeiung in den Pygmalion-

Effekt münden⁵⁸. Ein alarmierendes Ergebnis liefert die Auswertung der Aussage, ob die Kritik manchmal herablassend ist. Dies legt den Vergleich mit der Aussage von WINTER (2016) nahe, dass Beurteiler ihre Machtposition missbrauchen können⁴⁹. Entsprechend dazu gibt eine 2/3-Mehrheit der Studenten an, sich unwohl beim Vorzeigen einer Arbeit zu fühlen. Es ist demnach also wahrscheinlich, dass nicht nur einzelne Studenten von einer ungeeigneten Umgangsweise betroffen sind. Weitere Untersuchungen zur Identifizierung subjektiver Einflusstendenzen bei der Bewertung des Arbeitsklimas sollten durchgeführt werden.

Der Prozess des Lernens oder die Entwicklung feinmotorischer Fähigkeiten kommt in so einem Arbeitsumfeld jedoch zu kurz. Bereits mehrere Studien beschreiben den negativen Einfluss eines solchen Arbeitsumfeldes auf die Studenten^{10, 13}. Als möglicher Lösungsansatz – zur Einhaltung einer Gleichbehandlung aller Studenten – ist die Einrichtung einer Beschwerdestelle mittels Beschwerdebeauftragtem für betroffene Studenten denkbar. Bei dem Beschwerdebeauftragten sollte es sich um eine kursexterne Person handeln, bei der Beschwerden ausschließlich persönlich vorgetragen werden können. Auf diese Weise wird einem möglichen durch Anonymität bedingten negativen Verzerrungseffekt entgegengewirkt.

Umgang mit prepCheck 1.1:

Die Mehrheit der Befragten stuft Computerurteile nicht generell als vertrauenswürdiger im Vergleich zu zahnärztlichen Beurteilungen ein. Eine gewisse Skepsis gegenüber technischen Beurteilern ist erkennbar, obwohl die meisten Studenten computerbasierte Urteile als gerecht empfinden. Dies basiert vermutlich auf der von den Befragten angenommenen Objektivität digitaler Systeme. Diese Einstellung gegenüber computerbasierten Auswertungssystemen wird in der Literatur empfohlen⁴⁶. Während der Betreuung im Laboralltag bevorzugen knapp 60% der Studenten einen wissenschaftlichen Mitarbeiter. Dies liegt vermutlich auch an der noch langsamen Auswertung von bis zu fünf Minuten für eine Präparation, welche bei Verwendung von prepCheck 1.1 im Laboralltag, bei der ohnehin Wartezeiten entstehen, zu lang ist. Auch die individuelle Betreuung und Hilfestellung durch wissenschaftliche Mitarbeiter kann durch prepCheck 1.1 nicht adäquat ersetzt werden. Demnach bietet sich der Einsatz von prepCheck 1.1 zurzeit nur als zusätzliche Maßnahme zur Unterstützung der wissenschaftlichen Mitarbeiter im Laboralltag an. Bei Präparationstestaten sprechen sich jedoch mehr als 75% für die Bewertung durch einen Computer aus. Hier eignet sich die objektive und als gerecht empfundene Bewertungsweise der Analysesoftware optimal zur Verbesserung der vermuteten willkürlichen Benotung durch wissenschaftliche Mitarbeiter. Der Einsatz von digitalen Lehr- und Bewertungssystemen als mögliche Lösung für die defizitäre derzeit bestehende Situation setzt

allerdings eine Kooperationsbereitschaft der Studenten mit solchen Technologien voraus. Die erkennbare Akzeptanz der Studenten der Umfrage gegenüber dem Einsatz computerbasierter Ausbildungssysteme entspricht den Ergebnissen der aktuellen Studienlage^{34, 37-43}.

In Anbetracht einer fortschreitenden Digitalisierung der Umwelt bieten neue leistungsfähige Technologien sinnvolle Alternativen, welche auch zunehmend in der zahnärztlichen Ausbildung Anwendung finden können und sollen.

6.6. Diskussion der Fragebögen der Zahnärzte

Die Auswertung der Fragebögen in der vorliegenden Arbeit soll die Sicht der betreuenden Zahnärzte darstellen. Die Unkenntnis der befragten Zahnärzte hinsichtlich der dargestellten beeinflussenden Faktoren auf eine objektive Beurteilung wurde genutzt, um anhand gegebener Antworten mögliche Benotungsfehler ableiten zu können. Es ist jedoch darauf hinzuweisen, dass aufgrund von ablehnenden Antworten nicht von einem Fehlen des jeweils untersuchten Beurteilungsfehlers ausgegangen werden kann, da Beobachtungs- und Beurteilungsfehler unbewusst begangen werden. Eine positive Antwort auf gestellte Fragen liefert dagegen eindeutige Hinweise auf die Existenz eines vorliegenden Fehlers, da dieser den Befragten bereits bewusst aufgefallen ist.

Wenn man davon ausgeht, dass die Daten der Umfragen der Zahnärzte repräsentativ sind, dann lassen die folgenden Ergebnisse den jeweils dargestellten Schluss zu.

Allgemeiner Zustand des Betreuungsverhältnisses:

Anders als bei den Ergebnissen der Studenten weisen die Resultate darauf hin, dass es unter den Beurteilern deutliche Differenzierungen zur Einschätzung des gegenwärtigen Betreuungsschlüssels gibt. Diesen stuft über die Hälfte der Befragten als nicht ausreichend ein. KLEBER (1992) stellt fest, dass wachsende Gruppengrößen bei den Betreuern eine steigende Verunsicherung auslösen, welche zu einem autoritären Stil mit stereotypischen – verstärkt durch subjektive Tendenzen beeinflussten – Beurteilungen führt⁵². Es lassen sich demnach ein erhöhtes Stresslevel und eine Überforderung bei einem Großteil der Befragten ableiten. Dies ist die Basis für Beeinflussungstendenzen in den Beurteilungen wie Assimilation, Selektion oder Akzentuierung⁵². Die Ergebnisse hinsichtlich der empfundenen Stressbelastung in der studentischen Kursbetreuung führen zu dem Schluss, dass es unter den Befragten einzelne Personen gibt, welche Stress besser kompensieren können beziehungsweise diesen als nicht belastend wahrnehmen. Es scheint sich demnach um eine binäre Situation unter den Befragten hinsichtlich der Belastung zu handeln. Um diese zu optimieren, sollte auf einen von der

Gesamtheit als ausreichend wahrgenommenen Betreuungsschlüssel geachtet werden und diese darüber hinaus aufgefordert werden, bei Anzeichen von Überforderung kompensatorische Maßnahmen zu ergreifen (Pausenphasen oder personelle Aufstockung). Des Weiteren stellt der Einsatz computerbasierter Beurteilungseinheiten einen möglichen Lösungsansatz zur Entlastung der wissenschaftlichen Mitarbeiter dar.

Ein Teil der Befragten stellt außerdem fest, dass diese selbst dem Anspruch einer objektiven Beurteilung nicht immer gerecht werden können. Dieses Bewusstsein ist nach der Literatur die Voraussetzung zur Vermeidung von Beurteilungsfehlern^{51, 52}. Dies deutet auch darauf hin, dass die Befragten grundsätzlich gewillt sind, eine dargebotene Leistung möglichst neutral und objektiv zu betrachten. Gezielte Schulungen mit allen Beurteilern sollten eingerichtet werden, um ein umfassendes Wissen über die defizitäre Beurteilungslage zu schaffen, damit diese anschließend optimiert werden kann. Die gegenwärtig vorhandene Betreuungszeit bietet zwar die Möglichkeit für die Erteilung kurzer mündlicher Verbesserungsvorschläge; eine intensive individuelle Betreuung ist wie bereits beschrieben allerdings nicht zu leisten. Beim Erlernen komplexer praktischer Tätigkeiten wie beispielsweise der Präparation eines Zahnes kann jedoch angenommen werden, dass dieser Prozess von einer intensiveren und individuelleren Betreuung profitieren würde. Zum Zeitpunkt der Untersuchung schulten die Studenten ihre praktischen Fertigkeiten größtenteils selbstständig und erhielten Rückmeldungen über ihren Leistungsstand allein durch das Vorzeigen ihrer Arbeiten. Zum Verständnis der mündlich erteilten Rückmeldungen mangelt es vielen Studenten jedoch gerade am Anfang ihrer Ausbildung noch an Fachwissen in Bezug auf Verfahrenstechniken und die verwendbaren Materialien. Die daraus resultierenden Kommunikationsprobleme und Missverständnisse können dazu führen, dass die Ratschläge falsch oder gar nicht verstanden werden und die Studenten infolgedessen wiederholt die wissenschaftlichen Mitarbeiter konsultieren müssen. Dieser zeitliche Mehraufwand könnte reduziert werden, wenn kleinere Gruppen von Studenten eine intensivere Betreuung durch Tutoren (Kommilitonen aus höheren Semestern) erhalten, welche die Verbesserungsvorschläge direkt während der praktischen Tätigkeit einbringen können. Auch der Einsatz von CAL-Systemen zur Erlernung praktischer Fähigkeiten wäre denkbar. Ähnlich wie die Studenten empfinden auch die Befragten das verwendete Notensystem als sinnvoll; jedoch bietet die Einführung eines veränderten Notensystems eine Möglichkeit, alte Beurteilungsstrukturen aufzubrechen, eine einheitliche Kalibrierung der Mitarbeiter zu integrieren und die Verbindung zum Schulsystem mit den damit einhergehenden Assoziationen zu unterbinden. Dieser Kalibrierungsprozess sollte nach jedem Wechsel der kursbetreuenden wissenschaftlichen Mitarbeiter stattfinden.

Objektivität:

Die Annahme einer Steigerung der Objektivität durch eine intersubjektive Übereinstimmung ist laut der Literatur kritisch zu betrachten und sollte hinterfragt werden, da die Basis einer gemeinsamen Fehleinschätzung ebenfalls zu einer Übereinstimmung unter den Beurteilern führt^{52, 70, 85}. Zusätzlich dazu geben alle Befragten an, sich nicht von kollegialen Ratschlägen beeinflussen zu lassen. Diese Methode bietet jedoch nach der Literatur die Möglichkeit zur Verbesserung der Beurteilungsqualität⁸⁶. Voraussetzung ist jedoch, dass es nicht als Schwäche ausgelegt wird, wenn Mitarbeiter bei unsicherer Notengabe Hilfe im Kollegium suchen⁵¹. Eine regelmäßige Weiterbildung der in der Benotung tätigen Mitarbeiter bezüglich der Vorteile eines interkollegialen Austausches, aber auch des Risikos eines kollektiven Irrtums stellt eine empfehlenswerte Maßnahme zur Optimierung der Beurteilungsleistung dar. Offensichtlich sind Missstände in der Beurteilungsqualität bereits bekannt, da die Befragten zu der übereinstimmenden Aussage kommen, dass es Qualitätsunterschiede zwischen den Beurteilern hinsichtlich der Beurteilungsleistung gibt. Dies ist laut Literatur die primäre Voraussetzung zur Verbesserung der defizitären Beurteilungssituation⁵¹. Allerdings sollte sich jeder Beurteiler auch über eigene Beeinflussungstendenzen im Klaren sein und diese nicht unhinterfragt lassen⁵¹.

Anonymität:

In Anonymisierungsprozessen sieht eine Mehrheit der Befragten eine Steigerungsmöglichkeit des Beurteilungsprozesses in Bezug auf die Objektivität. Diese Annahme entspricht den Aussagen aus der pädagogischen Literatur^{52, 64}. Jedoch ist eine vollkommene Anonymität nicht gegeben, da sich auffällige Studenten bereits im Vorfeld während der Betreuung im Kurs mit ihrer Platznummer bei fast allen Befragten einprägen und dies infolgedessen zumindest teilweise zu von den von KLEBER (1992) beschriebenen Erwartungshaltungen und Wahrnehmungskompromissen während der Benotung führt⁵².

Wie bereits dargestellt, wurden zum Zeitpunkt der Untersuchung die Benotungen im Testatheft, welche eine namentliche Zuordnung und die Einsicht in den Notenspiegel ermöglichen, eingetragen. Eine verbesserte Anonymisierung ließe sich durch Einführung randomisiert erteilter Nummern für jedes Testat erreichen. Darüber hinaus sollten die erteilten Noten auf unbeschrifteten Listen verbindlich eingetragen und erst im Anschluss daran in den Testatheften vermerkt werden. Auf diese Weise ließen sich Fehler, die auf Vorwissen über die Studenten basieren, vermeiden. Darüber hinaus lassen sich dadurch Generalisierungsfehler und Akzentuierungen unterbinden, da ein Notenspiegel für bereits erbrachte Leistungen ebenfalls nicht vorliegt.

Beeinflussende Faktoren I:

Die Ergebnisse führen zu der Schlussfolgerung, dass auf Kausalattributionen basierende Annahmen (Anfängerfehler) gegenüber den Studenten vorliegen. Demnach kann angenommen werden, dass alle Beurteilungen unter Einfluss eines Referenzfehlers oder logischen Fehlers stattfinden, da die Befragten allen Studenten im Sinne eines Vorurteils Anfängerfehler unterstellen. Auch das von den Befragten angegebene Einschätzen der Studenten nach einer gewissen Zeitspanne lässt auf die Existenz von Interaktionsfehler und Haloeffekt schließen. Nach KLEBER (1992) beeinflussen Kenntnisse über die zu beurteilende Person stets die Leistungsbeurteilung⁵². Einem Vorliegen der implizierten Persönlichkeitstheorie (naiv-psychologische Eigenschaftstheorie) sind sich zumindest zwei der Befragten bewusst, da sie die Aussage, dass es mit Hilfe des ersten Eindrucks möglich ist, auf ein Zurechtkommen eines Studenten zu schließen, positiv beantworten. Die Aussage, dass ein sauberer Arbeitsplatz mit einer sauberen Arbeitsweise einhergeht, lässt auf eine weitere implizierte Persönlichkeitstheorie (naiv-psychologische Prozesstheorie) beziehungsweise auf den logischen Fehler schließen. Ein Befragter stimmt darüber hinaus der Aussage voll zu, dass Sympathie seine Benotung positiv beeinflusst, so dass sich hier ein auf dem Interaktionsfehler/Haloeffekt beruhender Mildefehler identifizieren lässt. Nach KLEBER (1992) bedarf die Widerlegung sehr starker Hypothesen, welche im Vorfeld über den Studenten aufgestellt worden sind und sich bereits eingepreßt haben, großer Wahrnehmungsreize⁵². Dass dieser Beurteilungsfehler auch bei den befragten Zahnärzten vorliegt, zeigt sich in der großen Zustimmungsrate zu der Aussage, dass plötzlich sehr gute Leistungen eines leistungsschwachen Studenten die Beurteiler misstrauisch machen. Nach KLEBER (1992) wird die zu beurteilende Arbeit infolgedessen einer besonders intensiven Kontrolle unterzogen, um auch nur geringfügige Hinweise zur Erfüllung der vorher aufgestellten starken Hypothese zu finden⁵². Die Existenz eines Nähefehlers innerhalb derselben Leistungsbeurteilung lässt sich anhand der Ergebnisse nicht bestätigen. Einer Einflussnahme bereits erbrachter Leistungen auf die Beurteilung gegenwärtiger Leistungen stimmen drei der Befragten zu und stützen das Vorliegen eines logischen Fehlers.

Zusammenfassend lässt sich sagen, dass die Ergebnisse der vorliegenden Befragung deutliche Hinweise liefern, dass die vorklinische Beurteilung einigen – schon aus der schulischen Leistungsbeurteilung bekannten – Beobachtungs- und Beurteilungsfehlern unterliegt. Die bereits angesprochene Anonymisierung der Leistungen würde alle auf Informationen über die Studenten basierenden Beurteilungsfehler verhindern (Interaktionsfehler I, logischer Fehler, Haloeffekt, Pygmalion-Effekt). Die Literatur empfiehlt zur Verringerung solcher Effekte die Ausführung konträrer Handlungen, um dem Niveau einer objektiven Beurteilung näher zu kommen⁶⁹. Von

maßgeblicher Bedeutung ist jedoch eine intensive Schulung der Beurteiler hinsichtlich der Beobachtungs- und Beurteilungsfehler sowie deren Auswirkungen⁷¹. Zusätzlich stellt die Kalibrierung der Beurteiler in Ergänzung zur Formulierung exakter Lernziele sinnvolle Ergänzungen dar⁵¹.

Beeinflussende Faktoren II:

KLEBER (1992) listet Stress und Bedürfnisabhängigkeit der Wahrnehmung als beeinflussende Faktoren für eine Beurteilung⁵². Die Ergebnisse der Fragebögen der Zahnärzte zeigen, dass „Ermüdung“ und „Stress“ häufig genannte Faktoren sind. Der Faktor der „Betriebsblindheit“ stellt eine weitergehende Form der Ermüdung dar. Aufgrund dessen sollte zur Verringerung des Einflussgrades dieser Faktoren auf eine intermittierende Belastung der Beurteiler geachtet werden. Bemerkenswert ist zudem, dass „eigene Erfahrungswerte“ und „Sympathie/Antipathie“ häufig genannte beeinflussende Faktoren sind. Dies lässt den Schluss zu, dass den Beurteilern zumindest partiell bewusst zu sein scheint, dass ihre individuellen Beurteilungsmaßstäbe dem Anspruch einer fairen Beurteilung nicht gerecht werden. Eine ausführliche Kalibrierung der Betreuer und die Einführung eines neuen Notensystems lassen sich erneut als mögliche Lösungsansätze aufführen. Auch die Rahmenbedingungen („Ablenkungen“, „Lichtverhältnisse“ und „Lärm“) spielen bei der Beeinflussung des Benotungsprozesses eine Rolle. Hier sollte diskutiert werden, ob bauliche Veränderungen (optimierte Lichtverhältnisse und/oder Schallschutz in der Decke) in Betracht gezogen werden können.

Eigene Meinung:

Zusammenfassend sehen alle Beurteiler die vorklinische Ausbildung als einen unverzichtbaren Ausbildungsanteil an. Die korrekte Beurteilung praktischer Arbeiten, respektive von Präparationen, stellt einen der wichtigsten Bestandteile in der gesamten zahnmedizinischen Ausbildung dar. Die Bereitschaft, die vorklinische universitäre Ausbildung zu verändern, ist zumindest partiell bei den Befragten vorhanden, da diese die Situation als verbesserungswürdig ansehen. Einer Legimitation zur Herbeiführung einer stressbehafteten Ausbildungssituation im Sinne einer naiv-psychologischen Prozesstheorie (implizierte Persönlichkeitstheorie, „eine harte Ausbildung bringt gute Zahnärzte hervor“) findet nur teilweise statt. Die Ergebnisse deuten darauf hin, dass ein Teil der Befragten eine Veränderung der gegenwärtigen Situation wünscht, wobei der Fokus zum Großteil auf einer Reduktion des Stresslevels liegt. Hier setzt sich das Ergebnis der bereits festgestellten zweigeteilten Meinung unter den Befragten fort. Zur

genaueren Analyse der Beurteilungsfehler sowie der möglichen Lösungsstrategien sollten in Zukunft weitere Untersuchungen folgen.

6.7. Fehler im Umgang mit CEREC/prepCheck 1.1

6.7.1. Folgen selbst eingestellter Parameter

Die im Folgenden dargestellten Auswirkungen der fehlerhaft eingestellten Parameter Aufnahmequalität, Einschubrichtung, Okklusallinie und Präparationsrand wurden im Rahmen der Analyse mit dem Goldstandard entdeckt.

- *Aufnahmequalität (CEREC):*
Von entscheidender Bedeutung ist die Qualität der Kameraaufnahme der Präparation. Erfolgt diese unzureichend, werden fehlende Abschnitte durch fehlerhafte Segmente ersetzt. Bei der Auswertung eines unzureichend erfassten Zahnstumpfes findet sich häufig eine mangelhafte Erfassung der zirkulären Flächenanteile, die sich im weiteren Analyseverlauf als Erhebungen auf der Oberfläche darstellen. Dies beeinflusst die Analyse der Kategorien *Hinterschnitt* und *Präparationswinkel*.
- *Einschubrichtung (prepCheck 1.1):*
Bereits minimale Abweichungen der Einschubrichtung einer Außenfläche – in der Darstellung nur als vereinzelte gelbe Störstellen dargestellt – wirken sich sehr deutlich in der anschließenden Analyse des *Präparationswinkels* aus.
- *Okklusallinie:*
Eine nach zervikal verlegte Okklusallinie zeigt in der Analyse des *Präparationswinkels* eine nach zervikal verlegte und verschmälerte Analysefläche, welche die Ergebnisse der Kategorie *Präparationswinkel* deutlich beeinflusst.
- *Präparationsrand:*
Nicht korrekt festgelegte Verläufe des Präparationsrandes führen zu Beeinflussungen der Kategorien *Randabschlussstyp*, *Präparationsrand* und *Hinterschnitt*.

6.7.2. Weitere Limitationen der computergestützten Beurteilung

Eingeschränkte Rechenleistung:

Die Zeitspanne vom Aufnahmevorgang im CEREC-Programm bis zur Präsentation der Ergebnisse mit prepCheck 1.1 beträgt unter der im Versuchsaufbau vorhandenen Rechnerkonfiguration für einen präparierten Zahn etwa fünf Minuten.

Kameraqualität:

Bei nur geringfügig gelöstem Approximalkontakt zwischen der Präparation und dem Nachbarzahn kann die Omnicam trotz ausführlicher Scanvorgänge die Separation nicht als solche identifizieren. Im weiteren Verlauf werden die betreffenden Bereiche falsch dargestellt und erscheinen nicht separiert, sondern zusammenhängend (siehe Abbildung 22). Dies beeinflusst maßgeblich die Analyseergebnisse der Parameter *Präparationsrand* und *Randabschlusstyp*.

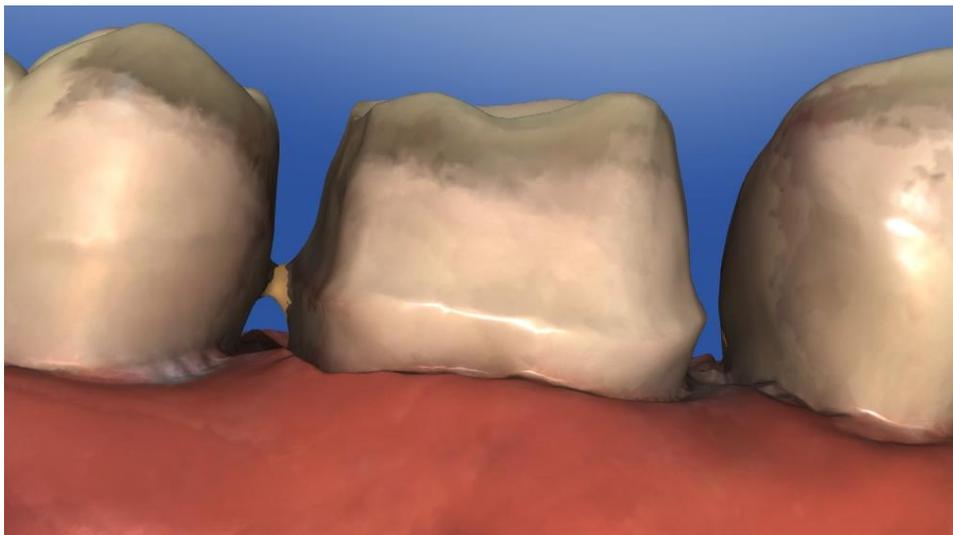


Abbildung 22: Falsch dargestellter zervikaler Interdentalraum

Sichtkontrolle am Kamerakopf:

Es stellt sich heraus, dass ein Führen der Omnicam unter gleichzeitig geforderter Sichtkontrolle am Monitor – trotz akustischer Rückmeldefunktion – als schwierig einzustufen ist.

Zirkulärer und okklusaler Abtrag:

Bei der Einstellung des Parameters *Substanzabtrag* im Konfigurationsprogramm prepCheck 1.1-Wizard liegt keine Unterscheidung zwischen okklusaler und zirkulärer Fläche vor. In Anbetracht der Tatsache, dass eine Vollgusskrone einen zirkulären Abtrag von etwa 0,8 mm, okklusal aber eine Reduktion von 1 mm benötigt, wird deutlich, dass die Spanne der Toleranzwerte entsprechend insgesamt größer gewählt werden muss, um für beide Flächen die zu erreichenden Toleranzwerte zu erhalten. Die zirkulären Flächen weisen daraufhin höhere Toleranzspannen auf als die okklusale Fläche. Fehler im *Substanzabtrag* der zirkulären Reduktion werden entsprechend weniger sensitiv erfasst.

Analyse des Präparationswinkels am Höcker/Einstellungen für Ober- und Unterkiefer:

Im Molarenbereich wird durch prepCheck 1.1 die Höckerabschrägung stets in Bezug auf die vestibuläre Fläche gemessen. Der palatinale Höcker im Oberkiefer ist jedoch im Vergleich zum Vestibulären deutlicher abschrägt. Darüber hinaus finden sich trotz der Auswahl des Zahnes 26 im prepCheck 1.1-Wizard-Programm die gleichen voreingestellten Werte für die *Präparationswinkel* der Höckerabschrägung wie für Unterkiefermolaren. Die vestibuläre Höckerabschrägung im Unterkiefer – auf die sich die voreingestellten Werte beziehen – ist jedoch deutlich stärker ausgeprägt als im Oberkiefer.

Unzureichende Abschrägung:

Eine unzureichend ausgeprägte Abschrägung der Höcker führt partiell zu einer fehlerhaften Analyse der zirkulären Außenfläche, da die Analyse der Höckerabschrägung ausbleibt.

Versatz:

Bei einem Mangel an Referenzpunkten – beispielweise bei ausgeprägter Schädigung der Nachbarzähne sowie nicht ausreichender Erfassung umgebender Strukturen – entstehen beim Vergleich von übereinandergelegter Präparation mit dem Goldstandard sichtbare Versatze. Um diesen Effekt mittels einer Abbildung verdeutlichen zu können, wurde der Goldstandard mit beschädigten Nachbarzähnen erfasst und mit dem Goldstandard und intakten Nachbarzähnen verglichen (siehe Abbildung 23 und 24). Es konnten Versatze bis zu 0,8 mm festgestellt werden. Dies führt zu erheblichen Beeinflussungen in der Kategorie des *Substanzabtrages*: sichtbar wird dies in Abbildung 23 mit 20% rot markierten und als zu klein identifizierten Flächen; dabei ist zu berücksichtigen, dass es sich um denselben Goldstandard handelt.

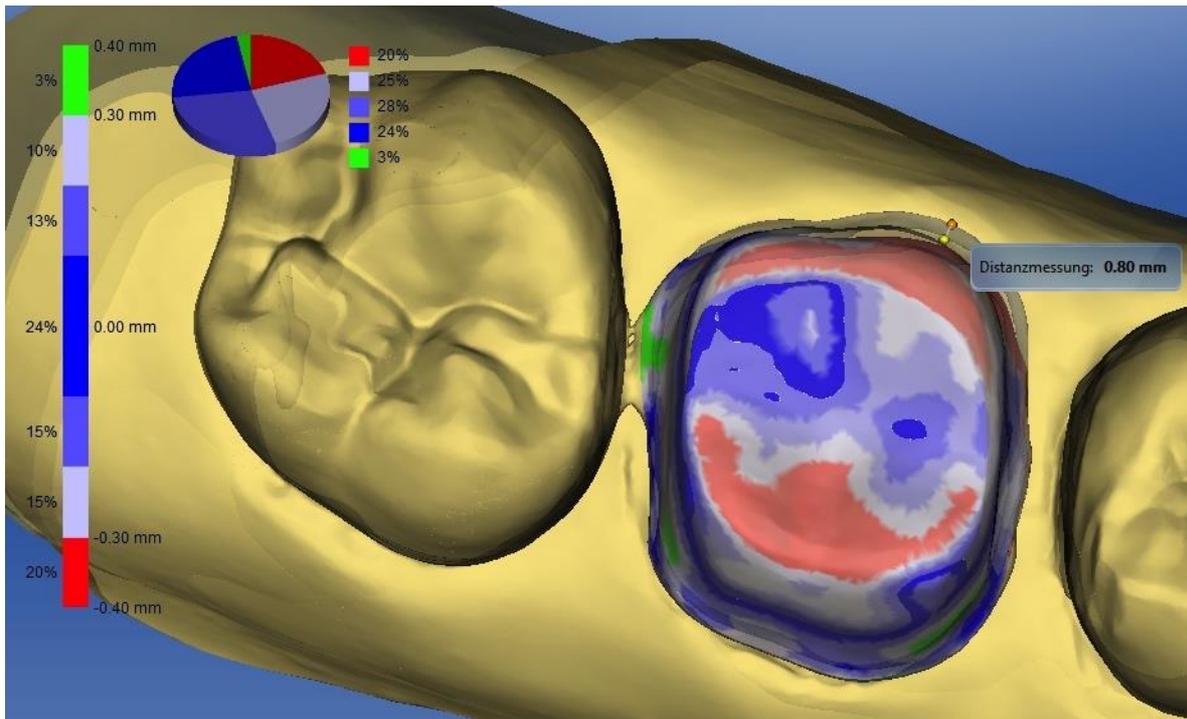


Abbildung 23: Versatz zum Goldstandard mit Distanzmessung

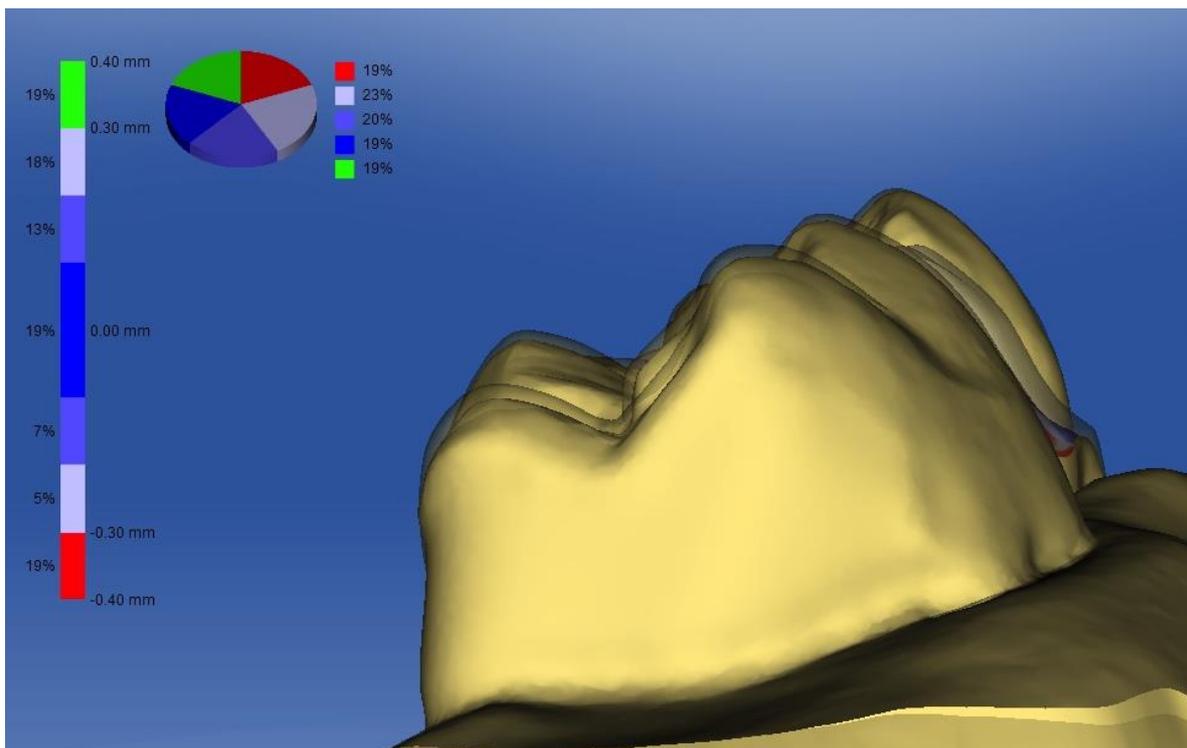


Abbildung 24: Versatz zum Goldstandard

Nachbarzahnsschädigung:

Eine Nachbarzahnsschädigung wird durch prepCheck 1.1 weder als solche erfasst noch gewertet. Lediglich ein zu geringer approximaler *Substanzabtrag* im Vergleich mit dem Goldstandard ermöglicht einen Hinweis auf eine vorliegende Nachbarzahnsschädigung, wobei dies als ein noch

zu korrigierender grün dargestellter Fehler dargestellt wird. Dies stellt einen entscheidenden Nachteil der Analysesoftware dar, da die Nachbarzahnschädigung ein wichtiges Kriterium bei der Beurteilung von Präparationen in Kurstestaten darstellt und später bei der Arbeit am Patienten im Sinne einer Körperverletzung zu werten ist. Die Verletzung von Nachbarzähnen mit dem daraus resultierenden erheblich erhöhten Kariesrisiko weist eine hohe Inzidenz auf^{87, 88}. Deswegen sollte das Erlernen einer verletzungsfreien vorsichtigen Separation der Zähne im Approximalbereich bereits Bestandteil von Präparationsbeurteilungen sein.

Lage der Präparationsgrenze:

Die optische Erfassung einer isogingival oder subgingival verlaufenden Präparationsgrenze ist nicht möglich, sodass eine Beurteilung solcher Präparationen mit prepCheck 1.1 nicht realisierbar ist.

6.8. Optimierungsoptionen für CEREC/prepCheck 1.1

Entsprechend der oben dargestellten Fehler sind folgende Lösungsansätze denkbar.

Leistung/Ausstattung:

Um die Wartezeit während der Analyse der Präparationen zu verkürzen, sollte die Leistung der Aufnahmeeinheit optimiert werden. Eine unangemessene Länge der Auswertungszeit kritisiert auch eine weitere Studie⁴⁴. Eine gesteigerte Leistungsfähigkeit der Omnicam hinsichtlich der visuellen Erfassung bietet sich als Lösungsansatz zur Vermeidung von falschen oder ungenauen Darstellungen an. Durch die Möglichkeit der direkten Sichtkontrolle über einen Kontrollbildschirm am Kamerakopf ließe sich ferner die Handhabung der Omnicam anwenderfreundlicher gestalten. Darüber hinaus sollte der Einfluss des Anwenders auf die selbst eingestellten Parameter reduziert werden.

Einstellungen der Konfigurationen:

Eine Differenzierung der Flächen (zirkulär/vestibulär/okklusal) bei der Analyse des *Substanzabtrages* würde dem Anwender eine erweiterte Analyse mit individuell einstellbaren Toleranzspannen ermöglichen. Eine Einstellungsfunktion, welche Höckerabschrägung (vestibulär oder oral) eines Molars gemessen werden soll, sowie die Einführung angepasster *Präparationswinkel* für die Höckerabschrägungen im Oberkiefer erscheinen sinnvolle Optimierungsmaßnahmen, um auch geeignete Analysen der Oberkiefermolaren zu erhalten. Bei einer fehlerhaften Analyse einer unzureichenden Höckerabschrägung stellt eine entsprechende Festlegung der zu analysierenden Bereiche – die der Beurteilungssoftware als Referenz dienen

und nicht umgangen werden können – einen möglichen Lösungsansatz dar.

Anwendung im Analyseprogramm:

Sinnvoll wäre eine Funktion, bei welcher der Versatz im Vergleich mit dem Goldstandard erkannt wird und dies dem Anwender mitgeteilt wird, sodass dieser die Analyseergebnisse der Kategorie *Substanzabtrag* kritisch hinterfragt und eine erneute Erfassung des zu beurteilenden Zahnes vornehmen kann. Eine Erfassung der Nachbarzähne inklusive der Schädigungsgrade sollte als selbstständiger Bewertungsparameter mit aufgenommen werden, um eine frühzeitige Vermittlung dieser Problematik bei den Studenten zu gewährleisten.

Darüber hinaus erscheinen folgende zwei Erweiterungsvorschläge sinnvoll:

Prozentuales Vermessen von Flächenanteilen:

Um die Parameter *Oberflächenbeschaffenheit* und *Präparationsrand* reliabler messen zu können, stellt eine prozentuale Messung vorhandener Störstellen – wie sie in der Kategorie *Hinterschnitt* zur Anwendung kommt – eine empfehlenswerte Ergänzungsmaßnahme dar.

Farbliche Angleichung der Abstufungen:

Eine farbliche Angleichung der verwendeten Blauabstufungen „Mittelblau“ und „Hellblau“ an die betreffende Richtung (grüner oder roter Bereich) ist zu empfehlen. Darüber hinaus sollte auch eine Differenzierung der verschiedenen Toleranzwerte im Kreisdiagramm stattfinden, da dem Anwender auf diese Weise der ermittelte Bereich der Toleranzskala gezeigt wird und ihm dies Hinweise liefert, ob er eher zu einem zu geringen oder einem erhöhten Abtrag neigt. Auf diese Weise können Studenten auch zu bereits guten Arbeiten weitere Verbesserungsvorschläge erhalten.

6.8.1. Analyse der Hypothesen

Bezüglich der anfangs aufgestellten Hypothesen lässt sich abschließend Folgendes feststellen:

1. Hypothese 1 hat sich bestätigt: die Reliabilität der zahnärztlichen Beurteiler (Intrarater-Reliabilität) ist signifikant niedriger als die der prepCheck 1.1 Analysesoftware.
2. Hypothese 2 hat sich bestätigt: Die Reliabilität aller zahnärztlichen Beurteiler (Interrater-Reliabilität) weist einen Reliabilitätskoeffizienten im unteren Bereich auf.
3. Hypothese 3 hat sich nicht bestätigt: Die Reliabilität der zahnärztlichen Beurteiler ist individuell verschieden und weist erhebliche Unterschiede in den Reliabilitätskoeffizienten auf.
4. Hypothese 4 hat sich nicht bestätigt: Es konnte kein eindeutiger Zusammenhang zwischen dem Grad der Berufserfahrung und der Reliabilität nachgewiesen werden.
5. Hypothese 5 hat sich unter Hinzunahme einer Ergänzung bestätigt: Die unterschiedlichen Beurteilungskategorien weisen ähnliche Reliabilitätskoeffizienten auf und unterliegen größtenteils nur geringfügigen Schwankungen. Bei allen Beurteilern – ausgenommen „L“ – setzt sich jedoch eine Kategorie als deutlich beste Kategorie ab.
6. Hypothese 6 hat sich mit einer Einschränkung bestätigt: Die Studenten bewerten die gegenwärtige Situation als verbesserungswürdig; allerdings bewerten sie den allgemeinen Zustand des Betreuungsverhältnisses positiv.
7. Hypothese 7 hat sich ebenfalls nur zum Teil bestätigt: Nur ein Teil der Zahnärzte bewertet die gegenwärtige Situation als defizitär.
8. Hypothese 8 hat sich bestätigt: Die Ergebnisse deuten darauf hin, dass es auch unter den zahnärztlichen Beurteilern zu den gleichen Beobachtungs- und Beurteilungsfehlern wie in der schulischen Leistungsbeurteilung kommt.
9. Hypothese 9 hat sich nur teilweise bestätigt: Einige Parameter der Analysesoftware müssen vom Anwender eingestellt werden. Zudem unterliegt die Beurteilung zum Teil – mangels geeigneter Skalen – der subjektiven Beurteilung durch den Anwender.

6.9. Schlussfolgerung

Zusammenfassend lässt sich feststellen, dass die Reliabilität der zahnärztlichen Beurteiler in der vorliegenden Arbeit deutlich geringer als die Beurteilungsqualität der Analysesoftware prepCheck 1.1 ist. Festzuhalten bleibt zudem, dass die Reliabilität innerhalb der zahnärztlichen Beurteiler erheblichen Qualitätsschwankungen unterliegt, sodass dies die Schlussfolgerung zulässt, dass eine erteilte Note in einem größeren Ausmaß von der Beurteilungsqualität als von der Qualität der erbrachten Leistung abhängt.

Die Auswertungen der Fragebögen der vorliegenden Arbeit liefern Hinweise, dass sowohl die Studenten als auch die an der Ausbildung beteiligten Zahnärzte die bestehende Situation zum Teil als Belastung empfinden. Mehrere Studien weisen darauf hin, dass ähnliche Rahmenbedingungen auch an anderen Universitäten für Zahnheilkunde bestehen. Respektive steht dabei das hauptsächlich auf Angestellten der jeweiligen Universitäten basierende Betreuungssystem und die damit einhergehende inkonstante und oft vermutet willkürliche Benotungsweise ebenso wie angenommene vereinzelte Diskriminierungen gegenüber Studenten im Fokus. Ziel zukünftiger Studien sollte daher die Untersuchung der bis dato lediglich auf subjektiven Einschätzungen beruhenden Ergebnisse hinsichtlich des Arbeitsklimas an vorklinischen Kursen sein.

Kalibrierungshilfen, schriftlich fixierte Leistungsanforderungen und Goldstandards stellen einfach umzusetzende Maßnahmen dar, welche jedoch die bestehende Problematik nicht gänzlich unterbinden. Aus den Ergebnissen wird deutlich, dass der Einsatz computerbasierter Beurteilungssysteme mit einigen Einschränkungen verbunden ist, welche trotzdem eine Entlastung der zahnärztlichen Betreuer ermöglichen und eine diskutabile ergänzende Maßnahme in der zahnärztlichen Ausbildung darstellen können. Jedoch wird die Betreuung durch wissenschaftliche Mitarbeiter und die damit einhergehende Möglichkeit der individuellen Ausbildung auch in Zukunft unverzichtbar sein.

Zusätzlich scheint die Orientierung an bereits erarbeiteten Grundlagen und Lösungsansätzen aus der schulischen Leistungsdiagnostik in der zahnmedizinischen Ausbildung sinnvoll.

Die Ermittlung von Beurteilungsmethoden – welche objektiver und reliabler sind, um den Lernprozess zu fördern und Unstimmigkeiten in der Notengebung effektiv zu reduzieren – sollte als Ziel zukünftiger Untersuchungen angestrebt werden.

7. LITERATURVERZEICHNIS

1. Pospiech P, Poliklinik für Zahnärztliche Prothetik, Pleicherwall 2, 97080 Würzburg, Persönliches Interview. Würzburg 24.04.2015, 12:00 Uhr.
2. Fuller JL. The effects of training and criterion models on interjudge reliability. *J Dent Educ.* 1972;36(4):19-22.
3. Jenkins SM, Dummer PM, Gilmour AS, Edmunds DH, Hicks R, Ash P. Evaluating undergraduate preclinical operative skill; use of a glance and grade marking system. *J Dent.* 1998;26(8):679-684.
4. Kateeb ET, Kamal MS, Kadamani AM, Abu Hantash RO, Abu Arqoub MM. Utilising an innovative digital software to grade pre-clinical crown preparation exercise. *Eur J Dent Educ.* 2017;21(4):220-227.
5. Kwon SR, Restrepo-Kennedy N, Dawson DV, Hernandez M, Denehy G, Blanchette D, Gratton DG, Aquilino SA, Armstrong SR. Dental anatomy grading: comparison between conventional visual and a novel digital assessment technique. *J Dent Educ.* 2014;78(12):1655-1662.
6. Lilley JD, ten Bruggen Cate HJ, Holloway PJ, Holt JK, Start KB. Reliability of practical tests in operative dentistry. *Br Dent J.* 1968;125(5):194-197.
7. Renne WG, McGill ST, Mennito AS, Wolf BJ, Marlow NM, Shaftman S, Holmes JR. E4D compare software: an alternative to faculty grading in dental education. *J Dent Educ.* 2013;77(2):168-175.
8. Salvendy G, Hinton WM, Ferguson GW, Cunningham PR. Pilot study on criteria in cavity preparation--facts or artifacts? *J Dent Educ.* 1973;37(11):27-31.
9. Taylor CL, Grey NJ, Satterthwaite JD. A comparison of grades awarded by peer assessment, faculty and a digital scanning device in a pre-clinical operative skills course. *Eur J Dent Educ.* 2013;17(1):e16-21. doi: 10.1111/j.1600-0579.2012.00752.x. PubMed PMID: 23279405
10. Brown RK. Research in the Use of a Rating Scale as a Means of Evaluating the Personalities of Senior Dental Students. *Journal of Dental Research.* 1930;10(3):271-279.
11. Licari FW, Knight GW, Guenzel PJ. Designing evaluation forms to facilitate student learning. *J Dent Educ.* 2008;72(1):48-58.
12. Myers B. Beliefs of dental faculty and students about effective clinical teaching behaviors. *J Dent Educ.* 1977;41(2):68-76.

13. Sharaf AA, AbdelAziz AM, El Meligy OA. Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. *J Dent Educ.* 2007; 71(4):540-544.
14. Al Amri MD, Sherfudhin HR, Habib SR. Effects of Evaluator's Fatigue and Level of Expertise on the Global and Analytical Evaluation of Preclinical Tooth Preparation. *J Prosthodont.* 2016. doi: 10.1111/jopr.12558. PubMed PMID: 27809403.
15. Haj-Ali R, Feil P. Rater reliability: short- and long-term effects of calibration training. *J Dent Educ.* 2006;70(4):428-433.
16. Mays KA, Crisp HA, Vos P. Utilizing CAD/CAM to Measure Total Occlusal Convergence of Preclinical Dental Students' Crown Preparations. *J Dent Educ.* 2016;80(1):100-107.
17. Patridge MI, Mast TA. Dental clinical evaluation: a review of the research. *J Dent Educ.* 1978;42(6):300-305.
18. Dhuru VB, Rypel TS, Johnston WM. Criterion-oriented grading system for preclinical operative dentistry laboratory course. *J Dent Educ.* 1978;42(9):528-531.
19. Goepferd SJ, Kerber PE. A comparison of two methods for evaluating primary class II cavity preparations. *J Dent Educ.* 1980;44(9):537-542.
20. Robertello FJ, Pink FE. The effect of a training program on the reliability of examiners evaluating amalgam restorations. *Oper Dent.* 1997;22(2):57-65.
21. Scheutzel P. Einfluss des Bewertungssystems auf Objektivität und Reliabilität der Benotung zahnmedizinischer Studentarbeiten am Phantompatienten. *GMS Z Med Ausbild.* [Internet]. 2007 [cited 2018 Jul 17];24(1): Doc67. Available from: <http://www.egms.de/static/en/journals/zma/2007-24/zma000361.shtml>
22. Schmitt L, Moltner A, Ruttermann S, Gerhardt-Szep S. 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 J Med Educ.* 2016;33(4):Doc61. doi: 10.3205/zma001060. PubMed PMID: 27579361; PubMed Central PMCID: PMC5003144
23. Scott BJ, Evans DJ, Drummond JR, Mossey PA, Stirrups DR. An investigation into the use of a structured clinical operative test for the assessment of a clinical skill. *Eur J Dent Educ.* 2001;5(1):31-37.
24. Kerschbaum T, Esser C, Winkelmann V, Mehl A, Gloger W. Computergestützte Untersuchung von Zahnpräparationen. *Dtsch Zahnarztl Z.* 2004(59):518-522.
25. Ahmed SN, Sturdevant J, Wilder R, Kowlowitz V, Boushell L. Development and Assessment of Discrimination Exercises for Faculty Calibration in Preclinical Operative Dentistry. *J Dent Educ.* 2016;80(8):994-1003.

26. Vann WF, Machen JB, Hounshell PB. Effects of criteria and checklists on reliability in preclinical evaluation. *J Dent Educ.* 1983;47(10):671-675.
27. Schiff AJ, Salvendy G, Root CM, Ferguson GW, Cunningham PR. Objective evaluation of quality in cavity preparations. *J Dent Educ.* 1975;39(2):92-96.
28. Rosenberg H, Grad HA, Matear DW. The effectiveness of computer-aided, self-instructional programs in dental education: a systematic review of the literature. *J Dent Educ.* 2003;67(5):524-532.
29. Cardoso JA, Barbosa C, Fernandes S, Silva CL, Pinho A. Reducing subjectivity in the evaluation of pre-clinical dental preparations for fixed prosthodontics using the Kavo PrepAssistant. *Eur J Dent Educ.* 2006;10(3):149-156.
30. Kournetas N, Jaeger B, Axmann D, Groten M, Lachmann S, Weber H, Geis-Gerstorfer J. Assessing the reliability of a digital preparation assistant system used in dental education. *J Dent Educ.* 2004;68(12):1228-1234.
31. El Tantawi MM, Abdelsalam MM, Mourady AM, Elrifae IM. e-Assessment in a Limited-Resources Dental School Using an Open-Source Learning Management System. *J Dent Educ.* 2015;79(5):571-583.
32. Lee C, Kobayashi H, Lee SR, Ohyama H. The Role of Digital 3D Scanned Models in Dental Students' Self-Assessments in Preclinical Operative Dentistry. *J Dent Educ.* 2018; 82(4):399-405.
33. McPherson KR, Mennito AS, Vuthiganon J, Kritzas YG, McKinney RA, Wolf BJ, Renne WG. Utilizing self-assessment software to evaluate student wax-ups in dental morphology. *J Dent Educ.* 2015;79(6):697-704.
34. Nagy ZA, Simon B, Toth Z, Vag J. Evaluating the efficiency of the Dental Teacher system as a digital preclinical teaching tool. *Eur J Dent Educ.* 2018;22(3):e619-e623. doi: 10.1111/eje.12365. PubMed PMID: 29797383.
35. Nishida M, Sohmura T, Takahashi J. Training in tooth preparation utilizing a support system. *J Oral Rehabil.* 2004;31(2):149-154.
36. Sampaio-Fernandes MA, Sampaio-Fernandes MM, Fonseca PA, Almeida PR, Reis-Campos JC, Figueiral MH. Evaluation of occlusal rest seats with 3D technology in dental education. *J Dent Educ.* 2015;79(2):166-176.
37. Gratton DG, Kwon SR, Blanchette DR, Aquilino SA. Performance of two different digital evaluation systems used for assessing pre-clinical dental students' prosthodontic technical skills. *Eur J Dent Educ.* 2017;21(4):252-260.

38. Hamil LM, Mennito AS, Renne WG, Vuthiganon J. Dental students' opinions of preparation assessment with E4D compare software versus traditional methods. *J Dent Educ.* 2014;78(10):1424-1431.
39. Park CF, Sheinbaum JM, Tamada Y, Chandiramani R, Lian L, Lee C, Da Silva J, Ishikawa-Nagai S. Dental Students' Perceptions of Digital Assessment Software for Preclinical Tooth Preparation Exercises. *J Dent Educ.* 2017;81(5):597-603.
40. Welk A, Maggio MP, Simon JF, Scarbecz M, Harrison JA, Wicks RA, Gilpatrick RO. Computer-assisted learning and simulation lab with 40 DentSim units. *Int J Comput Dent.* 2008;11(1):17-40.
41. Welk A, Splieth C, Seyer D, Rosin M, Siemer M, Meyer G. German dental faculty attitudes towards computer-assisted simulation systems correlated with personal and professional profiles. *Eur J Dent Educ.* 2006;10(2):87-95.
42. Welk A, Splieth C, Wierinck E, Gilpatrick RO, Meyer G. Computer-assisted learning and simulation systems in dentistry--a challenge to society. *Int J Comput Dent.* 2006;9(3):253-265.
43. El Tantawi MM, Saleh SM. Attitudes of dental students towards using computers in education--a mixed design study. *East Mediterr Health J.* 2008;14(3):675-685.
44. Lenherr P, Marinello CP. [prepCheck computer-supported objective evaluation of students preparation in preclinical simulation laboratory]. *Swiss Dent J.* 2014;124(10):1085-1092.
45. Arnetzl G, Dornhofer R. PREPAssistant: a system for evaluating tooth preparations. *Int J Comput Dent.* 2004;7(2):187-197.
46. Maier U. Leistungsdiagnostik in Schule und Unterricht : Schülerleistungen messen, bewerten und fördern. Bad Heilbrunn: Klinkhardt, 2015.
47. Ingenkamp K. Die Fragwürdigkeit der Zensurengebung : Texte und Untersuchungsberichte / Zusammengestellt und kommentiert von Karlheinz Ingenkamp. Weinheim: Beltz, 1971.
48. Pfistner H-J. Lehrer beurteilen Schüler : Beiträge zur Beurteilungspraxis und zur Theorie der Leistung. Bern: Lang, 1981.
49. Winter F. Leistungsbewertung : eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen. 5. überarbeitete und erweiterte Auflage. Baltmannsweiler: Schneider-Verlag Hohengehren, 2012.
50. Kornadt H-J. Lehrziele, Schulleistung und Leistungsbeurteilung : Beiträge zur Analyse ihrer theoretischen Bedeutung, ihrer schulischen Funktion und ihrer Wechselwirkungen. Düsseldorf: Pädagogischer Verlag Schwann, 1975.

51. Jürgens E, Sacher W. Leistungserziehung und Pädagogische Diagnostik in der Schule : Grundlagen und Anregungen für die Praxis. Stuttgart: Kohlhammer, 2008.
52. Kleber EW. Diagnostik in pädagogischen Handlungsfeldern : Einführung in Bewertung, Beurteilung, Diagnose und Evaluation. Weinheim: Juventa-Verlag, 1992.
53. Paradies L, Wester F, Greving J. Leistungsmessung und -bewertung. 1. Auflage. Berlin: Cornelsen Scriptor, 2005.
54. Sacher W. Leistungen entwickeln, überprüfen und beurteilen : Bewährte und neue Wege für die Primar- und Sekundarstufe. 4. überarbeitete und erweiterte Auflage. Bad Heilbrunn/Obb.: Klinkhardt, 2004.
55. Atkinson JW, McClelland DC. The projective expression of needs. II. The effect of different intensities of the hunger drive on Thematic Apperception. Journal of Experimental Psychology. 1948;38(6):643-658.
56. Graumann CF. Nicht-sinnliche Bedingungen des Wahrnehmens. In: Metzger W, editor. Handbuch der Psychologie in 12 Bänden: 1. Band Allgemeine Psychologie, I. Der Aufbau des Erkennens / 1. Halbband: Wahrnehmung und Bewußtsein. 2. Auflage. Göttingen: Verlag für Psychologie Hogrefe, 1966:1031-1096.
57. Bruner JS. Personality dynamics and the process of perceiving. In: Blake RR, Ramsey GV, eds. Perception: An approach to personality. New York, NY, US: Ronald Press Company; 1951:121-147.
58. Mietzel G. Interpretation von Leistungen : Dargestellt aus der Sicht der Attribuierungstheorie. Opladen: Leske und Budrich, 1982.
59. Hesse I, Latzko B. Diagnostik für Lehrkräfte. 3. vollständig überarbeitete und erweiterte Auflage. Opladen: Barbara Budrich, 2017.
60. Schwarzer C. Lehrerurteil und Schülerpersönlichkeit : Kognitive Stile und Sozialschicht als Einflussgrößen für die Beurteilung bei Schulbeginn. München: Kösel, 1976.
61. von Saldern M. Schulleistung in Diskussion. Baltmannsweiler: Schneider-Verlag Hohengehren, 1999.
62. Betz D. Rhythmische Schwankungen als Fehler in der Notengebung bei mündlichen Prüfungen. Psychol. in Erzieh. u. Unterr. 1974;21. Jg.(1):1-14.
63. Jürgens E, Sacher W. Leistungserziehung und Leistungsbeurteilung : schulpädagogische Grundlegung und Anregungen für die Praxis. Neuwied: Luchterhand, 2000.
64. Sacher W. Prüfen - Beurteilen - Benoten : Theoretische Grundlagen und praktische Hilfestellungen für den Primar- und Sekundarbereich. Bad Heilbrunn: Klinkhardt, 1994.

- 65.** Kleber EW, Meister H, Schwarzer C, Schwarzer R. Beurteilung und Beurteilungsprobleme : Eine Einführung in Beurteilungs- und Bewertungsfragen in der Schule. Weinheim: Beltz, 1976.
- 66.** Rosemann H. Arbeitshefte für Psychologie; Band 14 Schülerbeurteilung; Methoden der Gruppen- und Individualbeurteilung: Beobachtung, Befragung, Test, Soziometrie, Interaktionsanalyse, Aufsatz- und Dokumentenanalyse u. a.; Urteilsbildung und Urteilsfehler; Beurteilungsarten: Benotung (Aufsatz- und Diktatbeurteilung), Kurzcharakteristik, Persönlichkeitsbild; Beurteilungsanlässe: Schulfähigkeitsdiagnose, Sonderschulauslese, Übertrittsauslese. Bissendorf: Polerz, 1975.
- 67.** Faßnacht G. Systematische Verhaltensbeobachtung : Einführung in die Methodologie und Praxis. München: Ernst Reinhardt, 1979.
- 68.** Ziegenspeck JW. Handbuch Zensur und Zeugnis in der Schule : Historischer Rückblick, allgemeine Problematik, empirische Befunde und bildungspolitische Implikationen ; Ein Studien- und Arbeitsbuch. Bad Heilbrunn: Klinkhardt, 1999.
- 69.** Bohl T. Prüfen und Bewerten im Offenen Unterricht. Weinheim: Beltz, 2006.
- 70.** Helmke A. Unterrichtsqualität erfassen, bewerten, verbessern. Seelze: Kallmeyer, 2003.
- 71.** Hasemann K. Verhaltensbeobachtung und Verhaltensbeurteilung in der psychologischen Diagnostik. Göttingen: Verlag für Psychologie Hogrefe, 1964.
- 72.** <http://manuals.sirona.com/home.HomeDmsDocument.download.html?id=5591>, (letzter Zugriff: 17.07.2018, 11:39 Uhr).
- 73.** <http://manuals.sirona.com/home.HomeDmsDocument.download.html?id=4177>, (letzter Zugriff: 17.07.2018, 11:40 Uhr).
- 74.** Wirtz M, Caspar F. Beurteilerübereinstimmung und Beurteilerreliabilität : Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe, Verlag für Psychologie, 2002.
- 75.** Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13:61. doi: 10.1186/1471-2288-13-61. PubMed PMID: 23627889; PubMed Central PMCID: PMC3643869
- 76.** Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174.
- 77.** Abderhalden C, Schori E, Kersten B. Wie reliabel ist der Fragebogen zur Analyse der Pflegedokumentation (IzEP-Dok©) im Instrument zur Erfassung von Pflegesystemen

- IzEP©?. PrInterNet [Internet]. 2006 [cited 2018 Jul 17]. Available from: http://www.pflegeportal.ch/pflegeportal/pub/Schori_et_al_2006_1618_1.pdf
- 78.** Callan RS, Cooper JR, Young NB, Mollica AG, Furness AR, Looney SW. Effect of Employing Different Typodonts When Using E4D Compare for Dental Student Assessment. *J Dent Educ.* 2015;79(6):705-710.
- 79.** Callan RS, Blalock JS, Cooper JR, Coleman JF, Looney SW. Reliability of CAD CAM technology in assessing crown preparations in a preclinical dental school environment. *J Dent Educ.* 2014;78(1):40-50.
- 80.** Raab-Steiner E, Benesch M. Der Fragebogen : Von der Forschungsidee zur SPSS-Auswertung. 3. aktualisierte und überarbeitete Auflage. Wien: Facultas.wuv, 2012.
- 81.** Smock CD. The influence of stress on the perception of incongruity. *The Journal of Abnormal and Social Psychology.* 1955;50(3):354-356.
- 82.** Gansky SA, Pritchard H, Kahl E, Mendoza D, Bird W, Miller AJ, Graham D. Reliability and validity of a manual dexterity test to predict preclinical grades. *J Dent Educ.* 2004; 68(9):985-994.
- 83.** Esser C, Kerschbaum T, Winkelmann V, Krage T, Faber FJ. A comparison of the visual and technical assessment of preparations made by dental students. *Eur J Dent Educ.* 2006;10(3):157-161.
- 84.** Gratton DG, Kwon SR, Blanchette D, Aquilino SA. Impact of Digital Tooth Preparation Evaluation Technology on Preclinical Dental Students' Technical and Self-Evaluation Skills. *J Dent Educ.* 2016;80(1):91-99.
- 85.** Schröder H. Leistungsmessung und Schülerbeurteilung. 1. Auflage. Stuttgart: Klett, 1974.
- 86.** Feil PH. An analysis of the reliability of a laboratory evaluation system. *J Dent Educ.* 1982;46(8):489-494.
- 87.** Moopnar M, Faulkner KD. Accidental damage to teeth adjacent to crown-prepared abutment teeth. *Aust Dent J.* 1991;36(2):136-140.
- 88.** Strübig W, Opitz J. Präparationsdefekte an Nachbarzähnen bei Inlay- und Kronenversorgungen. *Dtsch Zahnärztl Z.* 2000(02):101-103.

8. EIDESSTATTLICHE VERSICHERUNG

„Ich, Hannah Grabowski, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Untersuchungen zur möglichen Steigerung der Beurteilungsqualität von Präparationsergebnissen mittels einer Analysesoftware im Vergleich zur klassischen Beurteilung durch wissenschaftliche Mitarbeiter in der studentischen Ausbildung“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -www.icmje.org) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o.) und werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Betreuer/in angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o.) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilerklärung an etwaigen erfolgten Publikationen
entfällt

9. LEBENSLAUF

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

10. DANKSAGUNG

Herrn Prof. Dr. Peter Pospiech danke ich für die freundliche Überlassung des Themas und die Unterstützung bei der Durchführung der Untersuchungen an der Klinik für Zahnärztliche Prothetik der Julius-Maximilians-Universität Würzburg.

Ich bedanke mich bei den Studenten sowie allen zahnärztlichen Beurteilern für die Teilnahme an dieser Studie.

Der Firma Sirona Dental Systems GmbH, Bensheim, Deutschland danke ich für die beiden zur Verfügung gestellten Beurteilungseinheiten.

Ganz herzlich danke ich darüber hinaus Carolin Malsch vom Institut für klinische Epidemiologie und Biometrie sowie der Diplomstatistikerin Dr. Silke Lange für die Unterstützung bei der statistischen Auswertung.

Herrn Thorsten Blum, Frau Anna-Christina Wieneke und Frau Barbara Fink-Abicht möchte ich sehr herzlich für ihre konstruktiven Ratschläge und die Unterstützung bei der Formatierung der Arbeit am Computer danken.

Meiner Familie, insbesondere meinen Eltern, danke ich für ihre weit über den Rahmen der Doktorarbeit hinausreichende Unterstützung und widme ihnen diese Arbeit.

11. APPENDIX

Anhang A – Fragebogen der Studenten Gruppe prepCheck

Fragebogen Gruppe prepCheck

Dieser Fragebogen dient zur Verbesserung der vorklinischen Ausbildung.
Die Auswertung wird ausschließlich von Studierenden vorgenommen und ist anonymisiert.

1. Kreuzen Sie das am besten Zutreffende an.

	stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu
○ Das Betreuungsverhältnis Assistenten zu Studierenden ist gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ein Unterschied in der Kritik / Benotung zwischen den einzelnen Assistenten ist vorhanden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Assistenten wirken manchmal überfordert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik der Assistenten ist nachvollziehbar / gut erklärt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Aussagen und Verbesserungsvorschläge der Assistenten verwirren mich manchmal.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Objektivität ist mir bei Benotungen sehr wichtig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Wartezeit zum Vorzeigen einer Arbeit ist in der Regel zu lang.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ein Assistent ist während der Kurszeit immer anwesend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik von unerfahrenen Assistenten hat weniger Qualität als die von Erfahrenen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich fühle mich manchmal unwohl beim Vorzeigen einer Arbeit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik der Assistenten ist tagesformabhängig (z.B. morgens strenger als nachmittags).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Noten im Kurs haben entscheidenden Einfluss auf meine Motivation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Notenvergabe empfinde ich als gerecht und nachvollziehbar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Das Bewertungssystem (Note 1-6) ist gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu
○ Ich vertraue dem Urteil der Assistenten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich habe das Gefühl, dass einzelne Studierende in der Kritik/ Bewertung bevorzugt werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Für gute Leistungen werde ich gelobt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Dieselbe Arbeit wird von verschiedenen Assistenten unterschiedlich bewertet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Dieselbe Arbeit wird vom <i>gleichen</i> Assistenten unterschiedlich bewertet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik der Assistenten ist manchmal herablassend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich habe den Eindruck, dass der Assistent sich ein Bild von mir gemacht hat.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich habe das Gefühl, dass einzelne Studierende in der Kritik/ Bewertung benachteiligt werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Bewertung durch einen Computer empfinde ich als gerecht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Bewertung des Computers ist nachvollziehbar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Computerurteile sind generell vertrauenswürdiger als menschliche Bewertungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Bewertung durch einen Computer empfinde ich als unpersönlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Bei Nichtbestehen eines praktischen Kurses/ Exmatrikulation könnte ich mir vorstellen, rechtlich dagegen vorzugehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Objektivere Verhältnisse würden das Arbeitsklima verbessern.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Ein Scheitern in einer Prüfungssituation bringe ich in Verbindung mit...

Ordnen Sie die Antworten nach Wichtigkeit.

Wichtigstes mit 1, Unwichtigstes mit 4.

- mangelnder Begabung
- Zufall (z.B. schlechter Tag/ Assistent zu streng)
- Anstrengung/ Fleiß (z.B. nicht genug geübt)
- Aufgabenschwierigkeit (Leistungsanforderung war zu hoch)

3. Ein Erfolg in einer Prüfungssituation bringe ich in Verbindung mit...

Ordnen Sie die Antworten nach Wichtigkeit.

Wichtigstes mit 1, Unwichtigstes mit 4.

- Begabung
- Zufall (z.B. Glück gehabt/ Assistent war gnädig)
- Anstrengung/ Fleiß (z.B. viel geübt)
- Aufgabenschwierigkeit (Leistungsanforderung war zu leicht)

4. Kreuzen Sie das am besten Zutreffende an.

	stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu
◦ Visuelle Rückmeldung (prep-Check) ist motivierender als mündliche Assistentenbewertung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
◦ Ich habe gemachte Fehler durch die Visualisierung besser verstanden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
◦ Die Darstellungen des Computerprogramms haben mich verwirrt und nicht geholfen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
◦ Die Art der Fehlerdarstellungen im Programm habe ich gut verstanden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
◦ Glauben Sie, dass Ihre heutigen Präparationen durch Assistenten schlechter bewertet worden wären?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Die Darstellung welches Kriteriums fanden Sie im Programm prepCheck am eingängigsten / besten verständlich? Drei Kreuze setzen.

- Präparationsgrenze
- Oberfläche
- Unterschnitt
- Abtrag zur ideellen Präparation
- Präparationswinkel

6. Kreuzen Sie das am besten Zutreffende an.

	ja	eher ja	eher nein	nein
○ Fällt es Ihnen schwer, sich alle Mängel an der Präparation im Gespräch mit dem Assistenten zu merken?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Haben Ihnen heute die Tipps der Assistenten zur Verbesserung der Präparation gefehlt (z.B. Walze steiler halten)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Fällt es Ihnen schwer, sich die Ergebnisse der Computerbewertung zu merken?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Motiviert Sie positive Rückmeldung (Lob) der Assistenten?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. Was halten Sie für die lehrreichste Methode eine Präparation zu erlernen?

*Ordnen Sie die Antworten nach Wichtigkeit.
Wichtigstes mit 1, Unwichtigstes mit 5.*

- Lehrfilm
- Demonstration am Phantomkopf
- Step-by-Step Abbildungen
- Assistentenkritik
- Computeranalyse (z.B. prepCheck)

8. Ihre Meinung: Welches System würden Sie bevorzugen?

	Computer	Assistent
○ Im Laboralltag	<input type="checkbox"/>	<input type="checkbox"/>
○ Bei Bewertung von Präptestaten	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 25: Fragebogen der Studenten Gruppe prepCheck

Anhang B – Fragebogen der Studenten Gruppe Assistent

Fragebogen Gruppe Assistent

Dieser Fragebogen dient zur Verbesserung der vorklinischen Ausbildung.
Die Auswertung wird ausschließlich von Studenten vorgenommen und ist anonymisiert.

1. Kreuzen Sie das am besten Zutreffende an.

	stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu
○ Das Betreuungsverhältnis Assistenten zu Studierenden ist gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ein Unterschied in der Kritik / Benotung zwischen den einzelnen Assistenten ist vorhanden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Assistenten wirken manchmal überfordert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik der Assistenten ist nachvollziehbar / gut erklärt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Aussagen und Verbesserungsvorschläge der Assistenten verwirren mich manchmal.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Objektivität ist mir bei Benotungen sehr wichtig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Wartezeit zum Vorzeigen einer Arbeit ist in der Regel zu lang.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ein Assistent ist während der Kurszeit immer anwesend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik von unerfahrenen Assistenten hat weniger Qualität als die von Erfahrenen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich fühle mich manchmal unwohl beim Vorzeigen einer Arbeit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik der Assistenten ist tagesformabhängig (z.B. morgens strenger als nachmittags).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Noten im Kurs haben entscheidenden Einfluss auf meine Motivation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Kreuzen Sie das am besten Zutreffende an.

	stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu
○ Die Notenvergabe empfinde ich als gerecht und nachvollziehbar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Das Bewertungssystem (Note 1-6) ist gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich vertraue dem Urteil der Assistenten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich habe das Gefühl, dass einzelne Studierende in der Kritik/ Bewertung bevorzugt werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Für gute Leistungen werde ich gelobt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Dieselbe Arbeit wird von verschiedenen Assistenten unterschiedlich bewertet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Dieselbe Arbeit wird vom <i>gleichen</i> Assistenten unterschiedlich bewertet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Kritik der Assistenten ist manchmal herablassend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich habe den Eindruck, dass der Assistent sich ein Bild von mir gemacht hat.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ich habe das Gefühl, dass einzelne Studierende in der Kritik/ Bewertung benachteiligt werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Das Erstellen einer guten Präparation mit dem Feedback der Assistenten ist leicht erlernbar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Fällt es Ihnen schwer, sich alle Mängel an der Präparation im Gespräch mit dem Assistenten zu merken?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Motiviert Sie positive Rückmeldung (Lob) der Assistenten?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Bei Nichtbestehen eines praktischen Kurses/ Exmatrikulation könnte ich mir vorstellen, rechtlich dagegen vorzugehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Objektivere Verhältnisse würden das Arbeitsklima verbessern.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Ein Scheitern in einer Prüfungssituation bringe ich in Verbindung mit...

Ordnen Sie die Antworten nach Wichtigkeit.

Wichtigstes mit 1, Unwichtigstes mit 4.

- mangelnder Begabung
- Zufall (z.B. schlechter Tag/ Assistent war zu streng)
- Anstrengung/ Fleiß (z.B. nicht genug geübt)
- Aufgabenschwierigkeit (Leistungsanforderung war zu hoch)

4. Ein Erfolg in einer Prüfungssituation bringe ich in Verbindung mit...

Ordnen Sie die Antworten nach Wichtigkeit.

Wichtigstes mit 1, Unwichtigstes mit 4.

- Begabung
- Zufall (z.B. Glück gehabt/ Assistent war gnädig)
- Anstrengung/ Fleiß (z.B. viel geübt)
- Aufgabenschwierigkeit (Leistungsanforderung war zu leicht)

5. Was halten Sie für die lehrreichste Methode eine Präparation zu erlernen?

Ordnen Sie die Antworten nach Wichtigkeit.

Wichtigstes mit 1, Unwichtigstes mit 4.

- Lehrfilm
- Demonstration am Phantomkopf
- Step-by-Step Abbildungen
- Assistentenkritik

Abbildung 26: Fragebogen der Studenten Gruppe Assistent

Anhang C – Fragebogen der Zahnärzte

Fragebogen Assistent

Dieser Fragebogen dient zur Verbesserung der vorklinischen Ausbildungssituation.
Die Auswertung erfolgt anonym und vertraulich.

1. Kreuzen Sie das am besten Zutreffende an.

	stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu
○ Das Betreuungsverhältnis Assistent - Studierende ist gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Betreuung der vorklinischen Kurse ist mit Stress verbunden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Meine Kritik während der Kurszeit beinhaltet Hinweise zur Verbesserung der Präparation (sich auf der Zahnreihe abstützen usw.).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Würden Sie sagen, dass es "typische Anfängerfehler" gibt, die nahezu immer vorkommen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Würden Sie sagen, dass Sie die Studierenden eines Kurses nach einer gewissen Zeit einschätzen können (strebsam, faul, geschickt, ungeschickt, usw.)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Das Bewertungssystem (Note 1-6) ist sinnvoll.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Intensive Betreuung (z.B. in Form von Demonstrationen) eines einzelnen Studierenden ist zeitlich möglich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Eine Gesamtnote im Testat für einen präparierten Zahn ist sinnvoll.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Das durchschnittliche Niveau eines Kurses beeinflusst die Benotung des einzelnen Studierenden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Kreuzen Sie vier Faktoren an, durch welche Ihre objektive Beurteilung am stärksten beeinflusst wird.

- Ermüdung
- Stress
- Frust
- "Betriebsblindheit" (viele Zähne in kurzer Zeit)
- Ablenkung
- Lärm
- Lichtverhältnisse
- eigene Erfahrungswerte
- Sympathie / Antipathie

3. Ordnen Sie folgende Attribute weiblichen Studierenden zu (vier ankreuzen):

- intelligent
- handwerklich ungeschickt
- räumliches Vorstellungsvermögen
- handwerklich geschickt
- kritikfähig
- unsicher
- setzen Instruktionen schnell um
- langsame Arbeitsweise (im Sinne von: faul, unmotiviert, resigniert)

4. Ordnen Sie folgende Attribute männlichen Studierenden zu (vier ankreuzen):

- setzen Instruktionen schnell um
- kritikfähig
- räumliches Vorstellungsvermögen
- handwerklich geschickt
- langsame Arbeitsweise (im Sinne von: faul, unmotiviert, resigniert)
- intelligent
- unsicher
- handwerklich ungeschickt

5. Schreiben Sie die aus Ihrer Sicht wichtigsten Ziele der vorklinischen Ausbildung auf:

6. Kreuzen Sie das am besten Zutreffende an.

	ja	eher ja	eher nein	nein
○ Der Anspruch einer objektiven studentischen Betreuung überfordert Sie manchmal.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Das plötzlich sehr gute Abschneiden eines schwachen Studierenden im Testat macht Sie misstrauisch.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Eine z.B. unzureichende Oberflächenglättung hat einen negativen Einfluss auf die Beurteilung der Präpgrenze.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Besonders auffällige Studierende prägen sich mit ihrer Platznummer ein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Eine Zuordnung von Platznummer und Studierenden in der Beurteilung ist anhand von Modellen möglich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Glauben Sie, dass die Reihenfolge der Beurteilung eine Rolle spielt (z.B. zuerst benotete Zähne strenger bewertet werden)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Liegt bei manchen Platznummern während des Benotungsprozesses eine gewisse Erwartungshaltung vor?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Würden Sie sagen, dass Sie eine Note auffallend häufiger vergeben als andere?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ein unsauberer Arbeitsplatz geht auch mit einer ungenauen Arbeitsweise einher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Glauben Sie, dass es mit Hilfe des "ersten Eindrucks" möglich ist, einzuschätzen wie gut ein Studierender im Kurs zurecht kommt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Würden Sie sagen, dass die Studierenden in den letzten Semestern immer ungeeigneter für das Zahnmedizinstudium geworden sind?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. Kreuzen Sie das am besten Zutreffende an.

	ja	eher ja	eher nein	nein
○ War Ihre eigene vorklinische Ausbildung im Hinblick auf die Leistungsbewertung objektiv?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Finden Sie, dass die vorklinische Ausbildung (wie sie jetzt ist) einen wichtigen, unverzichtbaren Bestandteil der zahnmedizinischen Ausbildung darstellt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Denken Sie, dass die jetzige vorklinische Ausbildung unter anderem der "Abhärtung" dienen sollte?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. Ordnen Sie folgende Attribute nach ihrer Wichtigkeit:

(für Kursabschluss (TPK/ Phantom) bedeutsam)
Wichtigstes mit 1, Unwichtigstes mit 6

- Intelligenz
- handwerkliches Geschick
- Kritikfähigkeit
- Verständnis für Zusammenhänge
- räumliches Vorstellungsvermögen
- Selbstvertrauen

9. Welche Attribute sind Ihnen bei Studierenden beim alltäglichen Umgang besonders wichtig? (drei ankreuzen)

- respektvolles Verhalten
- gepflegtes Äußeres
- Ausdrucksweise / Wortschatz
- Motivation
- Kritikfähigkeit
- _____

10. Kreuzen Sie das am besten Zutreffende an.

	ja	eher ja	eher nein	nein
○ Würden Sie sagen, dass eine Nichteinhaltung der unter 9. genannten Attribute ihre Kritik/ Benotung negativ beeinflusst?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	ja	eher ja	eher nein	nein
○ Beeinflusst Sympathie Ihre Notengebung positiv?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Beeinflusst Antipathie Ihre Notengebung negativ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Lassen Sie sich von zuvor erbrachten Leistungen (z.B. Aufwachsübung) eines Studierenden bei der Bewertung beeinflussen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Qualität der objektiven Bewertung ist bei allen Assistenten gleich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Lassen Sie sich bei Bewertungen von Kollegenmeinungen beeinflussen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Die Leistungen von Studierenden bei vorangegangenen Übungen lassen auf Zukünftige schließen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Wenn ein präparierter Zahn von mehreren Assistenten mit der gleichen Note beurteilt wurde, ist er objektiv bewertet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Eine Anonymisierung steigert die Objektivität.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
○ Ihre persönliche Meinung: Besteht Verbesserungsbedarf bei der vorklinischen Bewertung?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 27: Fragebogen der Zahnärzte

Anhang D – Konfiguration der Parameter

Substanzabtrag (Distanz zwischen Präparation und Goldstandard):

Beginn des Toleranzbereichs:	-0,300 mm
Ende des Toleranzbereichs:	0,300 mm
Beginn des Messbereichs:	-0,400 mm
Ende des Messbereichs:	0,400 mm
Vorzeichen:	Ja

Oberflächenbeschaffenheit:

Winkel einer scharfen Kante:	39,0°
------------------------------	-------

Randabschlusstyp:

Vertikaler Messbereich:	2 mm
Anzahl Messpunkte, an denen die Auswertung erfolgt:	72
Toleranzbereich:	0,4 mm

Präparationsrand:

Halbe Fenstergröße:	0,42 mm
Röhrenradius:	0,053 mm

Präparationswinkel:

Tabelle 20: Konfiguration der Präparationswinkel

Variable	Präparationswinkel Außenfläche	Präparationswinkel Höckerabschrägung
Aktiviert	Ja	Ja
Maximaler Präparationswinkel (°)	15	34
Minimaler Präparationswinkel (°)	5	18
Maximale vertikale Höhe (%)	60	95
Minimale vertikale Höhe (%)	25	80
Abstand vom Präparationsrand (mm)	1,0	-
Messpunkte (Anzahl)	180	-
Maximaler Sektor	180	140
Minimaler Sektor	0	40

Randabschlusstyp:

Hohlkehle

Vertikaler Messbereich:

2 mm

Anzahl der Messpunkte, an denen die Auswertung erfolgt:

72

Toleranzbereich:

0,4 mm

Anhang E – Kalibrierungshilfe für die Bewertung der Oberflächenbeschaffenheit

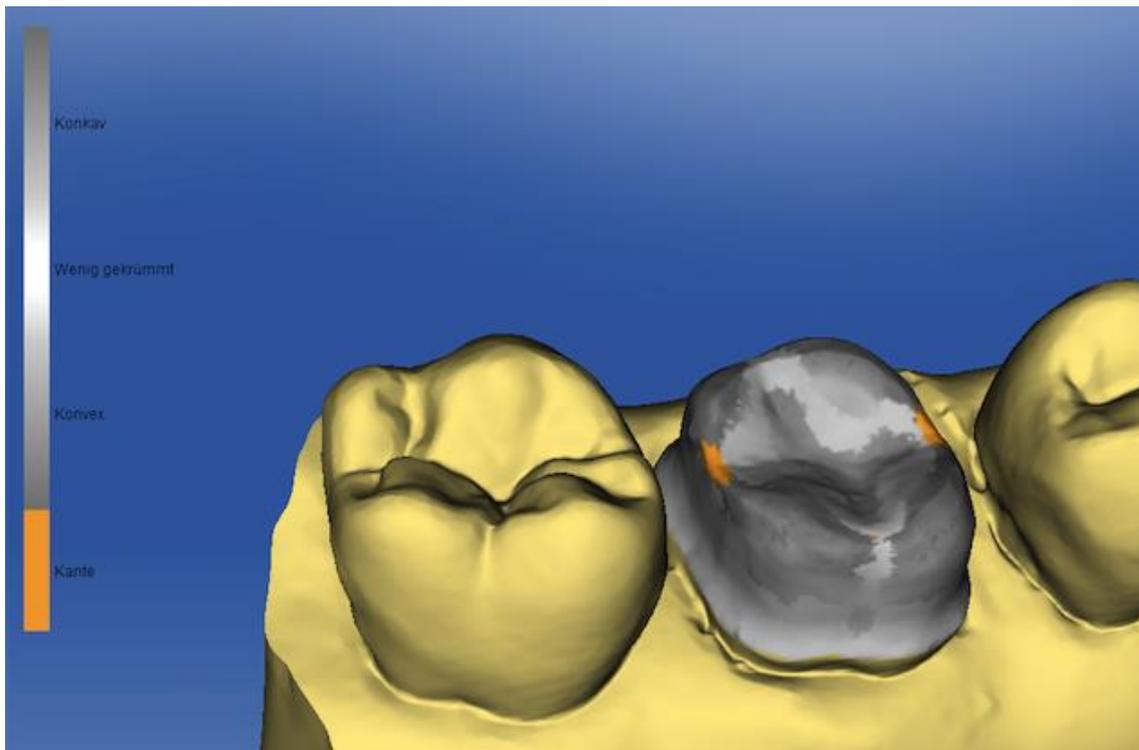


Abbildung 28: Kalibrierungshilfe für die Note sehr gut

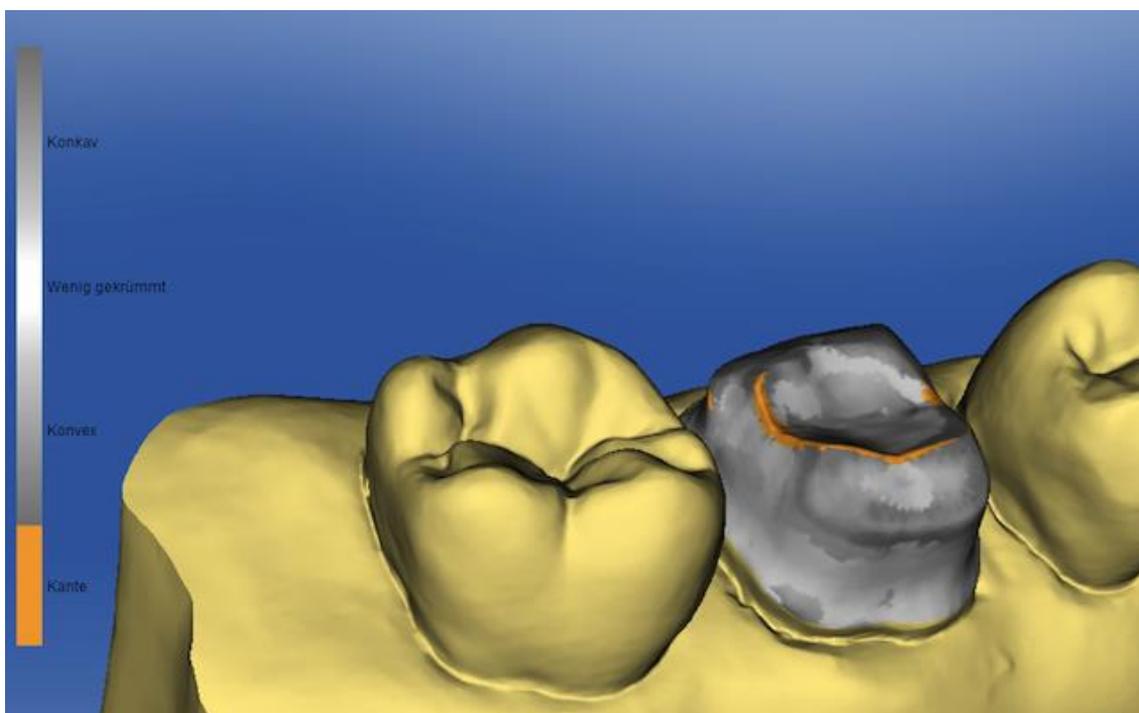


Abbildung 29: Kalibrierungshilfe für die Note akzeptabel

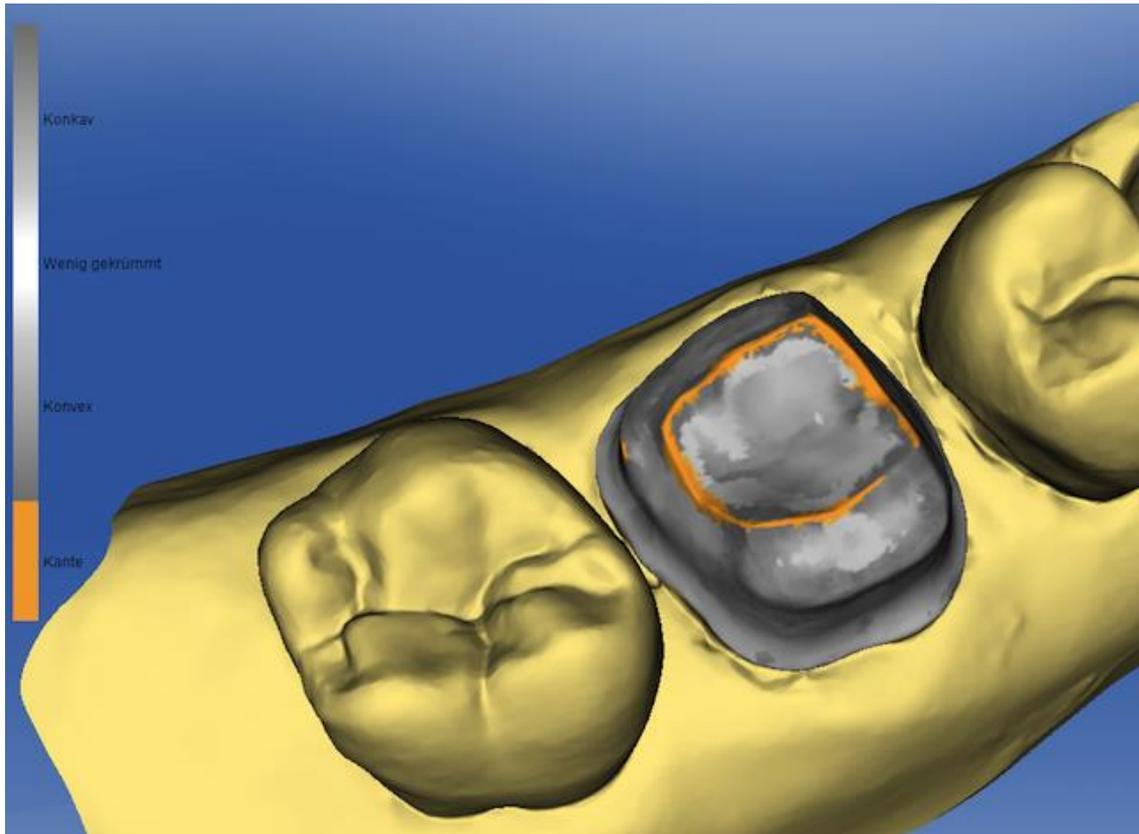


Abbildung 30: Kalibrierungshilfe für die Note verbesserungswürdig