

## Anhang 1: Ergänzendes Material

### **Bisherige Evidenzen**

Bislang gibt es wenige vielversprechende Studien, die die technische Umsetzung der Webcam Eye-Tracking Methode untersucht und eine zufriedenstellende Datenqualität gezeigt haben. Zum Beispiel zeigten Semmelmann und Weigelt [50] eine akzeptable Datenqualität während Fixations-, Verfolgungs- und freien Betrachtungsaufgaben. Bott et al. [3] berichteten von hohen Korrelationen von  $r=0.81$ , wenn Webcam Eye-Tracker und Labor-Eye-Tracker gleichzeitig verwendet wurden. Bánki et al. [2] verglichen Webcam mit Labor Eye-Trackern, um das Blickverhalten von Säuglingen zu erfassen, und berichteten nach sorgfältiger Qualitätskontrolle positive Ergebnisse. Obwohl diese ersten Studien ermutigend sind, gibt es keine Forschung über den Einsatz solcher Technologien im Bildungskontext oder über potenzielle Anwendungsbereiche in der medizinischen Ausbildung. In diesem Zusammenhang ergeben sich mehrere Fragen. Sind die Studierenden bereit, die Webcam-Augenverfolgungsmethode zu verwenden und die Anweisungen ordnungsgemäß zu befolgen? Da die Studierenden unterschiedliche Hardware (Laptops und Webcams) verwenden und diese Studien in nicht überwachten Umgebungen durchführen, wie hoch ist die Datenqualität? Wenn die Qualität geeignet ist, können die Ergebnisse verwendet werden, um die Entwicklung der visuellen Expertise bei Medizinstudierenden („innerhalb der Probanden“) zu beschreiben? Wie gut können die Ergebnisse zwischen verschiedenen Lernenden („zwischen den Probanden“) unterscheiden?

### **Poweranalyse zur Bestimmung der Stichprobengröße**

Unter der Annahme einer mittleren Effektstärke von  $d=0.5$ , einem Signifikanzniveau von  $\alpha=0.05$  und einer Teststärke ( $1-\beta$ ) von 0.80 ergibt die Schätzung der Stichprobengröße eine Mindestanzahl von  $N=34$  Teilnehmenden pro Gruppe für t-Tests. Für multiple lineare Regressionen mit einer mittleren Effektstärke von  $f^2=0.15$ , einem Signifikanzniveau von  $\alpha=0.05$  und einer Teststärke ( $1-\beta$ ) von 0,80 wurden 68 Teilnehmende benötigt.

### **Details zu den teilnehmenden Studierenden**

Von den etwa 180 eingeschriebenen Studierenden des Kurses haben 74 Teilnehmende (=100%) an t1 die Kalibrierungsphase ordnungsgemäß abgeschlossen, 63 Teilnehmer haben

den Test abgeschlossen und Ergebnisse bereitgestellt (=85%). An t2 haben 115 Teilnehmer (=100%) die Kalibrierungsphase bestanden, und N=95 haben den Test abgeschlossen (=83%). Insgesamt haben N=42 Studierende (durchschnittliches Alter  $21.49 \pm 1.92$  Jahre; 28 weiblich) an beiden Messzeitpunkten t1 und t2 teilgenommen. Die Teilnahme an dieser Studie war freiwillig, anonym und hatte keinen Einfluss auf die Noten der Studierenden.

### **Beschreibung des online Histologie-Curriculums**

Aufgrund der Covid-19-Pandemie wurde der gesamte Kurs in einem synchronen Online-Format unterrichtet. Insgesamt gab es 20 Kurstage mit jeweils drei Stunden Unterricht pro Kurstag. Neben dem Inhalt der Zellbiologie und grundlegenden Gewebetypen (Kurstage 1-4) wurden fast alle Organe systematisch an den folgenden Kurstagen behandelt (7-20). Der Kurs wurde ausschließlich online abgehalten, wobei Videokonferenz-Tools und virtuelle Mikroskopie verwendet wurden [10]. Die Studierenden hatten außerhalb des Campus Zugriff auf die virtuelle Mikroskopie, die alle relevanten Kurspräparate enthielt [22]. Dadurch erwarteten wir eine intensivere Auseinandersetzung mit den Kurspräparaten und eine Verbesserung der visuellen Expertise zum Ende des Kurses.

### **Maßnahmen zur Sicherstellung einer hohen Datenqualität**

Da die Qualität der Daten bei der Webcam-Augenverfolgungsmethode eine zentrale Rolle spielt, wurde besonderes Augenmerk auf die Steigerung der Qualität gelegt [25]. Eine Pilotstudie wurde mit zwei Personen durchgeführt, um die Studienumgebung zu optimieren. Dabei lag der Schwerpunkt darauf, die Dauer der Präsentationszeit anzupassen und das Verhalten der Teilnehmenden während der Studie einzuschätzen. Die Teilnehmenden wurden insgesamt dazu ermutigt, auf eine gute Beleuchtung und eine stabile Kopfposition zu achten. Wenn sich der Kopf des Teilnehmers zur Seite bewegte, wurde die Studie automatisch pausiert und die Teilnehmenden wurden aufgefordert, sich neu zu positionieren. Die Teilnehmenden wurden gebeten, aufrecht zu sitzen, einen konstanten Abstand zur Webcam einzuhalten, für ausreichend Licht zu sorgen, den Kopf ruhig zu halten und mögliche Ablenkungen zu vermeiden.

Die Webcam Eye-Tracking Technologie geht davon aus, dass der Blickfokus der Position des Mausclicks entspricht [44], [45]. Am Anfang der Studie wurde so eine 40-Punkt-Kalibrierung durchgeführt, indem 40 Klicks und Cursorbewegungen mit drei verschiedenen

Hintergrundfarben (je 13 Punkte mit RGB #FFFFFF, #000000, #919191) verwendet wurden. Anschließend wurde ein 4-Punkt-Test zur Genauigkeit durchgeführt. Um den Genauigkeitstest zu bestehen, mussten der visuelle Fokus der Teilnehmenden für jeden Punkt zu mehr als 50% innerhalb eines Toleranzradius von 200 px um diesen Punkt liegen. Der Versuch wurde beendet, wenn die Kalibrierung nach fünf Versuchen nicht erfolgreich war. Aufgrund der kurzen Studiendauer (ca. 10 Minuten) wurde keine Rekalibrierung durchgeführt und es wurden keine Pausen eingelegt. Zwischen zwei Stimuli wurde ein Separator verwendet, der ein Fadenkreuz in der Mitte des Bildschirms anzeigte, um die Netzhaut zu entspannen und die Augenpositionen auf die Mitte zu fixieren. Das Interstimulusintervall betrug 1.5 s. Es wurde eine ausreichende Datenqualität der Augenverfolgung in Bezug auf Genauigkeit, Webcam Bildrate, Datenintegrität und Blick-auf-Bildschirm-Rate erzielt.

Genauigkeit wurde als durchschnittlicher Unterschied zwischen dem Mausklick und der gemessenen Blickposition in Pixeln definiert (entspricht dem Konzept der internen Validität). Präzision wurde als Streuungsmaß der Standardabweichung definiert (entspricht dem Konzept der Reliabilität). Die Bildrate gibt die Anzahl der pro Sekunde in Hz gemessenen Blickpositions-Messungen an. Datenintegrität bezieht sich auf die Vollständigkeit der Daten als Prozentsatz der Stichprobe, der Augenbewegungsdaten liefert. Fehlende Daten wurden linear interpoliert und Blickpositionen wurden mit einem Reduktionsniveau von 21 entauscht (d.h., der Median wurde für 21 aufeinanderfolgende Punkte berechnet). Visuelle Winkel konnten nicht berechnet werden, da keine genauen Schätzungen für die Entfernung der Teilnehmer zum Bildschirm vorlagen. Es wurden keine Kinnstützen verwendet. Die Blickpositionsdaten wurden in Pixeln gemessen. Die meisten Studierenden verwendeten eine Bildschirmauflösung von 1440x900 px (t1=22%; t2=36%), 1536x864 px (t1=11%; t2=16%) oder 1280x720 px (t1=9%; t2=16%). Da sich die Auflösungen der Bildschirme der Teilnehmer unterschieden, wurden die Blickpositionen auf Höhe und Breite (in %) normalisiert. Die ersten 0.5 Sekunden der Daten wurden für statistische Tests weggelassen, um eine zentrale Bildschirmverzerrung zu vermeiden. Fixationsevents wurden mit einem Algorithmus ähnlich dem I-VT Fixationsfilter (unter Verwendung von Winkelwerten, ausgedrückt als Prozentsatz der Größe des Objekts, anstelle der Winkelgeschwindigkeit) berechnet [41]. Durch den Algorithmus automatisch erkannte Events wurden manuell überprüft. Die minimale Fixationsdauer wurde auf 100 ms festgelegt [24].

### **Details zu den Stimuli**

Aus einem Pool histologischer Präparate (= Stimuli) haben wir diejenigen ausgewählt, die bereits zuvor im Kurs besprochen wurden und die basierend auf unserer Erfahrung besonders trennscharf waren. Der Pool bestand daher zum Zeitpunkt 1 aus etwa 60 verschiedenen Präparaten und zum Zeitpunkt 2 aus 200 (60+140) potentiellen Präparaten. Eine hohe Anzahl möglicher Präparate kann die Schwierigkeit Identifikationsaufgabe erhöhen, da es mehr differentialdiagnostische Möglichkeiten gibt. Die Schwierigkeit für die Studierenden zum Zeitpunkt 2 war somit höher, da theoretisch eine größere Auswahl zur Verfügung stand. Dies kann wiederum zu einer Unterschätzung der in dieser Studie gefundenen Effekte führen.

### **Details zur Datenanalyse**

Um eine Multikollinearität unter den Prädiktoren zu überprüfen, wurde der Varianzinflationsfaktor (VIF) bestimmt, der keine Probleme mit Kollinearität zeigte (VIFs <6). Es wurde keine vorherige Transformation der Variablen durchgeführt. Stattdessen wurden Bootstrap-Analysen mit BCa-Korrektur an n=5000 Stichproben durchgeführt, um Nichtnormalverteilungen zu berücksichtigen [53]. Bei unbekanntem Verteilungen kann das Bootstrapping als Resampling-Ansatz verwendet werden.

### **Beschreibung der Eye-Tracking Variablen**

Um die Effektivität der Prozessierung zu messen, wurden bestimmte Augenbewegungen, wie die Fixationsanzahl, Fixationsdauer und der Scanpfadlänge (die Summe der Länge zwischen zwei Fixationen in % des Bildschirms) in die Untersuchung einbezogen (siehe Tabelle S1). Um die Vergleichbarkeit sicherzustellen, wurde die Fixationsanzahl von Studierenden, die die Aufgaben früher abgeschlossen hatten, linear extrapoliert (= an die Betrachtungszeit angepasste Werte). Dadurch werden Unterschiede in der Betrachtungszeit berücksichtigt und ermöglichen einen fairen Vergleich. Für die Regressionsmodelle, die zur Beantwortung der Forschungsfrage 3 erforderlich waren, wurden die unadjustierten Werte verwendet, um für die unterschiedliche Betrachtungszeit zu kontrollieren. Sakkaden (schnelle Augenbewegungen von einem Punkt zum anderen) werden typischerweise für die Erfassung von holistischer Verarbeitung genutzt [51]. Aufgrund einer niedrigen Bildrate unter 30 Hz (siehe Abbildung 3) haben wir jedoch auf die Detektion von Sakkaden oder anderen schnellen Ereignissen verzichtet [13].

## Definition von relevanten und irrelevanten Bildbereichen

Unter Berücksichtigung vergleichbarer Aufgabenschwierigkeit und Färbung an beiden Zeitpunkten zeigten die Präparate Regionen mit unterschiedlichen Vergrößerungen (siehe Abbildung 2). Bestimmte Bildbereiche (*areas of interest*; AOIs) wurden verwendet, um Augenbewegungen bestimmten Abschnitten des Bildes zuzuordnen. Die Präparate und AOIs wurden zuvor in einem Expertenkonsens (vier Histologiedozierende mit über 15 Jahren Erfahrung) ausgewählt, und es wurden nur solche Folienbereiche ausgewählt, die 1-3 verschiedene AOIs enthielten. Die vAOIs wurden unter Verwendung automatisch generierter Salienzkarten (OpenCV Saliency Detection) (Rosebrock, 2018) ausgewählt und manuell überprüft. Alle AOIs in den Bildern hatten vergleichbare Größen (zwischen 15-20% der Folie), und die dAOIs und vAOIs auf denselben Präparaten hatten die gleiche Größe (siehe Abbildung 5a). Insgesamt waren die dAOIs größer als die potenzielle Zielstruktur, um Messungenauigkeiten auszugleichen. Wir verwendeten keine Masson-Goldner-Färbung, um Farbverläufe zwischen Rot und Grün für Studierende mit Farbenblindheit zu berücksichtigen, und wir wählten nur diejenigen Präparate aus, die zuvor im Kurs unterrichtet wurden. Die Identifikation der Präparate war eindeutig, und die dAOI und die vAOI überlappten sich nicht. Fixationen wurden als Teil eines AOIs betrachtet, wenn die Fixation im Bereich des AOIs begann. Webcam Eye-Tracking zeigt die niedrigste Genauigkeit in den Ecken des Bildschirms. Darüber hinaus wird der zentrale Punkt des Bildschirms zu Beginn der Bildbetrachtung artifiziell fixiert (zentrale Verzerrung oder zentraler Fixationsbias) [13]. Um beide potenziellen Störfaktoren auszugleichen, wurden die AOIs zwischen der Mitte und Ecke des Bildschirms platziert (siehe Abbildung 5a).

Supplementäre Tabelle S1: Item Analyse für die Stimuli

Zeitpunkte	Itemschwierigkeit	95% CI	SD	Korrigierte Item- Total Korrelation	Cronbach's alpha
t1 (n = 6)					0.66
Präparat 1	0.33	0.21;0.46	0.48	0.34	
Präparat 2	0.23	0.12;0.34	0.43	0.34	
Präparat 3	0.17	0.07;0.26	0.38	0.51	
Präparat 4	0.15	0.06;0.24	0.36	0.54	
Präparat 5	0.30	0.18;0.42	0.46	0.31	
Präparat 6	0.35	0.23;0.47	0.48	0.40	
t2 (n = 6)					0.47
Präparat 1	0.82	0.74;0.89	0.39	0.28	
Präparat 2	0.62	0.21;0.40	0.49	0.16	
Präparat 3	0.31	0.21;0.40	0.46	0.20	
Präparat 4	0.20	0.12;0.28	0.40	0.34	
Präparat 5	0.84	0.84;0.77	0.37	0.17	
Präparat 6	0.47	0.37;0.57	0.50	0.27	

Abkürzungen: M = Mittelwert; CI = Konfidenzintervall; SD = Standardabweichung

Supplementäre Abbildung S1

