

Das LUCA-Dashboard im Usability-Test – Eine gaze-cued retrospective Think-Aloud-Studie

GEORG DARIUSH GORSHID, CHRISTIAN MAYER, ANDREAS RAUSCH, JÜRGEN SEIFRIED

Zusammenfassung

Learning Analytics Dashboards (LADs) eröffnen Lehrpersonen die Möglichkeit, Lernprozesse in computergestützten Lernumgebungen zu beobachten und zu analysieren. Ob ein LAD als praktikabel wahrgenommen wird, hängt hauptsächlich vom Design des LAD sowie der Erfahrung der nutzenden Person (bzw. der Interaktion beider Komponenten) ab. Im vorliegenden Beitrag wurden im Rahmen eines Usability-Testing-Ansatzes mittels Eye-Tracking und gaze-cued retrospective Think-Aloud zwei Prototypen eines LAD getestet. Acht Lehrende bearbeiteten mehrere Aufgaben. Im Anschluss wurden die Testpersonen aufgefordert, anhand der Eye-Tracking-Aufzeichnungen ihren Gedankengang rückblickend zu verbalisieren. Ergänzend wurden mittels eines standardisierten Fragebogens Daten zur Beurteilung der Nutzererfahrung erhoben. Durch den Mixed-Method-Ansatz konnten knapp 150 Usability-Probleme identifiziert werden. Im Beitrag berichten wir die empirischen Befunde und diskutieren das methodische Vorgehen.

Schlagworte: Learning Analytics Dashboard, Usability-Testing, Eye-Tracking, Gaze-cued-retrospective-think-aloud

Abstract

Learning analytics dashboards (LADs) enable teachers to monitor and analyze individual learning progresses in computer-based learning environments. Whether a LAD is perceived as viable depends primarily on users' experiences interacting with the LAD's design. In a usability testing approach, eye-tracking and gaze-cued retrospective think-aloud are used to test two prototypes of a LAD. Eight teachers were asked to complete several tasks in the prototypes while their eye movements were recorded. Subsequently, the subjects were asked to verbalize their thought process retrospectively based on the recordings. To complement this, a standardized questionnaire was used to assess the user experience. Almost 150 usability problems could be identified. In this paper, we present the empirical findings and discuss the methodological approach. Finally, implications and limitations are pointed out.

Keywords: learning analytics dashboard, usability testing, eye-tracking, gaze-cued-retrospective-think-aloud

1 Ausgangslage

Digitale Lernumgebungen stehen mittlerweile für viele Bildungsbereiche in großer Auswahl zur Verfügung. Für die kaufmännische Aus- und Weiterbildung wurde solch eine Lehr-Lern-Umgebung jüngst im Rahmen des Projekts „Problemlöseanalytik in Bürosimulationen“ (PSA-Sim) in der vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungs- und Transfer-Initiative „Technology-based Assessment of Skills and Competences in VET (ASCOT+)“ entwickelt. Die webbasierte Bürosimulation LUCA¹ zielt darauf ab, den Erwerb beruflicher Handlungskompetenz durch die Bearbeitung komplexer authentischer Arbeitsszenarien mittels typischer kaufmännischer Softwarewerkzeuge wie beispielsweise einer Tabellenkalkulation oder eines Enterprise-Resource-Planning-Systems zu fördern (Rausch, Deutscher, Seifried, Brand & Winther, 2021). Um Immersion und Authentizität sicherzustellen, können Lernszenarien in ein fiktives Modellunternehmen eingebettet werden, in dem die Lernenden in kaufmännischen Settings agieren. Die Plattform besteht aus vier Komponenten: Das LUCA-Office ermöglicht den Lernenden die Bearbeitung authentischer Arbeitsszenarien mithilfe bürotypischer Softwarewerkzeuge, der LUCA-Editor ermöglicht Lehrenden die Erstellung und Bearbeitung von Arbeitsszenarien, der LUCA-Manager dient dem Erstellen und Verwalten von Lernprojekten und enthält das LUCA-Dashboard, das Lehrenden während der Bearbeitungsprozesse das Monitoring der Lernenden ermöglicht. Sobald Lernende mit der LUCA-Office-Oberfläche interagieren, werden kontinuierlich Daten generiert und anschließend durch eine Problemlöseanalytik in Echtzeit verarbeitet. Lehrkräfte verfolgen und begleiten die Lernprozesse in einem dynamischen Learning Analytics Dashboard (LAD), das aggregierte Informationen zu den aktuellen Problemlöseschritten der Teilnehmenden in unterschiedlichen Ansichten zur Verfügung stellt. Von zentraler Bedeutung sind hierbei u. a. die Benutzerfreundlichkeit des Dashboards (Usability) und das mit der Nutzung einhergehende positive Erleben (User Experience; UX).

Um Usability und User Experience des LUCA-Dashboards sicherzustellen, wurden bereits in der Entwicklungsphase Lehrkräfte aus kaufmännischen Schulen einbezogen. Im Zuge der hier präsentierten Studie wurden im Softwareentwicklungsprozess zwei Dashboards entworfen und in Form von Click-Dummys, die das produktive System simulieren, implementiert. Dashboard 2 (DB 2) stellt dabei ein alternatives Design von Dashboard 1 (DB 1) dar, bei der Endanwenderinnen- und Endanwenderfeedback zu DB 1 berücksichtigt wurde. Zudem wurde hier noch stärker darauf geachtet, aktuelle LAD-Designprinzipien (Klerkx, Verbert & Duval, 2017) zu realisieren (s. u., Abschnitt 3.2). Zur Analyse von Usability und UX wurden die Interaktionen mit dem System durch potenzielle Nutzende aufgezeichnet und ausgewertet (zu dieser Vorgehensweise siehe z. B. Bojko, 2013; Klein & Ettinger, 2019; Kuniavsky & Moed, 2012). Während die Testpersonen typische Dashboard-Aufgaben bearbeiteten, wurden

¹ Die LUCA-Office-Simulation wurde in Kooperation mit dem Softwarehaus Cap3 (www.cap3.de) entwickelt. Der Name der Software ist an Luca Pacioli (1445–1514), einen italienischen Mathematiker und Pionier der kaufmännischen Didaktik, angelehnt.

Befragungsdaten sowie Blickbewegungsdaten (Eye Tracking) und Think-Aloud-Daten erhoben. Im vorliegenden Beitrag werden folgende Fragestellungen bearbeitet:

1. Inwiefern unterscheidet sich das User-Verhalten in den beiden Dashboard-Varianten?
2. Welche Usability-Probleme lassen sich für die beiden Dashboard-Varianten identifizieren?

Im Folgenden werden zunächst die Grundlagen zum Usability-Testing sowie der aktuelle Stand der Forschung zur Usability von Learning Analytics Dashboards umrissen (Abschnitt 2). Darauf aufbauend werden das methodische Vorgehen erläutert (Abschnitt 3) sowie die empirischen Befunde berichtet (Abschnitt 4). Abschließend werden in Abschnitt 5 die zentralen Ergebnisse sowie Limitationen und Implikationen der Untersuchung diskutiert.

2 Usability und User Experience (UX) von Dashboards

2.1 Ziele und Funktionen von Learning Analytics Dashboards (LADs)

Ein LAD wird definiert als „Anzeige, die verschiedene Indikatoren über den/die Lernenden, den/die Lernprozess/e und/oder den/die Lernkontext/e in einer oder mehreren Visualisierungen zusammenfasst“ (Schwendimann et al., 2017, S. 38). LADs stellen in komprimierter Form Informationen bereit (z. B. mit Blick auf Monitoring, Reflexion und Regulierung des Lernens), die aus verschiedenen Ressourcen generiert werden (Schwendimann et al., 2017). Hierzu zählen u. a. Logdaten aus den Interaktionen der Lernenden mit dem System, verwendete oder erstellte Lernartefakte, Fragebogen- oder Interviewdaten, physische Daten von Sensoren oder Daten externer Quellen. Zentrale Kennwerte (Indikatoren) werden in grafischer Form (Glahn, 2009) zum Zwecke der Reduktion der kognitiven Belastung der Nutzenden aufbereitet (Ali, Hatala, Gašević & Jovanović, 2012). So können Aspekte des Lernverhaltens (Einflussfaktoren, Prozesse und Ergebnisse) dokumentiert und analysiert sowie durch gezieltes Feedback gefördert werden (Verbert, Duval, Klerkx, Govaerts & Santos, 2013). LADs bieten somit vertiefte Einblicke in individuelle Lernprozesse und zeigen den Bedarf an möglichen Interventionen auf, um Lernende kurzfristig zu unterstützen. Die Möglichkeit, das Lernverhalten zu analysieren und entsprechende Daten mit den Lernergebnissen in Beziehung zu setzen, kann mittel- bis langfristig gewünschtes Verhalten fördern. Verbert et al. (2013) beschreiben LADs vor diesem Hintergrund als Anwendungen, die Werkzeuge zur Überprüfung und Analyse des Verhaltens und der Handlungen der Nutzenden bereitstellen.

Nutzende von LADs können Lehrende, Lernende, Administratorinnen und Administratoren oder Forschende sein, wobei sich die Ausgestaltung von LADs an der jeweiligen Zielgruppe orientieren sollte (Schwendimann et al., 2017). LADs für Lernende sollen z. B. die Leistung und Retention oder die Mitarbeit in Gruppen verbessern, indem sie Informationen über das eigene Handeln bereitstellen und dieses in

Verhältnis zum Verhalten anderer Personen setzen (Park & Jo, 2015). Varianten für Lehrende zielen dagegen auf eine Verbesserung der Instruktion ab. Dazu werden z. B. Aktivitäten und Ergebnisse einer Lerngruppe angezeigt, unterdurchschnittlich aktive Lernende identifiziert oder potenzielle Interventionen vorgeschlagen (Park & Jo, 2015). Indikatoren können in verschiedene Kategorien unterteilt werden (z. B. Lernende, Handlung, Inhalt, Ergebnis, Kontext, soziale Interaktion) und stehen auf Individual-ebene und stellenweise auch auf Gruppenebene zur Verfügung (Schwendimann et al., 2017). Schließlich sind auch verschiedene Einsatzkontexte wie z. B. Präsenzunterricht, Gruppenarbeiten in Präsenz und Online- oder Blended Learning zu unterscheiden (siehe Verbert et al., 2013). Die Lernwirksamkeit des Einsatzes von LADs wird in der Literatur indes auch kritisch diskutiert. Eine systematische Literaturanalyse von Valle, Antonenko, Dawson und Huggins-Manley (2021) kommt zu dem Schluss, dass die Annahme einer lernförderlichen Wirkung von LADs nicht theoretisch fundiert sei. Zudem fehle es auch an empirischer Evidenz.

2.2 Usability und User Experience (UX)

Faktoren, die die Nutzung einer Softwareanwendung beeinflussen, werden seit den 1980er-Jahren beispielsweise im Kontext des Technology Acceptance Model (TAM) thematisiert. Im Zentrum vieler Ansätze stehen dabei die wahrgenommene Nützlichkeit und der wahrgenommene Grad der Schwierigkeit der Bedienbarkeit (Davis, 1985). Das etwas jüngere Konzept der User Experience (UX) beschreibt „die Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder erwarteten Nutzung eines Systems, Produkts oder einer Dienstleistung resultieren“ (ISO 9241–11, 2018, S. 33). Hassenzahl (2003) schlägt ein hedonisch-pragmatisches Modell der UX vor. Während sich die hedonische Ebene auf die Wirkung der Interaktion auf die Emotionen und Einstellungen der Nutzenden bezieht, geht es bei der pragmatischen Ebene um die Fähigkeit eines Systems, aufgabenbezogene Bedürfnisse zu befriedigen (Hassenzahl, 2003). Die pragmatische Ebene von UX überschneidet sich somit mit dem Konzept von Usability, verstanden als das „Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung von bestimmten Nutzenden verwendet werden kann, um bestimmte Ziele mit Effektivität, Effizienz und Zufriedenheit in einem bestimmten Nutzungskontext zu erreichen“ (ISO 9241–11, 2018, S. 9). Effektivität, Effizienz und Zufriedenheit werden dabei wie folgt umrissen:

- Effektivität beschreibt die Präzision und Vollständigkeit, mit der User:innen ein bestimmtes Ziel erreichen kann.
- Effizienz setzt die zur Erreichung eines bestimmten Ziels aufgewendeten Ressourcen ins Verhältnis zur Effektivität.
- Zufriedenheit ist definiert als „Freiheit von Unbehagen und positive Einstellung gegenüber der Nutzung des Produkts“ (ISO 9241–11, 2018, S. 11).

Die genannten drei Faktoren werden nach Thoma und Dodd (2019) maßgeblich von den Nutzenden (z. B. in Abhängigkeit der Vorerfahrungen bzw. der Expertise), den Zielen (z. B. in Abhängigkeit des Schwierigkeitsgrades, Vergleichbarkeit der Aufga-

ben) und dem spezifischen Kontext der Nutzung beeinflusst (z. B. Einsatzbereich des Produkts). Dementsprechend werden Usability-Probleme definiert als „eine Reihe von negativen Phänomenen, wie die Unfähigkeit des Nutzenden, ein Ziel zu erreichen, ineffiziente Interaktion und/oder die Unzufriedenheit des Benutzers, verursacht durch eine Kombination von Faktoren der Benutzeroberfläche und des Nutzungskontextes“ (Manakhov & Ivanov, 2016, S. 3146). Einen Ansatz zur weiteren Operationalisierung der aufgeführten Kriterien legt Herczeg (2018) mit Bezugnahme auf diverse Kriterien, Modelle und Verordnungen vor (siehe auch Sangmeister et al., 2018). Herczeg (2018) bezieht sich bei der Beurteilung der wahrgenommenen Interaktion der Nutzenden mit Objekten und Funktionen eines Systems auf die ISO 9241-110-Dialogkriterien, die mittlerweile durch die ISO 9241-110:2020-Interaktionsprinzipien ersetzt wurden (ISO 9241-110, 2020). Eingabe- und Ausgabekriterien beurteilen die Handhabbarkeit durch die Nutzenden und die Darstellung von Informationen durch das System und beziehen sich auch auf die Grundsätze der Informationsdarstellung (ISO 9241-112, 2017). Grundsätzlich wird für die Sicherstellung der Usability empfohlen, potenzielle Nutzende von Anfang an in den Designprozess einzubeziehen, die Anwendung testen zu lassen und iterativ vorzugehen (User-Centered-Design, siehe Klerkx et al., 2017). Zur Evaluation der Usability können verschiedene Methoden Anwendung finden bzw. kombiniert werden (u. a. Eye-Tracking und Befragung, siehe Kuhnel, Seiler, Honal & Ifenthaler, 2018).

3 Methodisches Vorgehen

3.1 Stichprobe und Studiendesign

Potenzielle Endanwendende des hier im Blickpunkt stehenden Dashboards der LUCA-Bürosimulation (siehe Abschnitt 1) sind Lehrkräfte an kaufmännischen Schulen mit den unterrichtlichen Schwerpunkten Betriebswirtschaftslehre, Rechnungswesen, Volkswirtschaftslehre, Datenverarbeitung und Informationsmanagement. Der Usability-Test wurde mit acht Personen dieser Zielgruppe zum Ende des Schuljahres 2021 durchgeführt. Dieser Stichprobenumfang kann für die Fragestellung der Studie als angemessen erachtet werden (siehe hierzu Thoma & Dodd, 2019). Die Teilnehmenden waren zum Erhebungszeitpunkt zwischen 28 und 52 Jahre alt ($M=40.25$, $SD=9.9$), überwiegend männlich (sechs der acht Testpersonen) und mit vollem Deputat beschäftigt (ebenfalls sechs von acht). Die Teilnahme an der Untersuchung erfolgte auf freiwilliger Basis und wurde nicht vergütet.

Die Testung erfolgte anhand eines standardisierten Ablaufplans mit konkreten Verhaltensvorgaben für die Versuchsleiter und standardisierten Hilfestellungen für die Probandinnen und Probanden. Nachdem sich die Testpersonen mit der Bürosimulation bzw. dem Dashboard vertraut gemacht hatten und auf das Eye-Tracking-System kalibriert wurden, bearbeiteten sie zwei interaktive Prototypen (Click-Dummys bzw. Medium-Fidelity-Prototypen). Diese stellen partiell funktionelle klickbare Ansichten dar, welche aber noch nicht das voll funktionsfähige Endprodukt umfassen

(zur weiteren Unterscheidung von Prototypen siehe Olmsted-Hawala, Holland & Quach, 2014). In den Prototypen wird eine Gruppe von fiktiven Lernenden simuliert (sämtliche Namen in den nachstehenden Abbildungen sind zu Illustrationszwecken frei erfunden), die simultan ein komplexes domänenspezifisches Problem in der Bürosimulation lösen. Die beiden Entwürfe (Abbildung 1 und 2) umfassen Übersichtsseiten, zwischen denen auf unterschiedlichen Wegen navigiert werden kann. Dashboard 1 (DB 1) basiert auf Modellen und Entwürfen des Entwicklers, die in einer früheren Phase des Entwicklungsprozesses erstellt wurden. Im Laufe des Entwicklungsprozesses von DB 1 wurden potenzielle Schwächen des Prototyps identifiziert, die zusammen mit dem Feedback der Zielgruppe die Grundlage für die Entwicklung von Dashboard 2 (DB 2) bildeten. Für DB 2 wurde das Information-Seeking-Mantra nach Shneiderman (1996) berücksichtigt, das beschreibt, in welcher Reihenfolge dem User bzw. der Userin Interaktionsoptionen bereitgestellt werden. Shneidermans Design-Richtlinien folgend sollen Informationen zunächst in einer allgemeinen Übersicht dargestellt werden, welche dann durch gezieltes Filtern und Vergrößern die Darstellung sinnvoll gegliederter Teilinformationen erlaubt. Details werden zudem nur auf Anfrage dargestellt. Im vorliegenden Fall wurde zunächst eine aggregierte Übersicht anhand von Grafiken dargestellt und bei Bedarf relevante Details angezeigt. Dabei wurden Fotos der Lernenden ergänzt und zusätzliche Farben und Symbole eingesetzt, um Informationen zu codieren. Aufgrund der funktionalen Einschränkungen der Prototypen und der begrenzten Testzeit wurden zudem standardisierte Hilfestellungen entwickelt (wenn z. B. eine Testperson innerhalb eines definierten Zeitraums nicht in der Lage war, eine relevante Information zu finden). Es wurden jeweils drei unterschiedliche authentische Rechercheaufgaben präsentiert (Tabelle 1), die sämtliche Unterseiten der jeweiligen Prototypen abdecken (Riihiäho, 2017). Die Aufgaben waren unabhängig voneinander und wurden nacheinander angezeigt. Aufgrund der Bedeutung der Aufmerksamkeit für die Verhaltensregulierung zielten die Aufgaben jeweils auf unterschiedliche Aspekte ab (beispielsweise Fehleranalyse, Prozessfortschritt etc.). Die Formulierung wurde so gestaltet, dass sie eindeutig war, ohne mögliche Lösungen zu suggerieren (Riihiäho, 2017).

Tabelle 1: Zu bearbeitende Aufgabenstellungen je Dashboard

DB1	DB2
Welches Projektmodul bearbeitet Lerner 1 aktuell?	Welche Frage hat Lerner 1 im Fragebogen 1 falsch beantwortet?
Welche Frage hat Lerner 2 im Fragebogen 1 falsch beantwortet?	Welche Frage beantwortet Lerner 3 vermutlich gerade im Fragebogen 1?
Welcher Lernende hat die meisten Fragen des Fragebogen 1 beantwortet?	Welcher Lernende hat bei der Anpassung der Nutzwertanalyse im Projektmodul 1 Lieferantenauswahl 1 Batterie einen Fehler gemacht?

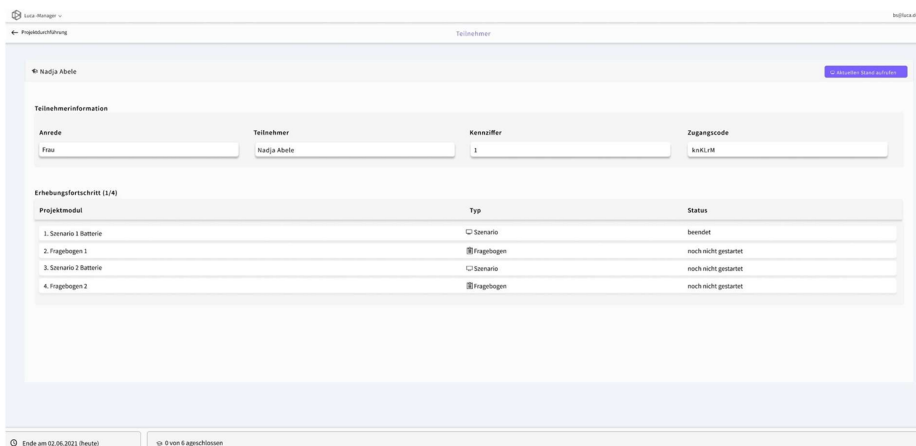


Abbildung 1: Dashboard-Variante 1

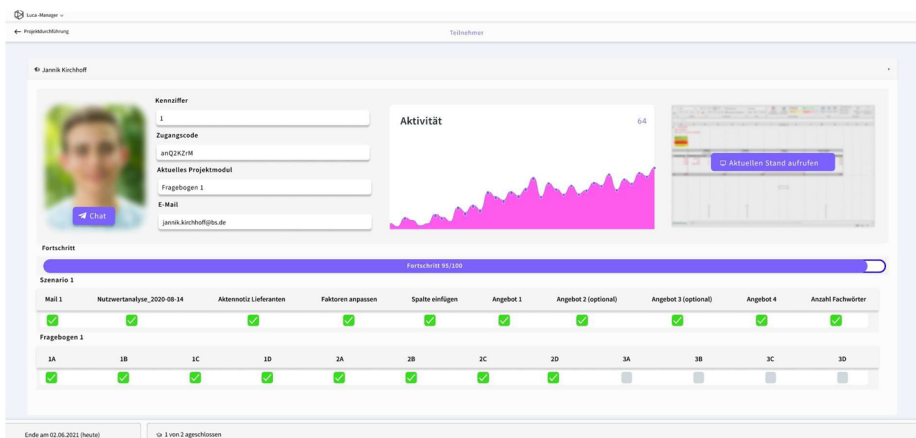


Abbildung 2: Dashboard-Variante 2

Die Usability-Testung mittels Eye-Tracking geht mit einigen Limitationen einher (Thoma & Dodd, 2019). Die Messung des Blickverhaltens liefert zwar umfangreiche Informationen; Analyse und Interpretation von Heatmap- und Eye-Tracking-Metriken können jedoch auch fehlerbehaftet sein. So kann eine längere Fixierung von Bereichen des Bildschirms einerseits auf tiefere Verarbeitungsprozesse hindeuten, andererseits ist aber auch Tagträumerei oder Überforderung eine denkbare Erklärung (Abele, Ostertag, Peissner & Schuller, 2017). Um über die Deskription von Verhaltensdaten hinauszukommen, wird die ergänzende Erfassung zusätzlicher Informationen (z. B. Log-Daten, Think-Aloud, Performance-Daten) empfohlen. Studien aus dem Bereich des computerbasierten Lernens verweisen diesbezüglich auf einen komplementären Charakter von Think-Aloud-Protokollen und Eye-Tracking-Daten (van Gog, Paas,

van Merriënboer & Witte, 2005; Stieff, Hegarty & Deslongchamps, 2011). Hier kann zwischen zeitgleichen und retrospektiven Think-Aloud-Protokollen (Concurrent bzw. Retrospective Think-Aloud: CTA bzw. RTA) unterschieden werden (Konrad, 2010). Die Nutzung der RTA-Variante führt zu einer geringeren kognitiven Belastung während des Versuchs (Thoma & Dodd, 2019), erhöht aber die Versuchsdauer für die Testperson. Angesichts der insgesamt moderaten Testdauer fiel die Entscheidung auf die RTA-Variante. Es bot sich an, bei der RTA-Variante den Testpersonen auch ihre Eye-Tracking-Daten zu zeigen (siehe auch Alhadreti, Elbabour & Mayhew, 2017). Hierzu wurde eine Bildschirmaufzeichnung herangezogen, die neben den Interaktionen mit der Maus auch den Blickverlauf wiedergibt (simultaner Einsatz von RTA und Eyetracking; gaze-cued RTA).

3.2 Messinstrumente

Befragungsdaten zu User Experience

Zur Erfassung der User Experience wurden Kurzskalen zur Messung des subjektiven Nutzererlebnisses (User Experience) eingesetzt (Schrepp, Hinderks & Thomaschewski, 2017a; 2017b). Dabei greifen wir auf eine Kurzform des User Experience Questionnaire (UEQ-S) zurück. Dieser eignet sich für die Testung von mehreren Anwendungen in kurzen zeitlichen Abständen und erfasst die pragmatische und hedonische Qualität der User Experience über je vier Items auf einer siebenstufigen Likert Skala (von -3 bis + 3). Dabei wird zwischen einer negativen und positiven Ausprägung unterschieden. Die Testpersonen wurden beispielsweise gefragt, inwiefern das Dashboard „verwirrend“ oder „übersichtlich“ (pragmatische Qualität) gestaltet ist bzw. inwiefern das jeweilige Dashboard als „langweilig“ oder „spannend“ (hedonische Qualität) erlebt wurde. Die Reliabilität der Kurzskalen erweist sich in den Studien von Schrepp et al. (2017b) als zufriedenstellend. Die Verwendung des UEQ bietet zudem den Vorteil, die User Experience über eine Benchmark (Schrepp, Hinderks & Thomaschewski, 2017c) vergleichen zu können.²

Blickbewegungsdaten

Die Blickbewegungsdaten wurden mit einem Gazepoint GP3 HD Eye-Tracker aufgezeichnet (150 Hz). Hierzu wurde die Software iMotions 9.1 genutzt und die Plattform Maze zur Präsentation der Click-Dummys herangezogen. Zur Auswertung wurden das Softwarepaket R (R Core Team, 2021) und die integrierte Entwicklungsumgebung R Studio (R Studio Team, 2021) sowie weitere Packages zur Datenaufbereitung und -analyse (insbesondere Tidyverse, siehe Wickham et al., 2019) verwendet. Jede einzelne Dashboard-Ansicht bildet eine dynamische Area of Interest (AOI) ab. So kann die visuelle Aufmerksamkeitssteuerung seitenspezifisch über Fixationen (das gezielte Betrachten eines Punktes) und Sakkaden (Bewegungen zwischen zwei Fixationspunk-

2 Für weiterführende Informationen siehe die Website des kostenfreien Instruments („User Experience Questionnaire UEQ“, 2021).

ten) erfasst werden (für eine Übersicht über Eye-Tracking-Metriken siehe Holmqvist et al., 2011; Holmqvist & Andersson, 2017).

Retrospective gaze-cued Think-Aloud (RTA)

Ergänzend wurde ein Retrospective (gaze-cued) Think-Aloud (RTA) zu beiden Dashboard-Anwendungen durchgeführt (Guan, Lee, Cuddihy & Ramey, 2006; van den Haak, De Jong & Schellens, 2003). Dabei diente ein in halber Geschwindigkeit abgespielter Screencast mit Maus- und Blickverhalten der Testpersonen als Stimulus bzw. Erinnerungsstütze für den kürzlich erlebten Bearbeitungsprozess. Die Auswertung erfolgt auf Basis eines umfassenden Codiermanuals unter Rückgriff auf die deduktiv-induktive Inhaltsanalyse nach Mayring (2016). Ein Auszug aus dem Codiermanual ist Tabelle 2 zu entnehmen.

Tabelle 2: Auszug aus dem Codiermanual zu (Usability-Problemen, Definitionen und Beispiele)

Kategorie	Definition	Ankerbeispiel
Layout	Die Versuchspersonen sind nicht in der Lage, relevante Informationen auf dem Bildschirm zu erkennen oder übersehen diese. Sie beschreiben ästhetische Probleme und weisen auf überflüssige Informationen hin.	„(...) ist alles sehr klein.“ (2-DB2)
Terminologie	Die verwendete Terminologie wird von den Versuchspersonen nicht verstanden.	„Was bedeutet Projektübersicht? Was kriege ich da?“ (5-DB1)
Feedback	Feedback, das die Versuchspersonen von der Benutzeroberfläche erhalten, ist nicht relevant oder stimmt nicht mit den Erwartungen der Nutzenden überein. Es wird ein Mangel an Feedback kritisiert.	„Oder wenn ich mit der Maus drüberfahre, dass es vielleicht auch nochmal irgendwie hervorhebt.“ (2-DB2)
Verständnis	Durch einen Mangel an Informationen sind die Versuchspersonen nicht fähig, das System effektiv zu erfassen und zu nutzen.	„Wie gesagt, mit den Balken bin ich nicht klargekommen, das habe ich auf die Schnelle jetzt nicht erkennen können, was die bedeuten.“ (2-DB1)
Navigation	Die Versuchspersonen äußern Probleme, im System zu navigieren.	„Was mich irritiert hat, ist, wenn ich bei dem einen Kandidaten zurückgesprungen bin, dass ich dann zwei Reiter zurückgesprungen bin“ (8-DB1)

Anmerkungen: Der erste Wert in Klammern identifiziert die Versuchsperson, während der zweite Wert das bearbeitete Dashboard kennzeichnet.

Das Codiermanual umfasst die Kategorien *Basiskriterien*, *Interaktionskriterien*, *Ein- und Ausgabekriterien* sowie *Usability-Problem-Kategorien*. Die Kategorien sind wie folgt operationalisiert:

Die Kategorie *Basiskriterien* umfasst die Kriterien *Effektivität*, *Effizienz* und *Zufriedenheit*. Informationen zu den Kriterien Effektivität und Effizienz werden nicht co-

diert, sondern direkt aus den Verhaltensdaten gewonnen.³ Zufriedenheit wird codiert, wenn Einstellungen oder Gefühle gegenüber den LADs genannt wurden. Diesbezüglich wird zwischen positiven und negativen Äußerungen unterschieden.

Bei der Kategorie *Interaktionsprinzipien* handelt es sich um Designprinzipien, deren Beachtung zu einer Verbesserung der Usability führen. Folgende Kriterien wurden hier berücksichtigt:

- *Aufgabenangemessenheit* gilt als gegeben, wenn das Dashboard als hilfreich oder unterstützend für die Bearbeitung der Aufgabe wahrgenommen wird. Dazu zählt u. a., dass überflüssige Interaktionen für die/den Nutzenden vermieden werden.
- *Selbstbeschreibungsfähigkeit* beurteilt die Repräsentation und Offensichtlichkeit von Elementen. Eine differenzierte Codierung erfolgt über die Ein- und Ausgabekriterien (s. u.).
- *Erwartungskonformität* beurteilt, inwiefern die Erwartungen der/des Nutzenden an die LADs und deren Funktionen erfüllt werden.
- *Erlernbarkeit* gilt als erfüllt, wenn die LADs Explorationen und Lernprozesse unterstützen, Lernaufwand reduziert und auf Funktionalitäten verwiesen wird.
- *Robustheit gegenüber Benutzerfehlern* evaluiert, wie die LADs auf Fehler der/des Nutzenden reagieren.
- *Benutzerbindung* bezieht sich auf Aspekte der Motivation und Vertrauenswürdigkeit, die sich im Umgang mit den LADs entwickeln.

Ein- und Ausgabekriterien beziehen sich auf die Interaktionsprinzipien und beschreiben Kriterien, wie Informationen dargestellt werden sollten.

- *Entdeckbarkeit* beurteilt, ob Elemente als vorhanden wahrgenommen werden.
- *Ablenkungsfreiheit* beschreibt einen Zustand ohne störende weitere Informationen.
- *Unterscheidbarkeit* liegt dann vor, wenn Elemente voneinander unterschieden werden können und anhand ihrer Eigenschaften zu Gruppen zugeordnet oder diskriminiert werden können.
- *Eindeutige Interpretierbarkeit* ist dann gegeben, wenn dargestellte Elemente wie vorgesehen verstanden werden.
- *Kompaktheit* beurteilt, ob lediglich relevante Informationen dargestellt werden.
- *Konsistenz* beurteilt die Einhaltung von Konventionen, d. h., ob vergleichbare Elemente ähnlich gestaltet sind.
- Das Kriterium *Zutraglichkeit zur Orientierung* besagt, dass ein System den Nutzenden eine umfassende Orientierung über Funktions- und Objektstrukturen ermöglicht.

Die Kategorien der Usability-Probleme beschreiben gängige Schwächen von entsprechenden Anwendungen (van den Haak et al., 2003; Olsen, Smolentzov & Strandvall, 2010).

³ *Effektivität* wurde über die Quote der erfolgreichen Aufgabenbearbeitung durch die Testpersonen erfasst. Das Kriterium der Effizienz wird mittels der zur Lösungsfindung notwendigen Zeit (time on task), der Anzahl an Klicks sowie der Anzahl an Wechseln zwischen den jeweiligen Ansichten der Dashboards erfasst.

4 Empirische Befunde

4.1 User-Experience-Befragung (Zufriedenheit)

Die UX für beide Dashboards wird anhand des UEQ-S für die pragmatische und hedonische Qualität ermittelt (siehe Abbildung 3). Die pragmatische Qualität von DB 1 wird mit einem Mittelwert von $-0,125$ und die hedonische Qualität mit $0,188$ bewertet. Auf einer Skala von -3 bis $+3$ bewegen sich die Werte im neutralen Bereich. Zieht man die Benchmark (s. o.) heran, so zeigt sich, dass die UX für beide Aspekte nicht zufriedenstellend ausgeprägt ist. Bei DB 2 liegt die pragmatische Qualität bei einem Mittelwert von $0,688$ und die hedonische Qualität bei einem Wert von $1,156$. Der Benchmark-Vergleich verweist erneut auf ein nicht zufriedenstellendes Ergebnis.

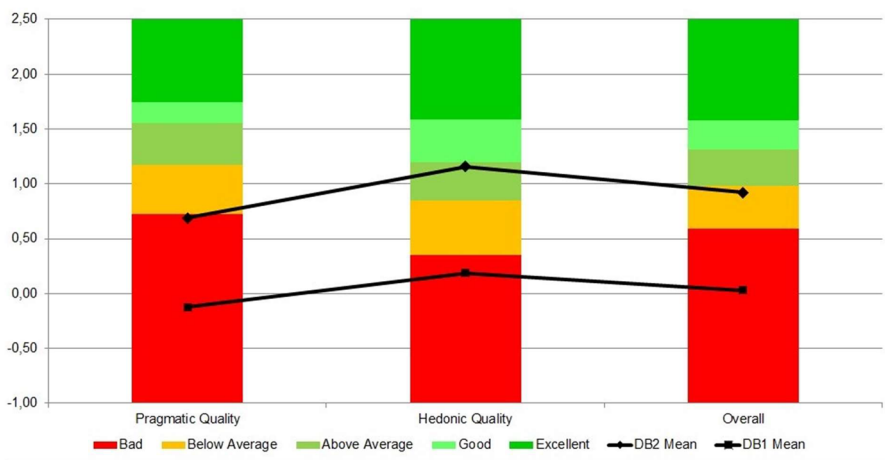


Abbildung 3: Pragmatische und hedonische Qualität (Mittelwerte, UEQ-S Fragebogen, $n = 8$)

4.2 Blickbewegungsdaten

Ergänzend zur Zufriedenheit (s. o.) werden zur Analyse der Ergebnisse der Blickbewegungsdaten die Merkmale Effektivität und Effizienz herangezogen:

- Bezüglich der Effektivität (Anteil erfolgreicher Aufgabenbearbeitung) lässt sich festhalten, dass sieben der acht Teilnehmenden die gestellten Aufgaben in beiden Dashboards lösen konnten. Allerdings musste im DB 1 in sieben der acht Fälle eine Hilfestellung gegeben werden. Bei DB 2 dagegen wurden keine Hilfestellungen benötigt.
- Die Effizienz der Dashboards wurde mittels der Kriterien time on task, Anzahl an Klicks und Anzahl an Wechseln zwischen den jeweiligen Ansichten der Dashboards erfasst. Die Bearbeitungszeiten für beide Prototypen sind Tabelle 3 zu entnehmen. Die erste von drei Aufgaben diente als Orientierungsaufgabe, die den Teilnehmenden dabei helfen sollte, sich mit dem Dashboard vertraut zu machen. Diese Orientierungsaufgabe verzerrt jedoch das quantitative Ergebnis und wird

daher in dieser Betrachtung (nicht jedoch in der qualitativen Auswertung des Think-Aloud) ausgeschlossen. Die durchschnittliche Bearbeitungszeit der beiden weiteren Aufgaben je Dashboard beträgt für den ersten Prototypen 128 Sekunden ($SD=76$ Sek.) und für den zweiten Prototypen 117 Sekunden ($SD=87$ Sek.). Die Anzahl an durchschnittlichen Klicks zwischen Dashboard 1 und Dashboard 2 ist nahezu gleich (33,99 vs. 36,31). Ein geringer Unterschied lässt sich bezüglich der aggregierten Anzahl an Transitionen (Wechsel zwischen Unterseiten) der Dashboards feststellen (164 vs. 125 Transitionen).

Tabelle 3: Bearbeitungszeiten je Aufgabe (time on task)

Prototyp	Anzahl an Klicks	Anzahl an Transitionen	Aufgabe	M Bearbeitungs-dauer (ms)	SD Bearbeitungs-dauer (ms)
Dashboard 1	33,99	164	Task 1	215625.50	77913.71
			Task 2	47311.00	31269.55
			Task 3	80772.25	44855.68
Dashboard 2	36,31	125	Task 1	53027.25	29897.07
			Task 2	33522.75	41068.13
			Task 3	83169.00	45862.07

Anmerkung: $N=8$ Lehrkräfte ($M=40.25$, $SD=9.9$). *Anzahl Klicks* = Durchschnittliche Anzahl an Mausklicks pro Dashboard. *Anzahl an Transitionen* = Aggregierte Anzahl an Wechslen zwischen den Unterseiten. *Bearbeitungs-dauer* = Durchschnittliche Dauer der Bearbeitung je Task in ms.

Tabelle 4 zeigt die Eye-Tracking-Daten je Dashboard.⁴ Die Datenqualität pro Teilnehmenden bzw. Teilnehmender und Dashboard-Seite liegt bei 92 bis 99 %, d. h., für bis zu maximal 8 % der AOI wurden die Augenbewegungen nicht erfasst. Dieser Datenverlust kann auf natürliche Ursachen wie Blinzeln zurückgeführt werden und ist noch akzeptabel (Holmqvist et al., 2011; Holmqvist & Andersson, 2017). Beide Prototypen bestehen aus 13 Ansichten. Alle Unterseiten lassen sich (im Sinne eines Problemlöseprozesses für Lehrkräfte) in für die Aufgabenstellung relevante Attraktoren bzw. irrelevante Distraktoren unterscheiden. Zu den Attraktoren (fett markiert in Tabelle 4) zählen folgende Ansichten: Dashboard-Start, Projektübersicht, Szenario 1, Fragebogen 1 bzw. Gruppenübersicht 1 und potenziell die Ansicht der einzelnen Schülerinnen und Schüler (hier sind die für die Lösung der Aufgabe relevanten Informationen häufig ebenfalls auffindbar). Weitere Ansichten sind Distraktoren im Sinne der gestellten Aufgaben.

Für DB 1 und 2 zeigen sich keine wesentlichen Abweichungen hinsichtlich der Aufrufe von Attraktoren und Distraktoren. Abweichungen sind in der durchschnittlichen Besuchszeit auf zentralen Seiten wie der Projektübersicht (34 Sek. vs. 16 Sek.), der Unteransicht Fragebogen 1 bzw. Gruppenübersicht Aufgabe 1 (20 Sek. vs. 8 Sek.) zu

4 Nachfolgend berichtete Werte sind von Millisekunden auf Sekunden übertragen worden und gerundet.

finden. Geringere Unterschiede lassen sich auch bezüglich der Ansichten „Szenario 1“ (11 Sek. vs. 16 Sek.) und jener des relevanten Schülers bzw. der relevanten Schülerin (6 Sek. vs. 8 Sek.) feststellen. Die Projektübersicht in DB 2 wird wesentlich seltener wiederholt angesteuert (6 vs. 2.5 Revisits, andere Attraktoren sind dagegen nahezu gleich häufig ausgeprägt). Wesentliche Unterschiede zeigen sich in der durchschnittlichen Anzahl und Dauer der Fixationen⁵ auf den Seiten der Projektübersicht (72 vs. 39 Fixationen, 9 Sek. vs. 5 Sek. mittlere Fixationsdauer) und der Sicht des Fragebogens bzw. der Gruppenübersicht (62 vs. 22 Fixationen, 5 Sek. vs. 3 Sek. mittlere Fixationsdauer), bei sonst erneut ähnlichen Werten für die Ansichten Szenario und Schülerinnen und Schüler. Dieser Befund findet sich in der Anzahl an Sakkaden für die Projektübersicht (426 vs. 190 Sakkaden) und der Sicht des Fragebogens bzw. der Gruppenübersicht (245 vs. 98 Sakkaden) entsprechend wieder.

Tabelle 4: Eye-Tracking Daten je Dashboard

Dashboard	Durchschnittliche Besuchszeit (ms)	Aufrufe in Relation der Tn. (%)	Anzahl an Wiederbesuchen	Anzahl an Fixationen	Fixationsdauer (ms)	Anzahl an Sakkaden	Anzahl an Klicks
<i>Dashboard 1</i>							
Start	21510.46	100.00	2.00	50.62	5984.18	268.25	5.00
Projektübersicht	33777.75	100.00	6.00	72.88	8714.77	426.38	8.12
SuS-Screen live	1686.19	25.00	2.00	11.00	1444.84	78.50	1.50
Fragebogen 1	20184.84	100.00	1.62	62.29	8000.87	245.12	3.00
Fragebogen 2	854.21	25.00	0.00	9.00	1098.73	37.50	1.00
Szenario 1	11341.75	87.50	1.86	35.29	4541.88	160.86	2.71
Szenario 2	1500.80	37.50	0.33	16.50	1830.12	51.00	1.33
SuS Dominik	626.59	12.50	1.00	11.00	1210.80	63.00	3.00
SuS Leo	2844.08	37.50	0.33	15.33	1990.17	98.67	2.00
SuS Nadja	7418.47	50.00	1.25	30.00	3605.52	180.00	2.75
SuS Roman	5859.57	75.00	2.00	13.83	1690.54	87.50	1.83
SuS Theresa	0.00	0.00	NA	NA	NA	NA	NA
SuS Thomas	5262.74	50.00	0.75	21.25	2120.27	128.75	1.75

⁵ Fixation gemessen ab 100 ms; vgl. Rayner, 1998

(Fortsetzung Tabelle 4)

Dashboard	Durchschnittliche Besuchszeit (ms)	Aufrufe in Relation der Tn. (%)	Anzahl an Wiederbesuchen	Anzahl an Fixationen	Fixationsdauer (ms)	Anzahl an Sakkaden	Anzahl an Klicks
<i>Dashboard 2</i>							
Start	7460.35	100.00	1.50	16.25	1984.09	92.75	2.25
Projektübersicht	15673.66	100.00	2.50	38.75	4808.87	189.88	4.38
SuS-Screen live	0.00	0.00	NA	NA	NA	NA	NA
Gruppenübersicht Fragebogen	4775.52	12.50	7.00	109.00	13599.13	479.00	7.00
Gruppenübersicht Aufgabe 1	8086.28	100	1.12	21.62	2931.81	97.62	2.25
Gruppenübersicht Aufgabe 2	14629.66	50.00	2.50	84.50	10661.92	358.00	3.00
Gruppenübersicht Aufgabe 3	4004.14	37.50	1.33	25.67	3326.05	134.67	3.00
Fragebogen 1	4040.41	50.00	0.75	24.50	2584.76	112.25	2.00
Szenario 1	15716.59	100.00	1.75	48.38	6263.85	195.38	3.00
SuS Alex	1919.76	12.50	1.00	30.00	3744.36	203.00	3.00
SuS Anna-Lena	859.82	12.50	0.00	9.00	935.28	102.00	1.00
SuS Jannik	8103.83	87.50	0.43	24.71	3183.20	113.29	1.43
SuS Tanja	2372.08	12.50	0.00	35.00	4755.56	227.00	4.00

Anmerkung: N = 8 Lehrkräfte ($M = 40.25$, $SD = 9.9$). *Dashboard* = Prototyp bzw. Unterseite. *Durchschnittliche Besuchszeit* pro Ansicht in ms. *Aufrufe in Relation der Tn.* = Relative Anzahl an Personen, die diese Sicht aufgerufen haben. *Anzahl an Wiederbesuchen* = Durchschnittliche absolute Häufigkeit, mit der die Sicht erneut aufgerufen wurde. *Anzahl an Fixationen* = Durchschnittliche absolute Anzahl an Fixationen (100ms). *Fixationsdauer* = Durchschnittliche Fixationsdauer in ms. *Anzahl an Sakkaden* = Durchschnittliche Anzahl an Sakkaden, die innerhalb der Seite erkannt wurden. *Anzahl an Klicks* = Durchschnittliche Anzahl der Mausklicks, die innerhalb der Seite erkannt wurden. NA steht für Not available (nicht verfügbar).

Die Probandinnen und Probanden berichteten insbesondere von Schwierigkeiten bei der Navigation und beim Erkennen relevanter Informationen. Die Zeit bis zur ersten Fixation der Navigation (time to first fixation: TTFF), also die durchschnittliche Zeit, die verging, bis die erste Fixierung innerhalb der AOI erkannt wurde (relativ zu Beginn der AOI), wird berechnet, um die Salienz der Navigationspfeile zu bestimmen (Thoma & Dodd, 2019). Die TTFF der Navigationspfeile konnte durchschnittlich um 32 Sekunden (77 Sek. vs. 45 Sek.) reduziert werden. Die wiederkehrende Betrachtung relevanter Inhalte dient als Proxy zur Bestimmung der kognitiven Verarbeitung der

Information (Thoma & Dodd, 2019). Die Anzahl an wiederholten Betrachtungen (Re-visits) konnte ebenfalls leicht reduziert werden (5,17 vs. 3,49).

4.3 Identifizierte Usability-Probleme aus RTA

Tabelle 5 zeigt die aggregierten Häufigkeiten der aufgetretenen Usability-Probleme. Insgesamt wurden im Rahmen der RTAs bei acht Testpersonen bzw. 16 Protokollen insgesamt 445 Codiereinheiten gesetzt. Davon entfallen 136 Codierungen auf vier Hauptkategorien (Layoutprobleme, Verständnisprobleme, Navigationsprobleme, sonstige Probleme, siehe Tabelle 5). Die weiteren 309 Codierungen beziehen sich auf Aspekte der Usability im weiteren Sinne (Zufriedenheit, Interaktionsprinzipien, Ein- und Ausgabekriterien, siehe Tabelle 6). In beiden Dashboards wurde in der Kategorie Layoutprobleme die Darstellung einzelner relevanter Elemente oder der gesamten Ansicht als zu klein beschrieben. Ebenso wurde bemängelt, dass Überschriften und Oberbegriffe nicht offensichtlich genug gestaltet sind. In der Kategorie Verständnisprobleme wurde das Fehlen von Informationen zur Orientierung am häufigsten genannt. Bei DB 2 wurde zudem oft angemerkt, dass Angaben zur Interpretation von Indikatoren fehlen. Bei DB 1 wurden in der Kategorie Navigationsprobleme besonders häufig Elemente genannt, deren Bezeichnung auf die gewünschte Seite hindeutet, aber entweder keine (13 Nennungen) oder nicht die gewünschte (11 Nennungen) Funktionalität aufwiesen. Zusätzlich wurde die indirekte Navigation bemängelt. Es wurde kritisiert, dass Seiten nicht direkt miteinander verbunden sind, sondern nur über andere Seiten angesteuert werden können (8 Nennungen). Vergleichbare Navigationsprobleme traten in geringerem Umfang bei DB 2 auf (10 Nennungen).

Tabelle 5: Usability-Probleme je Dashboard (aggregierte Darstellung)

Problemkategorien	DB1	DB2	Zitat
Layoutprobleme	24	23	„Dann kann ich das auch besser lesen, weil es doch sehr klein war“ (2-DB2)
Verständnisprobleme	22	15	„Ich hätte mir (...) eine klare Kennzeichnung [gewünscht], was auf den folgenden Seiten ist, nicht nur der kleine Pfeil.“ (1-DB1)
Navigationsprobleme	34	9	„(...), dass man gezielt nach einem Projekt suchen kann, ohne sich durchklicken zu müssen und durch die verschiedenen Seiten.“ (6-DB2)
Sonstige Usability-Probleme	3	6	„Ehrlich gesagt, auch die Begriffe finde ich nicht ganz eindeutig.“ (2-DB1)

Anmerkungen: Der erste Wert in Klammern identifiziert die Versuchsperson, während der zweite Wert das bearbeitete Dashboard kennzeichnet.

Tabelle 6: Codierungen der weiteren Kategorien: Zufriedenheit, Interaktionskriterien, Ein- und Ausgabekriterien (aggregierte Darstellung)

Kategorie	DB 1	DB 2	Zitat
<i>Aussagen zur Zufriedenheit</i>			
positive Aussagen	9	17	„Aber so ein Dashboard ist grundsätzlich hilfreich und wünschenswert.“ (2-DB1)
negative Aussagen	8	8	„Problematisch waren halt immer diese farblichen Absetzungen und dass alles so klein war“ (3-DB1)
<i>Interaktionsprinzipien (positive Aussagen)</i>			
Aufgabenangemessenheit	4	13	„Aber es ist wie eine Karteikarte, und ehrlich gesagt, für uns Lehrende ist es eine Hilfe (...) deswegen fände ich ein Photo immer ganz schön drinzuhaben“ (2-DB2)
Sonstige	1	3	„Es war natürlich auch motivierender und individueller, wenn ich da Bildchen sehe vom Schüler.“ (3-DB2)
<i>Interaktionsprinzipien (negative Aussagen)</i>			
Aufgabenangemessenheit	12	6	„(...), weil ich kann daraus vielleicht nicht unbedingt den Erfolg, die Information rausfiltern, die ich jetzt als Lehrer unbedingt brauche im Dashboard.“ (6-DB2)
Erwartungskonformität	8	4	„Und hier bin ich ein bisschen gefangen und weiß nicht, wohin ich gehen soll, weil da ist ja der Fragebogen, aber ich kann nicht draufklicken“ (3-DB1)
Sonstige	9	2	„Jetzt bin ich aber wieder nicht in der in der Fragebogen Übersicht, sondern in einer anderen Überübersicht. Also nochmal 2 Schritte zurück. Das wäre natürlich wünschenswert, wenn ich wieder in diese ... also wieder zu den Teilnehmern zurück switche“ (3-DB2)
<i>Ein- und Ausgabekriterien (positive Aussagen)</i>			
Eindeutige Interpretierbarkeit	11	17	„und da war es dann schnell ersichtlich, dass der Alex Scheibel einen Fehler gemacht hat und dementsprechend konnte ich das dann anklicken.“ (4-DB2)
Orientierungsförderlichkeit	5	6	„Aber dann doch deutlich schneller und zügiger als beim ersten Dashboard den Aufbau erkannt und die Vorgehensweise erkannt und auch deutlich besser dann in das jeweilige Projekt reingekommen“ (4-DB2)
Unterscheidbarkeit	4	3	„Ja, da mit richtig und falsch, das hat man natürlich dann gleich gesehen. Logisch.“ (2-DB2)
Sonstige	3	5	„Hier ist schön die Haken, dass man es gleich sieht, was da gelaufen ist.“ (2-DB2)
<i>Ein- und Ausgabekriterien (negative Aussagen)</i>			
Orientierungsförderlichkeit	47	14	„Da finde ich es praktischer, wenn die Pfeile gekennzeichnet wären. Also mit einer Beschriftung, wo komme ich da hin, wenn ich draufklicke (...)“ (1-DB1)
Entdeckbarkeit	17	10	„Ok, die Pfeile müssten, glaube ich, etwas größer sein, damit es offensichtlicher ist. (...)“ (1-DB1)

(Fortsetzung Tabelle 6)

Kategorie	DB 1	DB 2	Zitat
Eindeutige Interpretierbarkeit	16	8	„Also so ein Aktivitätsmodul vielleicht zu zeigen ist ganz gut, aber man sollte wissen, was es bedeutet.“ (2-DB2)
Unterscheidbarkeit	14	8	„(...), wenn es so wäre, dass die Zeilen farblich ein bisschen abgesetzt wären. Reicht ja jede zweite Zeile, dass man sich da entsprechend besser in der Zeile bewegen kann, ohne zu verrutschen mit dem Auge.“ (2-DB2)
Sonstige	8	9	„Also da haben die Pfeile rechts und links nicht die gleiche Funktion.“ (2-DB2)

Anmerkungen: Der erste Wert in Klammern identifiziert die Versuchsperson, während der zweite Wert das bearbeitete Dashboard kennzeichnet.

Die Äußerungen zur Zufriedenheit in DB 1 beziehen sich auf Aspekte der Darstellung und Funktionalität (5 bzw. 2 Nennungen). Negative Äußerungen (8 Aussagen von 5 Probandinnen und Probanden) betreffen zumeist die Aspekte der Orientierung (4Aussagen) und Darstellung (3 Aussagen). Bei DB 2 wurden 18 positive Segmente zur Zufriedenheit codiert. Am häufigsten wurden Aspekte der Darstellung positiv erwähnt (12). Negative Aussagen bezogen sich auf die Darstellung (3 Nennungen), Orientierung (3 Nennungen) sowie Navigation (2 Nennungen). Hinsichtlich der Interaktionsprinzipien findet sich bezüglich der Aufgabenangemessenheit für DB 1 negatives Feedback durch die Testpersonen, welches sich zumeist auf den Aspekt der indirekten Navigation (7 Aussagen) bezieht. Für DB 2 beziehen sich die meisten positiven Aussagen (7 Nennungen) auf Facetten der Darstellung.

5 Diskussion

In diesem Beitrag berichten wir Ergebnisse zur Usability des Learning Analytics Dashboards der digitalen Bürosimulation LUCA. Hierzu wurden zwei Dashboard-Entwürfe als Prototypen mittels Eye-Tracking durch acht Lehrkräfte getestet. Anschließend gaben die Lehrkräfte eine Einschätzung zur Usability ab. Mittels Cued-Retrospective-Think-Alouds konnte das Handling der Dashboards näher analysiert werden. Zunächst wurde hierzu mittels eines Fragebogens (UEQ-S) die User Experience erfasst. Der UEQ-S gibt Rückmeldung über die hedonische und pragmatische Qualität der Interaktion. DB 1 schneidet – insbesondere im Vergleich zu den Benchmark-Werten – weniger gut ab. Insbesondere die pragmatischen, aufgabenbezogenen Aspekte werden schlecht bewertet. Dies weist auf Schwächen bezüglich der Benutzerfreundlichkeit hin. DB 2 wurde bezüglich der pragmatischen Qualität positiver bewertet. Im Vergleich zur Benchmark zeigen sich jedoch auch hier noch Verbesserungspotenziale. Die hedonische Qualität, die sich auf Aspekte wie Freude oder Vergnügen an der Interaktion bezieht, wird für beide Dashboard-Varianten positiver bewertet als die pragmatische Qualität. Dies gilt insbesondere für DB 2. Aus den Ergebnissen des

UEQ-S und dem Vergleich mit dem Benchmark lässt sich schließen, dass DB 2 eine Verbesserung des Designs des LUCA-Dashboards in Bezug auf die UX darstellt. Beide Prototypen weisen in Bezug auf die pragmatische Qualität zwar Unzulänglichkeiten auf, diese können jedoch bis zu einem gewissen Grad auch auf die eingeschränkten Funktionalitäten der Prototypen zurückgeführt werden. Insgesamt zeigt sich, dass beide Varianten eine hinreichend gute Effektivität aufweisen – die Aufgaben in beiden Dashboards waren jeweils gut zu lösen. Hinsichtlich der Effizienz weist DB 2 (als Weiterentwicklung) Vorteile gegenüber DB 1 auf. Hier konnten die Lösungen schnell gefunden werden (geringe Bearbeitungsdauer) und auch die Navigation gestaltete sich vergleichsweise einfach (geringe Anzahl von Transitionen als Indikator für effiziente Navigation, siehe Jacob & Karn, 2003). Allerdings fallen die Unterschiede hier eher gering aus. Ein anderes Bild ergibt sich bei der Analyse der Eye-Tracking-Daten. Insgesamt verweisen die Daten auf spürbare Vorteile von DB 2 gegenüber DB 1, z. B. hinsichtlich der Verweildauer auf relevanten Unterseiten (Attraktoren im Sinne der Aufgabenstellung) oder der Anzahl der wiederkehrenden (redundanten) Besuche der zentralen Projektübersicht. Eine hohe Anzahl an wiederkehrenden Besuchen kann auch durch die Interessantheit der Inhalte hervorgerufen werden (Thoma & Dodd, 2019). Im vorliegenden Fall werten wir dies jedoch eher als ein Anzeichen für Verständnisprobleme. Diese Annahme wird durch die Daten der Analyse der grundlegenden Augenbewegungsmuster gestützt. Sowohl Anzahl als auch Dauer der Fixationen fallen für DB 2 deutlich geringer aus (zur Interpretation dieses Werts siehe auch Abele et al., 2017; Goldberg & Kotval, 1999; Pannasch, 2003; Thoma & Dodd, 2019). Abele et al. (2017, S. 104) verweisen aus Sicht der domänenspezifischen Problemlöseforschung auf teilweise ungewöhnlich „hohe kritische Fixationsdauern“. Damit ist der Übergang von einer tiefen kognitiven Verarbeitung zum Zustand der Konfusion gemeint. Just und Carpenter (1980) gehen unter Rückgriff auf Ergebnisse der Leseforschung von einer unmittelbaren und linearen Abhängigkeit von Fixationsdauer und kognitiver Verarbeitung aus (Eye-Mind-Assumption). Allerdings ist diese Vermutung empirisch nicht durchgängig belegt (Anderson, Bothell & Douglass, 2004; Underwood & Everatt, 1992). Die Annahme einer grundlegenden Verbindung zwischen visueller Aufmerksamkeitssteuerung und kognitiver Verarbeitung ist unumstritten (für eine domänenspezifische Analyse siehe Schindler & Lilienthal, 2019). Ähnlich ist die Anzahl der Sakkaden zu interpretieren (Notwendigkeit der Informationssuche vs. Wunsch, sich vertiefter mit den Inhalten auseinanderzusetzen; siehe Thoma & Dodd, 2019).

Alles in allem erweist sich das Design von DB 2 im Vergleich zu DB 1 somit als vorteilhaft. Ungeachtet dessen sind weitere Verbesserungen des Dashboards notwendig. Hinweise zur weiteren Verbesserung des Dashboards lassen sich aus Aussagen der Testpersonen gewinnen. Ein durchgängiges Problem beider Dashboards war das Layout. Offenbar waren viele Objekte nicht ausreichend wahrnehmbar. Folgendes Zitat einer Testperson unterstreicht dies: „Ich habe lange z. B. nach dem Wort Batterie gesucht, auch wenn es im Endeffekt oben schon dastand. Aber es ging für den ersten Blick so ein bisschen unter, weil ich es mir gewünscht hätte, dass es vielleicht irgend-

wie ein bisschen dicker hervorgehoben ist.“ Darüber hinaus werden fehlende Informationen zur Interpretation von Indikatoren bemängelt. Für die finale Gestaltung des Dashboards wurden diese Hinweise entsprechend berücksichtigt (siehe Abbildung 4).⁶

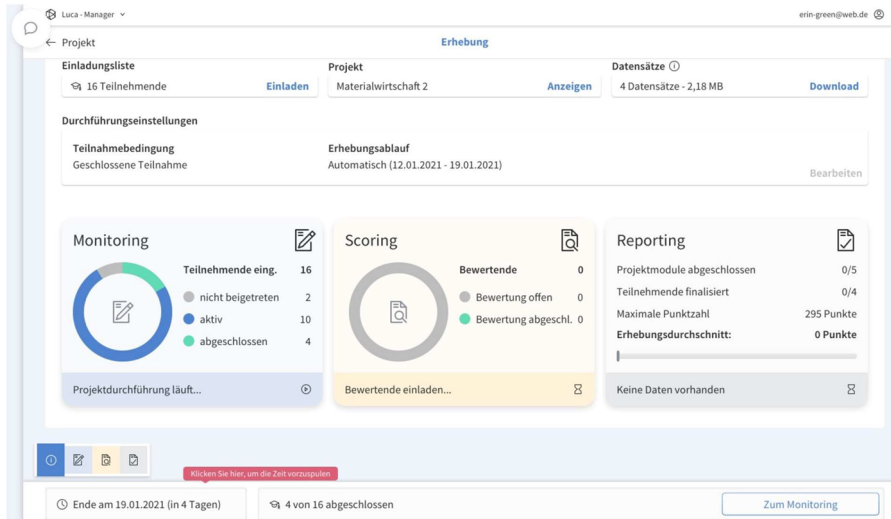


Abbildung 4: Startansicht des finalen Dashboards mit farblicher Trennung und Shortcut-Navigation zu den jeweiligen Grundfunktionen (Entwurf)

Bezüglich des methodischen Designs dieser Studie war es von Bedeutung, dass die verwendeten Aufgaben vergleichbarer Natur sind, da die Aufgabenstellung das Augenbewegungsverhalten maßgeblich beeinflusst (Yarbus, 1967; siehe auch Tatler, Wade, Kwan, Findlay & Velichkovsky, 2010). Zudem ist es wichtig, einen ökologisch validen Test zur Verfügung zu stellen. Abschließend ist davon auszugehen, dass der Rückgriff auf Eye-Tracking als Stütze der retrospektiven Think-Alouds die Validität positiv beeinflusst (Bojko, 2013; Thoma & Dodd, 2019, siehe auch van Gog et al., 2005; Olsen et al., 2010; Alhadreti et al., 2017; Elling, Lentz & de Jong, 2011).

Die vorgelegte Studie geht – wie jede Untersuchung – mit einigen Limitationen einher. So konnte lediglich ein Convenience-Sample rekrutiert werden. Zudem erfolgte die Teilnahme auf freiwilliger Basis, sodass ein Selection-Bias nicht ausgeschlossen werden kann. Zudem ist der Stichprobenumfang mit lediglich acht Testpersonen zwar noch als ausreichend für die Bearbeitung der Forschungsfrage zu bewerten, eine höhere Anzahl an Probandinnen und Probanden wäre jedoch insbesondere in Bezug auf den UEQ-S-Fragebogen wünschenswert gewesen. Aufgrund technischer Restriktionen konnte zudem eine randomisierte Zuweisung der Dashboards nicht stattfinden und somit ist ein Trainingseffekt nicht auszuschließen. Allerdings spricht wenig für solch einen Effekt, da Aufgaben und Aufbau beider Dash-

6 Mit freundlicher Genehmigung von Cap3

boards nicht übereinstimmen. Eine weitere Limitation ist darin zu sehen, dass wir zu dem Untersuchungszeitpunkt noch kein reales Dashboard, sondern lediglich Prototypen testen konnten. Ungeachtet dieser Einschränkungen förderte die Analyse einige wichtige Hinweise zum Dashboard-Design zutage, die wir für die Umsetzung des realen Dashboards in der LUCA-Office-Simulation sehr gut nutzen konnten.

Literaturverzeichnis

- Abele, S., Ostertag, R., Peissner, M. & Schuller, A. (2017). Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess. Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113(1), 86–109.
- Alhadreti, O., Elbabour, F. & Mayhew, P. (2017). Eye tracking in retrospective think-aloud usability testing: Is there added value? *Journal of Usability Studies*, 12(3), 95–110.
- Ali, L., Hatala, M., Gašević, D. & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1), 470–489. doi: 10.1016/j.compedu.2011.08.030
- Anderson, J. R., Bothell, D. & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science*, 15(4), 225–231.
- Bojko, A. (2013). *Eye tracking the user experience: a practical guide to research*. Brooklyn, New York: Rosenfeld Media.
- Cap3 (2021). LUCA-Office. *Cap3 Software*. Verfügbar unter <https://www.cap3.de/blog/luca> (Zugriff am: 14.11.2021).
- Davis, F. (1985). *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results* (Doctoral dissertation, Cambridge: Massachusetts Institute of Technology). Verfügbar unter <http://hdl.handle.net/1721.1/15192> (Zugriff am: 10.06.2022).
- Elling, S., Lentz, L. & de Jong, M. (2011). Retrospective think-aloud method: using eye movements as an extra cue for participants' verbalizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1161–1170. doi: 10.1145/1978942.1979116
- Glahn, C. (2009). *Contextual support of social engagement and reflection on the Web*. Heerlen: Open University of the Netherlands.
- Goldberg, J. H. & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24(6), 631–645. doi: 10.1016/S0169-8141(98)00068-7
- Guan, Z., Lee, S., Cuddihy, E. & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 1253–1262.

- Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product (Human-Computer Interaction Series). In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.), *Funology* (Vol. 3, pp. 31–42). Dordrecht: Springer Netherlands. doi: 10.1007/1-4020-2967-5_4
- Herczeg, M. (2018). *Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme*. De Gruyter: Oldenbourg. doi: 10.1515/9783110446869
- Holmqvist, K. & Andersson, R. (2017). *Eye tracking: a comprehensive guide to methods, paradigms, and measures* (2nd ed.). Charleston: CreateSpace.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- ISO 9241–11 (2018). *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. Retrieved June 10, 2022 from <https://www.iso.org/standard/63500.html>
- ISO 9241–110 (2020). *Ergonomics of human-system interaction – Part 110: Interaction principles*. Retrieved June 10, 2022 from <https://www.iso.org/standard/75258.html>
- ISO 9241–112 (2017). *Ergonomie der Mensch-System-Interaktion – Teil 112: Grundsätze der Informationsdarstellung*. Verfügbar unter <https://www.iso.org/standard/64840.html> (Zugriff am 10.06.2022).
- Jacob, R. J. K. & Karn, K. S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The Mind's Eye Cognitive and Applied Aspects of Eye Movement Research* (pp. 573–605). Amsterdam: Elsevier. doi: 10.1016/B978-0-444-51020-4/50031-1
- Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329–354.
- Klein, C. & Ettinger, U. (2019). *Eye movement research: an introduction to its scientific foundations and applications* (Studies in Neuroscience, Psychology and Behavioral Economics). Cham: Springer Nature.
- Klerkx, J., Verbert, K. & Duval, E. (2017). Learning Analytics Dashboards. In C. Lang, G. Siemens, A. F. Wise & D. Gašević (Eds.), *The Handbook of Learning Analytics* (pp. 143–150). Alberta: Society for Learning Analytics Research (SoLAR).
- Konrad, K. (2010). Lautes Denken. In G. Mey & K. Mruck (Hg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 476–490). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-92052-8_34
- Kuhnel, M., Seiler, L., Honal, A. & Ifenthaler, D. (2018). Mobile learning analytics in higher education: usability testing and evaluation of an app prototype. *Interactive Technology and Smart Education*, 15(4), 332–347. doi: 10.1108/ITSE-04-2018-0024
- Kuniavsky, M. & Moed, A. (2012). *Observing the User Experience a Practitioner's Guide to User Research*. San Diego: Elsevier Science & Technology Books.
- Manakhov, P. & Ivanov, V. D. (2016). Defining Usability Problems. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3144–3151. doi: 10.1145/2851581.2892387
- Mayring, P. (2016). *Einführung in die qualitative Sozialforschung*. Weinheim: Beltz.

- Olmsted-Hawala, E., Holland, T. & Quach, V. (2014). Usability Testing. In J. R. Bergstrom & A. J. Schall (Eds.), *Eye Tracking in User Experience Design* (pp. 49–80). doi: 10.1016/B978-0-12-408138-3.00003-0
- Olsen, A., Smolentzov, L. & Strandvall, T. (2010). Comparing different eye tracking cues when using the retrospective think aloud method in usability testing. *Proceedings of HCI 2010*, 24, 45–53. doi: 10.14236/ewic/HCI2010.8
- Pannasch, S. (2003). *Ereignisbezogene Veränderungen der visuellen Fixationsdauer* (Dissertation). Dresden: TU Dresden.
- Park, Y. & Jo, I.-H. (2015). Development of the Learning Analytics Dashboard to Support Students' Learning Performance. *Journal of Universal Computer Science*, 21(1), 110–133.
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S. & Winther, E. (2021). Die web-basierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick. *Zeitschrift für Berufs und Wirtschaftspädagogik*, 117(3), 372–394. doi: 10.25162/zbw-2021-0017
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372–422.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved June 12, 2022 from <https://www.R-project.org/>
- Riihiäho, S. (2017). Usability Testing. In K. L. Norman & J. Kirakowski (Eds.), *The Wiley Handbook of Human Computer Interaction* (pp. 255–275). Chichester: John Wiley & Sons. doi: 10.1002/9781118976005.ch14
- R Studio Team. (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Retrieved June 12, 2022 from <http://www.rstudio.com/>
- Sangmeister, J., Winther, E., Deutscher, V., Bley, S., Kreuzer, C. & Weber, S. (2018). Designing Competence Assessment in VET for a Digital Future. In D. Ifenthaler (Ed.), *Digital Workplace Learning* (pp. 65–92). Cham: Springer International Publishing. doi: 10.1007/978-3-319-46215-8_5
- Schindler, M. & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, 101(1), 123–139. doi: 10.1007/s10649-019-9878-z
- Schrepp, M., Hinderks, A. & Thomaschewski, J. (2017a). Konstruktion einer Kurzversion des User Experience Questionnaire. In M. Burghardt, R. Wimmer, C. Wolff & C. Womser-Hacker (Hg.), *Mensch und Computer 2017 – Tagungsband* (S. 355–360). Regensburg: Gesellschaft für Informatik e. V. doi: 10.18420/muc2017-mci-0006
- Schrepp, M., Hinderks, A. & Thomaschewski, J. (2017b). Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103–108. doi: 10.9781/ijimai.2017.09.001
- Schrepp, M., Hinderks, A. & Thomaschewski, J. (2017c). Construction of a benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4), 40–44. doi: 10.9781/ijimai.2017.445

- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D. & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30–41. doi: 10.1109/TLT.2016.2599522
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In P. van Zee, M. Burnett & M. Chesire (Eds.), *Proceedings 1996 IEEE Symposium on Visual Languages* (pp. 336–343). Boulder: IEEE Comput. Soc. Press. doi: 10.1109/VL.1996.545307
- Stieff, M., Hegarty, M. & Deslongchamps, G. (2011). Identifying Representational Competence with Multi-Representational Displays. *Cognition and Instruction*, 29(1), 123–145.
- Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M. & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, 1(1), 7–27. doi: 10.1068/i0382
- Thoma, V. & Dodd, J. (2019). Web Usability and Eyetracking. In C. Klein & U. Ettinger (Eds.), *Eye Movement Research* (pp. 883–927). Cham: Springer International Publishing. doi: 10.1007/978-3-030-20085-5_21
- Underwood, G. & Everatt, J. (1992). The Role of Eye Movements in Reading: Some Limitations of the Eye-Mind Assumption. In E. Chekaluk & K. Llewellyn (Eds.), *Advances in Psychology* (Vol. 88, pp. 111–169). Amsterdam: Elsevier. doi: 10.1016/S0166-4115(08)61744-6
- User Experience Questionnaire (UEQ) (2021). *User Experience Questionnaire*. Retrieved October 27, 2021 from <https://www.ueq-online.org/>
- Valle, N., Antonenko, P., Dawson, K. & Huggins-Manley, A. C. (2021). Staying on target: A systematic literature review on learner-facing learning analytics dashboards. *British Journal of Educational Technology*, 52(4), 1724–1748.
- van den Haak, M., De Jong, M. & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5), 339–351. doi: 10.1080/0044929031000
- van Gog, T., Paas, F., van Merriënboer, J. J. G. & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237–244. doi: 10.1037/1076-898X.11.4.237
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S. & Santos, J. L. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, 57(10), 1500–1509. doi: 10.1177/0002764213479363
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pederson, T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. 1–6. doi: 10.21105/joss.01686
- Yarbus, A. L. (1967). *Eye movements and vision*. Boston: Springer.

Autoren

Georg Dariush Gorshid (M. Sc.) ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Wirtschaftspädagogik mit dem Schwerpunkt „Berufliches Lehren und Lernen“ an der Universität Mannheim. Seine Masterarbeit beschäftigte sich mit der Entwicklung von Learning Analytics Dashboards.

Kontakt: gorshid@bwl.uni-mannheim.de

Christian Mayer (M. Sc.) ist Mitarbeiter am Lehrstuhl für Wirtschaftspädagogik mit dem Schwerpunkt „Berufliches Lehren und Lernen“ an der Universität Mannheim. Sein Dissertationsvorhaben beschäftigt sich mit Eye-Tracking im Problemlöseprozess in computerbasierten Lernumgebungen.

Kontakt: mayer@bwl.uni-mannheim.de

Prof. Dr. Andreas Rausch ist Inhaber des Lehrstuhls für Wirtschaftspädagogik mit dem Schwerpunkt „Lernen im Arbeitsprozess“ an der Universität Mannheim. Er forscht zu Fragen des Lernens und der Expertiseentwicklung an Arbeitsplätzen, des simulationsbasierten Lernens und der Analyse kaufmännischen Problemlösens.

Kontakt: rausch@uni-mannheim.de

Prof. Dr. Jürgen Seifried ist Inhaber des Lehrstuhls für Wirtschaftspädagogik mit dem Schwerpunkt „Berufliches Lehren und Lernen“ an der Universität Mannheim. Er beschäftigt sich hauptsächlich mit Fragen der Gestaltung und Evaluation von Lehr-Lern-Settings, dem Lernen am Arbeitsplatz sowie der Digitalisierung in der beruflichen Bildung.

Kontakt: juergen.seifried@uni-mannheim.de