

Evaluation eines Dashboard-Prototypen für Lehrende in der Alphabetisierung zur interaktiven Auswertung kompetenzdiagnostischer Datensätze

Imke A.M. Meyer
imeyer@uni-bremen.de

Universität Bremen

Abstrakt

Laut LEO-Studie 2018 leben in Deutschland 6,2 Millionen deutschsprachige Erwachsene mit geringer Literalität. Diagnose und Förderung der Betroffenen sind zeit- und personalintensiv, da literale Kompetenzen stark variieren und individuelle Förderansätze notwendig sind. Digitale Diagnoseinstrumente können Lehrende dabei unterstützen. Der vorliegende Artikel evaluiert die Nutzung eines interaktiven Dashboard-Prototypen zur Auswertung förderdiagnostischer Daten durch Lehrende hinsichtlich Gebrauchstauglichkeit, Verständlichkeit der Datenvisualisierungen und möglicher Einsatzszenarien. Der Prototyp wurde dazu mit Lehrkräften aus Alphabetisierungskursen und Berufsschulen mittels einer Think-Aloud-Testung getestet. Insgesamt konnten die Getesteten das Dashboard gut bedienen, doch die Ergebnisse unterstreichen ein grundlegendes Designdilemma zwischen Komplexitätsreduktion und der Bereitstellung umfassender Informationen. Die Analyse legt außerdem nahe, dass Fehlinterpretationen der Nutzer:innen weniger auf mangelnde Visualization Literacy als auf fehlendes domänenspezifisches Fachwissen zurückzuführen sind. Die Lehrenden identifizierten verschiedene Einsatzszenarien für die Unterrichtsplanung, Feedbackgestaltung und Diagnostik. Während Lehrkräfte aus Alphabetisierungskursen die mögliche Zeitersparnis betonten, nannten Berufsschullehrende gerade zeitliche Knappheit als Hindernis für den Einsatz. Insgesamt unterstreicht die Analyse der Ergebnisse die Notwendigkeit, bereits in der Entwicklung wie auch in der Evaluation nicht nur technische, sondern auch pädagogisch-didaktische sowie sozio-kulturelle Aspekte zu berücksichtigen und aufeinander zu beziehen.

Stichwörter: e-learning; Lehrenden-Dashboard; Geringe Literalität; Usability-Testung; Lehrenden-Dashboard Design

Abstract

According to the LEO 2018 study, there are 6.2 million German-speaking adults with low literacy skills living in Germany. Diagnosing and supporting those affected is time-consuming and labor-intensive, as literacy skills vary greatly and individual support approaches are necessary. This article evaluates an interactive dashboard prototype for the analysis of diagnostic data by teachers in terms of usability, comprehensibility of data visualizations, and possible application scenarios. The prototype was tested with teachers from literacy courses

and vocational schools using think-aloud testing. Overall, the test subjects were able to use the dashboard well, but the results highlight a fundamental design dilemma between reducing complexity and providing comprehensive information. The analysis also suggests that the misinterpretations made are less due to a lack of visualization literacy than to a lack of domain-specific expertise. Teachers saw various usage scenarios for lesson planning, feedback design, and diagnostics, with literacy teachers emphasizing the time savings, while vocational school teachers mentioned time constraints as an obstacle. Overall, the analysis of the results underscores the need to consider and relate not only technical aspects but also pedagogical, didactic and socio-cultural aspects during both development and evaluation.

Keywords: e-learning; teacher dashboard; low literacy; usability-testing; teacher dashboard design

1. Einleitung

„Leben mit geringer Literalität“ ist der Titel der LEO-Studie 2018, nach welcher 6,2 Millionen Erwachsene in Deutschland geringe Lese- und Schreibkompetenzen besitzen. Dies umfasst das Lesen und Schreiben einzelner Buchstaben und Wörter, jedoch nicht das Beherrschen zusammenhängender Texte (Grotlüschen u.a. 2020, 6). Auch wenn die erste LEO-Studie 2010 keinen generellen Teilhabeausschluss der Betroffenen festgestellt hat (Grotlüschen und Riekmann 2012), zeigt die LEO-Studie 2018, dass Erwachsene mit geringer Literalität mit verschiedenen gesellschaftlichen Ausschlussrisiken konfrontiert sind. Betroffene trauen sich z.B. die Nutzung von Online-Stellenbörsen oder Online-Wohnungsbörsen deutlich weniger zu als die Gesamtbevölkerung (Buddeberg und Grotlüschen 2020, 214). Ein ähnliches Bild zeigt sich in Bezug auf kritisch hinterfragende digitale Kompetenzen (Buddeberg und Grotlüschen 2020, 215), die Nutzung des Internets für gesundheitsbezogene Informationen (Heilmann 2020, 266) oder für Online-Banking (Grotlüschen, Buddeberg, Dutz, u. a. 2020, 42). Mit einer zunehmenden Digitalisierung ist somit auch eine Verschärfung der Ausschlussrisiken zu befürchten.

Betrachtet man nun die aktuelle Schülerschaft, zeichnet sich ein ähnliches Bild: Die PISA-Studie 2022 zeigt, dass 35,3% der 15-jährigen Schüler:innen nicht-gymnasialer Schulformen über nur geringe Lesekompetenzen (Stufe 2 oder niedriger) verfügen und somit als leseschwach gelten (Heine u. a. 2023, 159). Zudem haben sich die Ergebnisse von Schüler:innen aller Schulformen im Vergleich zur PISA-Studie 2018 signifikant verschlechtert. Auch der IQB-Bildungstrend 2022 macht deutlich, dass etwa 33% der Neuntklässler:innen im Bereich Lesen (Schneider und Boemmel 2023, 67) und 22% im Bereich Orthografie (ebd., 72) die Mindeststandards für einen mittleren Schulabschluss nicht erreichen. Um betroffene Personen insbesondere noch vor dem bzw. während des Einstiegs in das Berufsleben zu erreichen, bietet sich die Unterstützung von Menschen mit mangelnden Grundbildungskompetenzen auch in Berufsschulen an.

Zur Förderung der Betroffenen existieren Angebote von privaten und staatlichen Bildungsträgern. Im Jahr 2023 hatten Alphabetisierungskurse an Volkshochschulen 18.389 Belegungen (Horn u. a. 2024, 68). Dies erscheint in Relation zur Gesamtmenge der

Betroffenen gering. Da die Kompetenzen von (jungen) Erwachsenen mit geringen Lese- und Schreibkenntnissen jedoch stark variieren, bedarf es eines individuellen Förderansatzes. Dies führt zu einer personalintensiven Betreuung der Betroffenen (Wolf u. a. 2011). Lehrende in der Alphabetisierung sind häufig auf Honorarbasis oder aufgrund von Projektfinanzierungen befristet tätig, was zu Unsicherheit und einer schwierigen Planbarkeit führt. Zudem existieren keine bundesweiten Standards in Bezug auf die Qualität der Alphabetisierungsarbeit (Klemm u. a. 2023). Gleichzeitig basiert die Diagnostik schriftsprachlicher Kompetenzen auf komplexen Kompetenzmodellen (Schepers u. a. 2022; Koppel u. a. 2022). Ein möglicher Ansatz, um Lehrende in der Alphabetisierung zu unterstützen und eine größere Anzahl von Betroffenen zu fördern, ist der Einsatz einer digitalen Förderdiagnostik (Koppel u. a. 2022; Koppel und Wolf 2014).

Die Auswertung förderdiagnostischer Daten kann als Teilaspekt von Data-Driven Decision Making (DDDM) betrachtet werden. Gängige DDDM-Ansätze (Breiter und Light 2006; Schildkamp u. a. 2018; Schildkamp 2019) umfassen die Datenerhebung, -analyse und -nutzung zur Entscheidungsfindung auf Unterrichtsebene, Schulebene oder bildungspolitischer Ebene. Ergänzend dazu gibt es Ansätze wie das response-to-intervention Modell (RTI) aus dem Bereich der Sonderpädagogik, die die systematische Nutzung von Lernstandsdaten zur individuellen Förderung von Lernenden beschreiben (Grosche und Huber 2012; Souvignier und Förster 2025). Die regelmäßige förderdiagnostische Datenerhebung und -auswertung, beispielsweise zur zielgerichteten Anpassung einer Lerneinheit, kann der Entscheidungsebene der Schulklasse zugeordnet werden (Souvignier und Förster 2025). Zu berücksichtigen ist jedoch, dass Entscheidungen nicht ausschließlich auf Grundlage von Daten, sondern in Kombination mit professionellem Erfahrungswissen getroffen werden sollten (Mandinach und Schildkamp 2021). Die Autor:innen fordern zudem, dass Technologien zur Erfassung und Analyse von Bildungsdaten eine tiefgreifende Auseinandersetzung mit den Daten unterstützen, an die Bedürfnisse der Endnutzer:innen angepasst sind und dass in weiteren Studien die Konzeption, Entwicklung, Implementierung und Bewertung dieser Systeme untersucht wird.

Die Herausforderung besteht somit darin, ein System zu entwickeln, das komplexe förderdiagnostische Informationen für eine sehr heterogene Zielgruppe – Lehrende in der Alphabetisierung sowie in der beruflichen Bildung – bereitstellt und zugleich leicht zugängliche Auswertungen bietet und dennoch zu einer vertieften Interpretation anregt. Für die vorliegende Erhebung steht ein interaktiver Gestaltungsentwurf zur Verfügung, der im Rahmen einer Usability-Testung mit Lehrenden aus den Bereichen Alphabetisierung und Berufsschule untersucht wurde. In den folgenden Abschnitten wird zunächst ein Überblick über den Forschungsstand zur Entwicklung und Evaluation von Lehrenden-Dashboards gegeben. Darauf aufbauend werden die Forschungsfragen formuliert. Anschließend werden die Testumgebung otu.lea und der Aufbau des lea.Dashboard-Prototypen beschrieben, der für die Testung zur Verfügung steht. Danach wird das methodische Vorgehen für den Usability-Test sowie dessen Einbettung in den gesamten Entwicklungsprozess des Dashboards erläutert. Abschließend werden die Testergebnisse dargestellt, die Forschungsfragen beantwortet und die Ergebnisse im Hinblick auf aktuelle Forschungsergebnisse diskutiert.

2. Stand der Forschung: Dashboards für Lehrende

Ein Dashboard dient der Informationsdarstellung. Dabei werden die zentralen Informationen kompakt und übersichtlich auf einem Bildschirm visualisiert (Few 2006). Die Informationen sollten entsprechend den Zielen der Nutzenden leicht zugänglich aufbereitet sein. Etablierte Lehr-/Lernplattformen wie Bettermarks oder Khan Academy bieten Lehrenden Dashboard-Ansichten, in denen die Leistungen und der Lernprozess der Lernenden visualisiert werden. In Lehrenden-Dashboards werden Daten von und über Lernende im Kontext des Unterrichts aufbereitet, um Lehrende in ihrem pädagogischen und didaktischen Handeln zu unterstützen. Dies können zum einen Daten zu Leistungsindikatoren sein, um prädiktive Aussagen zu treffen. Dieser Bereich der Learning Analytics (LA) umfasst die algorithmische Auswertung digital erfasster Lern- und Prozessdaten (Van Leeuwen u. a. 2022). Andererseits können förderdiagnostische Daten aus Testergebnissen mit dem Ziel erhoben werden, konkrete Stärken und Schwächen einzelner Lernender zu beschreiben und daraus Fördermaßnahmen abzuleiten und aufzubereiten (Grosche und Huber 2012; Wolf u. a. 2011; Koppel 2017). Im Bereich der Alphabetisierung kommt diesem Bereich eine besondere Bedeutung zu, da aufgrund meist sehr heterogener Kompetenzstände und individueller Lernbiografien die individuelle Förderung der Betroffenen im Vordergrund steht (Ehmig 2025).

Die aktuellen Forschungsarbeiten zu Lehrenden-Dashboards zeigen ein unterschiedliches Bild hinsichtlich ihrer Nutzungsweisen und Effekte. Der Fokus liegt dabei mehrheitlich auf Dashboards aus dem Learning Analytics Bereich. Ez-Zaouia und Lavoué (2017) entwickelten das EMODA-Dashboard, das Lehrenden dabei unterstützt, die Emotionen von Lernenden zu erfassen und deren Lernaktivitäten besser zu verstehen. Die Datengrundlage zur Darstellung und Auswertung der Emotionen umfasst reale multimodale Datenquellen (Audio, Video, Selbstauskunft und Interaktionsspuren). In der Evaluation mit zwei Tutor:innen wurde das Dashboard als positiv und vielversprechend bewertet. Es war gut bedienbar und die Fragen zu den Visualisierungen konnten ohne Hilfestellung beantwortet werden, lediglich die Zeitleistenvisualisierungen bereiteten Verständnisprobleme. Safsouf u. a. (2021) untersuchten im Rahmen einer Studie mit Lehrenden und Lernenden ein Dashboard (TABAT), das den Lernfortschritt visualisiert und die Abbrecherquote der Lernenden vorhersagt. Die aufbereiteten Daten werden in Form von Datenvisualisierungen als Einzel- und Gruppenergebnisse für Lehrende und Lernende zur Verfügung gestellt. In einer achtwöchigen Studie mit 26 Studierenden zeigte sich ein positiver Effekt auf den Progress-Score sowie pünktlichere und längere Aufgabenbearbeitungen. In einer weiteren Studie wurde ein Dashboard in Zusammenarbeit mit Lehrenden entwickelt und in einem zehnwöchigen Kurs evaluiert (Nguyen 2024). Die Ergebnisse zeigen, dass die Lehrkraft eher an einer umfassenden Übersicht über den Kurs als an einzelnen Einblicken interessiert war. Die Funktion zur Voraussage von Entwicklungen wurde von der Lehrkraft nicht genutzt, da keine Handlungen daraus abgeleitet werden konnten. In einer aktuellen Studie untersuchten Borgards u. a. (2024) 462 Feedbacknachrichten, die Lehrende auf Basis von LA-Informationen in einem Dashboard an Lernende gesendet hatten. Lehrende stellten leistungsschwächeren Lernenden Feedback auf Aufgabenebene und leistungsstarken Lernenden Feedback auf Übersichtsebene zur Verfügung.

Isaias und Viana (2020) kamen in einer Befragung von 14 Hochschullehrenden zu dem Ergebnis, dass die Lehrenden nur über wenig Kenntnisse bezüglich der Nutzungsmöglichkeiten von Dashboards verfügen und diese eher als Kontrollinstrumente und nicht als Unterstützung des Lehr-Lern-Prozesses betrachten. Ein besseres Verständnis darüber, wie diese Technologien pädagogisch und didaktisch sinnvoll in Lehr-Lern-Kontexten eingesetzt werden können, ist somit vonnöten. Zudem stellten Kaliisa u. a. (2024) in einem Literaturreview heraus, dass der Einsatz von Learning-Analytics-Dashboards bislang zu keiner nachhaltigen Verbesserung von Lernleistungen geführt hat. Die meisten Arbeiten zeigen geringe Effekte, die auf die konkrete Untersuchung begrenzt sind. Ein weiteres Problem ist, dass es für Lehrende eine Herausforderung darstellt, korrekte Informationen und pädagogische Fragestellungen aus den Datenvisualisierungen abzuleiten bzw. diese zu erkennen (Jivet u. a. 2017; Borgards u. a. 2024). In einer Untersuchung von Pozdniakov u. a. (2024) interpretierten Lehrende mit geringer Visualization Literacy (VL) Datenvisualisierungen weniger tiefgehend als Lehrende mit hoher VL. Liu u. a. (2024) untersuchen die Auswirkungen verschiedener Gestaltungen von Datenvisualisierungen auf die kognitive Belastung (Sweller 1988) von Lehrenden. Dabei fanden die Autor:innen heraus, dass insbesondere Lehrende mit niedriger VL eine deutlich geringere kognitive Belastung hatten, wenn sie Dashboards mit Data-Storytelling-Elementen verwendeten. Data Storytelling ist ein Ansatz zur Vereinfachung komplexer Daten durch einen narrativen Aufbau der Datenvisualisierungen, der mit Kontextinformationen, wie beispielsweise Leitfragen, verbunden ist. Demgegenüber stehen Ergebnisse, die zeigen, dass die Verwendungen von Visualisierungsanleitungen nicht zu einer besseren Interpretation von Datenvisualisierungen durch Lehrende führt (Pozdniakov u. a. 2024).

3. Partizipative Entwicklungsansätze von Lehrenden-Dashboards

Eine Möglichkeit Dashboards effektiver und zielgerichteter zu gestalten, sind Entwicklungsansätze, die Lehrenden in den gesamten Entwicklungsprozess miteinbeziehen (Nguyen 2024; Pozdniakov u. a. 2022; Ez-Zaouia 2020; Karademir u. a. 2022; Kaliisa und Dolonen 2023). Pozdniakov u. a. (2022) beschreiben ein menschenzentriertes mehrstufiges Vorgehen zu Entwicklung eines Lehrenden-Dashboards. Auf Grundlage von Interviews mit Lehrenden wurden deren wichtigsten Fragen extrahiert. Entlang dieser Fragen wurden Daten zur Analyse von online Gruppenaktivitäten im Dashboard zur Verfügung gestellt. Die Evaluation zeigte, dass die Lehrenden die dargestellten Daten insgesamt gut interpretieren und in Handeln integrieren konnten, jedoch einzelne Verständnisfragen bzgl. mehrdeutiger Frageformulierungen und fehlender Transparenz von Schwellenwerten blieben. Die Ergebnisse zeigen das Potential von Entwicklungsansätzen, die Lehrenden in den Prozess integrieren. Karademir u. a. (2022) arbeiten nach einem von Forschungsfragen geleiteten, mehrstufigen Ansatz zur Dashboard-Entwicklung. Auch hier werden die Bedarfe der Lehrenden durch Interviews mit der Zielgruppe erfasst. In dem Modell wird die Interdisziplinarität zwischen Softwareentwicklung, pädagogischem Design und Usability-Forschung als notwendig hervorgehoben. Kaliisa und Dolonen (2023) haben das CADA-Dashboard zur automatisierten Analyse von Forenbeiträgen in einem

Lernmanagementsystem entwickelt. Lehrende sollen auf einen Blick die Beteiligung der Lernenden, Stimmungen und Schlüsselbegriffe erkennen können. Die Anwendung wurde nach dem Design-Based-Research-Ansatz iterativ und unter Einbezug von Lehrenden entwickelt. Dabei wurden verschiedene Formen von Prototypen zur Evaluation genutzt bevor die Anwendung programmiert wurde. Van Leeuwen u. a. (2023) argumentieren zudem, dass beim Einbezug von Lehrenden in den Entwicklungsprozess von Bildungstechnologien ein Gleichgewicht zwischen der Lehrendenperspektive und bildungstheoretischen Interessen gefunden werden muss. Menschenzentriertes Design kann laut den Autoren beide Perspektiven in Einklang bringen, indem die Rollen von Forschenden und Lehrenden klar definiert werden. So können Lehrende demnach mit ihrem Fachwissen Ziele und Nutzungsweisen eines Dashboards beschreiben. Die Forschenden sollten den Prozess steuern und die endgültige Entscheidung über Metriken und Funktionen des Dashboards treffen. Die Autor:innen plädieren zudem dafür, Lehrende durch Fortbildungsangebote zu unterstützen, um eine sinnvolle und nachhaltige Integration in den beruflichen Alltag zu ermöglichen. Ez-Zaouia (2020) beschreibt ein Modell zur lehrerzentrierten Entwicklung von Dashboards, das sich an gängigen mehrstufigen Designmodellen orientiert. In der ersten sogenannten Situate-Phase sollen alle Anforderungen an das Dashboard mittels Interviews oder kontextbezogener Analysen iterativ gesammelt werden. In der zweiten Phase Ideate werden erste Gestaltungsentwürfe erstellt, evaluiert und überarbeitet. In der dritten Phase Develop soll ein Prototyp in Pilotversuchen auf seine Benutzertauglichkeit und die Erfahrungen sowie Wahrnehmungen der Lehrenden hin untersucht werden. In der letzten Phase Evaluate wird die Anwendung im Realeinsatz getestet, um Ergebnisse zur Benutzerfreundlichkeit und zu pädagogischen Vorgehensweisen zu gewinnen.

4. Usability, User Experience und Learning Experience Design

Eine gute Benutzerfreundlichkeit und Benutzererfahrung kann für ein angenehmes Benutzergefühl sorgen (Molich und Nielsen 1990). Usability und User Experience sind Teilgebiete der Mensch-Computer-Interaktionsforschung. Usability wird als „Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung von bestimmten Nutzern verwendet werden kann, um bestimmte Ziele mit Effektivität, Effizienz und Zufriedenheit in einem bestimmten Nutzungskontext zu erreichen“ (International Organization for Standardization 2018) definiert. In der DIN-Norm EN ISO 9241-210 wird User Experience als „die Wahrnehmungen und Reaktionen einer Person, die sich aus der Nutzung und/oder der erwarteten Nutzung eines Produkts, Systems oder einer Dienstleistung ergeben“ beschrieben (International Organization for Standardization 2019). Vermeeren u. a. (2010) identifizierten in einer umfassenden Übersicht 96 verschiedene Methoden zur Evaluation von Usability und User Experience und kategorisierten diese nach Entwicklungsphase, Informationsquelle, Evaluationskontext, Datentyp und Zeitaufwand. Eine Methode, die mit klickbaren Prototypen verschiedener Stadien oder vollständiger Software zur formativen Evaluation von Usability-Problemen genutzt werden kann, ist der Think-Aloud-Test (Nielsen 2010; Vermeeren u. a. 2010; Rubin 2008). Während einer Think-Aloud-Testung müssen die Teilnehmenden Aufgaben mit einer Software erfüllen und ihre Gedanken und Gefühle währenddessen frei äußern. Dies gibt nicht nur Aufschlüsse darüber, wie Nutzende mit einem Interface

interagieren sondern auch warum sie dies tun (Goodwin 2011; Nielsen 2010). Die Think-Aloud-Methode erlaubt Einblicke in mentale Modelle, Entscheidungsprozesse und Usability-Probleme der Nutzer:innen (Schmidt, Earnshaw, u. a. 2020). Entgegen der Annahme, dass bereits sehr kleine Stichproben zur Identifikation von Usability-Problemen ausreichen (Nielsen und Landauer 1993; Nielsen 2010), zeigt Faulkner (2003), dass eine Gruppe von fünf Teilnehmenden nicht ausreichen, um Usability-Probleme verlässlich zu erkennen. So entdecken erst zehn Personen mindestens 80% und zwanzig Personen mindestens 95% der Usability-Probleme.

Nokelainen (2006) erweitert den Begriff der Usability um die pädagogische Usability. In diesem Konzept werden in Bezug auf Lernmaterialien Aspekte der pädagogischen Gestaltung, der Interaktionen der Lernenden mit dem Material sowie der Bewertung des Materials durch die Lernenden erfasst. Aufbauend auf der pädagogischen Usability umfasst die Learning Experience (LX) die Wirksamkeit einer Lernintervention, die Beziehung zwischen dem Lernenden/Lehrenden und der Lerntechnologie, die pädagogischen Strategien sowie den Lernkontext (Schmidt, Tawfik, u. a. 2020, 4). Liu u. a. (2023) heben die Multidimensionalität der Learning Experience hervor und fassen auf Basis eines Literaturreviews drei LX-Dimensionen zusammen: „Aktivitäten“, „Wahrnehmungen“ sowie „Einstellungen und Verhalten“. (Lern)technologien werden als Vermittler beschrieben, die bei der Bewertung von Lernerfahrungen in Verbindung mit pädagogischen Zielen berücksichtigt werden sollten. Der Zusammenhang zwischen technologischen und pädagogischen Aspekten wird in der Theorie des Learning Experience Designs noch deutlicher erfasst (Jahnke u. a. 2022). Learning Experience Design wird als ein multidimensionales Konstrukt aus einem pädagogischen, einem technischen und einem sozio-kulturellen Bereich verstanden. Im Mittelpunkt stehen dabei die Fragen, wie Lernende ihre (Lern)ziele erreichen, mit Technologie interagieren und mit anderen kollaborieren. Eine besondere Betrachtung gilt den Schnittstellen zur Mensch-Computer-Interaktion (HCI). Zu verstehen, wie und warum Lernende und Lehrende mit Software interagieren, kann Fehler bei der technischen, gestalterischen und pädagogischen Konzeption verhindern (Schmidt, Earnshaw, u. a. 2020, 21). Voraussetzung dafür ist, die Lernenden oder im vorliegenden Fall die Lehrenden als Nutzende zu verstehen. Angesichts der interdisziplinären Natur des Learning Experience Designs plädieren die Autor:innen dafür, Konzepte und Evaluationsmethoden aus der HCI-, Usability- und UX-Forschung mit Theorien und Prozessen des Lerndesigns zu verknüpfen.

5. Forschungsfragen

Obwohl Learning-Analytics-Dashboards in verschiedenen Kontexten entwickelt und getestet wurden, besteht aktuell weiterer Forschungsbedarf. Die beschriebenen Studien zeigen auf der einen Seite die Potentiale vom Einsatz von Lehrenden-Dashboards, wie z.B. gezielteres Feedback durch die Lehrenden oder Verbesserung von Lernleistungen. Eine Generalisierung über einzelne Untersuchungen hinaus sei bislang aber nicht möglich (Kaliisa u. a. 2024). Als eine zentrale Herausforderung stellt sich das Ableiten konkreter pädagogischer Handlungen dar, insbesondere bei Lehrenden mit geringer Visualization Literacy (Pozdniakov u. a. 2024; Liu u. a. 2024). Partizipative Entwicklungsmodelle, die Lehrende von Beginn an einbeziehen und so eine zielgruppen- und bedarfsgerechte Entwicklung ermöglichen sollen, zeigen sich

als vielversprechende Ansätze (Ez-Zaouia 2020; Liu u. a. 2024). Vor diesem Hintergrund erscheinen die iterative und partizipative Entwicklung und Evaluation eines Dashboards mit förderdiagnostischen Daten besonders sinnvoll. Zudem sind weitere Erkenntnisse zur Nutzung von Datenvisualisierungen durch Lehrende mit stark variierender Visualization Literacy über den Alphabetisierungsbereich hinaus relevant. Auf dieser Grundlage ergeben sich folgende Forschungsfragen für die Evaluation des vorliegenden Dashboard-Prototypen:

FF1: Welche Usability-Probleme treten bei der Nutzung des Dashboards auf und wie bewerten die Nutzer:innen die Usability des Dashboards?

FF2: Inwieweit sind die Lehrkräfte in der Lage, die im Dashboard dargestellten diagnostischen Informationen korrekt zu beschreiben und zu interpretieren?

FF3: Inwiefern entsprechen die Funktionalitäten des Dashboards den Bedarfen und Handlungsmöglichkeiten der Lehrenden?

6. Otu.lea und das lea.Dashboard

Im Rahmen des vom BMBF geförderten Projekts lea.online (Förderkennzeichen W143600) wurde der bestehende Kompetenztest otu.lea für Menschen mit Grundbildungsbedarf überarbeitet und ein Dashboard für Lehrende zur förderdiagnostischen Auswertung entwickelt (Koppel u. a. 2022). otu.lea ist eine Online-Testumgebung, mit der Lernende eigenständig einen Kompetenztest in den Dimensionen Lesen, Schreiben, Sprachgefühl und Rechnen absolvieren können. Diese Tests können als Selbsttest von zu Hause aus oder im Rahmen einer Bildungsmaßnahme, z.B. einem Alphabetisierungskurs, durchgeführt werden. Dabei kann otu.lea als Eingangs-, Verlaufs- oder Abschlussbewertung genutzt werden. Die Kompetenztests sind in verschiedenen Schwierigkeitsstufen unterteilt und basieren auf den lea.-Kompetenzmodellen (Schepers u. a. 2022; Grotlüschen u. a. 2010; Kretschmann und Wieken 2010a; 2010b). Die feingranulare Auswertung der einzelnen Kompetenzen ermöglicht eine Rückmeldung der Kompetenzstände auf übergeordneter Ebene in Form von Kompetenzstufen. Diese Stufen werden in den lea.-Kompetenzmodellen Alpha-Levels genannt. Jede Dimension ist in fünf Alpha-Level unterteilt, wobei Alpha-Level 1 die niedrigste und Alpha-Level 5 die höchste Stufe darstellt. Zudem ist eine Rückmeldung auf der Ebene detaillierter Einzelkompetenzen möglich, die in den lea.-Kompetenzmodellen Kann-Beschreibungen bezeichnet werden (Koppel 2017; Koppel u. a. 2022; Meyer 2024). Darüber hinaus werden keine Prozessdaten wie beispielsweise Lese- oder Tippgeschwindigkeit erfasst.

Mithilfe des entwickelten Dashboards sollen die Ergebnisse aller Dimensionen der otu.lea-Kompetenzdiagnostik (Schreiben, Lesen, Sprachgefühl und Rechnen) für Kursleitende übersichtlich, verständlich und interaktiv dargestellt werden. Ziel ist es, die individuellen Lernstände im Dashboard so aufzubereiten, dass Lehrende bei der gezielten Förderung der Lernenden in Einzelbetreuung oder in der Kursarbeit unterstützt werden. Das Dashboard soll Förderbedarfe der Betroffenen für die Lehrenden schnell und einfach sichtbar machen. Sie sollen konkrete Kompetenzdefizite identifizieren und darauf aufbauend passende Aufgaben und Materialien auswählen sowie gegebenenfalls Lerngruppen bilden können. Das lea.Dashboard dient somit der pädagogischen Förderdiagnostik und insbesondere der

kontinuierlichen Evaluation von Lernfortschritten auf Gruppen- und Personenebene. Die Anwendung ist für die Desktop-Nutzung konzipiert. Bei der iterativen Entwicklung des Dashboards wurde besonderer Wert darauf gelegt, zielgruppen- und domänenspezifische Bedarfe zu identifizieren und systematisch in den Entwicklungsprozess mit einfließen zu lassen (Meyer 2024). Die Anforderungen unterteilen sich in **domänenspezifische Anforderungen** mit den Unterkategorien *Aufbereitung der Testergebnisse* und *Inhalte der Datenvisualisierungen*, **zielgruppenspezifische Anforderungen** mit den Unterkategorien *Zielgruppengerechte Darstellung* und *Förderdiagnostische Unterstützung der Lehrenden* sowie **gestalterische Anforderungen** mit den Unterkategorien *Intuitive Nutzung*, *Unterstützung der Nutzer:innen*, *Daten und Informationen zu verstehen* und *Nutzerfreundliche Gestaltung*.

Auf dieser Grundlage wurde im Rahmen der gestalterischen Entwicklung ein sogenanntes Userflow erstellt. Userflows zeigen die typischen Wege, die Nutzende wählen könnten, um mit einer Software zu interagieren. Das Ziel besteht darin, die gewünschten Interaktionen einfach und mit möglichst wenigen Schritten umzusetzen. Dies soll eine intuitive und niedrigschwellige Bedienbarkeit des otu.lea-Dashboards ermöglichen. Weiterhin soll sichergestellt werden, dass alle notwendigen Interaktionen mit dem Dashboard absolviert werden können und keine überflüssigen Interaktionen durchgeführt werden müssen. Bei der Erstellung des Userflows wurde der Fokus auf drei Kerninteraktionsfolgen gelegt, die für die produktive Nutzung des Dashboards grundlegend sind. Diese lauten:

- (a) *Analyse von Testergebnissen einer bestimmten Person*,
- (b) *Analyse von Ergebnissen eines gesamten Kurses*,
- (c) *Feingranulare Analyse und Vergleich von Einzelergebnissen innerhalb eines Kurses*

Auf Basis dieser Kerninteraktionen wurde entschieden, dass das Dashboard aus einer **Personenansicht** (Ziel (a)), einer **Kursansicht** (Ziele (b) und (c)) sowie einer **Startseite** bestehen soll. Der grundsätzliche Aufbau soll so gestaltet werden, dass Nutzende von der Startseite aus Kurse und Lernende verwalten und zu Kurs- oder Lernendenergebnissen navigieren können (siehe Abbildung 1). Anhand der Userflow-Grafik (blauer Pfad) ist zu erkennen, dass es zwei Möglichkeiten gibt das Ziel (a) *Analyse eines Testergebnisses einer bestimmten Person* umzusetzen. Lehrende können auf der Startseite Lernende aus der Liste auswählen und gelangen direkt zu deren Testergebnissen. Alternativ können sie zunächst den Kurs auswählen, in dem die Lernenden Mitglied sind. In der Kursansicht können dann einzelne Lernende aus einer Liste ausgewählt werden, um ebenfalls zu den individuellen Testergebnissen zu gelangen. Auf der Personenseite können die Testergebnisse mithilfe eines Filtersystems detailliert analysiert werden. Die Ziele (b) *Analyse von Ergebnissen eines gesamten Kurses* und (c) *Feingranulare Analyse und Vergleich von Einzelergebnissen innerhalb eines Kurses* führen beide auf die Kursansicht (roter und grüner Pfad auf der Userflow-Grafik). Nutzende können auf der Startseite einen Kurs auswählen und gelangen so ebenfalls auf die Kursansicht. Hier werden alle Ergebnisse auf Alpha-Level-Ebene sowie bezogen auf einzelne Kann-Beschreibungen detailliert visualisiert. Die grauen, gepunkteten Linien zeigen, dass Nutzende die Navigationswege auch rückwärts verfolgen können.

- (a) Analyse eines Testergebnisses einer bestimmten Person
- (b) Analyse von Ergebnissen eines gesamten Kurses
- (c) Feingranulare Analyse und Vergleich von Einzelergebnissen innerhalb eines Kurses

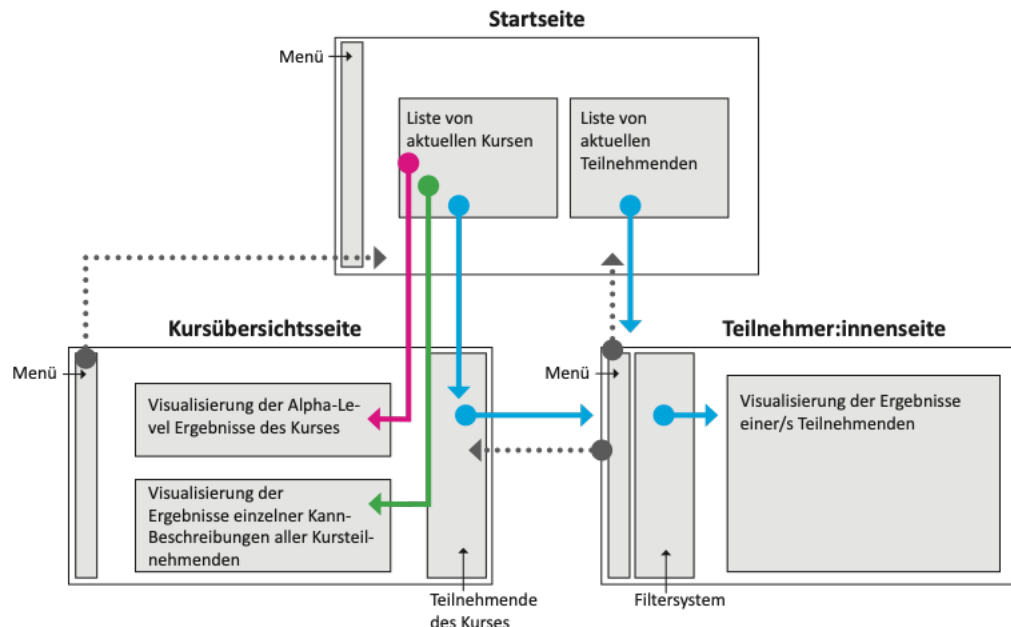


Abbildung 1: Userflow

Als nächster Schritt wurden auf Basis einer Literaturrecherche Gestaltungskriterien für Datenvisualisierungen in Lehrenden-Dashboards beschrieben und die Datenvisualisierungen im *lea*.Dashboard entwickelt (Meyer n.d., eingereicht). Parallel dazu wurden die Gestaltungsentwürfe der einzelnen Seiten des Dashboards erstellt und auf Basis von arbeitsgruppeninternen Iterationsschritten überarbeitet. Abschließend wurde ein interaktiver Prototyp mit der Software Adobe XD erstellt. Im Folgenden werden die Gestaltungsentwürfe der einzelnen Seiten des Dashboards mit Datenvisualisierungen vorgestellt und die für die Usability Testung umgesetzten Interaktionen beschrieben.

Über die Startseite (erster Bereich) können Personen und Kurse verwaltet werden (siehe Abbildung 2). Mit den Buttons „Neue Klasse anlegen“ und „Neu anlegen“ können Klassen und Teilnehmende hinzugefügt werden. Mit den Bearbeiten-Buttons (Stiftsymbol) können bestehende Daten verändert werden. Durch Klicken auf den Ansehen-Button (Augen-Symbol) können die Testergebnisse ganzer Kurse oder einzelner Lernender eingesehen werden. Darüber hinaus ist die Navigation zu den Testergebnissen durch Klicken auf den Namen des Kurses oder der Lernenden möglich. Abgeschlossene Kurse und frühere Lernende können über die Buttons „Archivierte Klassen“ und „Archivierte Teilnehmer:innen“ angesehen und nach Bedarf wieder aktiviert werden. Für die Usability-Testung wurden im interaktiven Prototyp alle auf der Seite sichtbaren Buttons mit Hover-Status versehen. Das bedeutet, dass die Buttons ihre Erscheinung verändern sobald sich der Mauszeiger über die Schaltfläche bewegt. Außerdem wurde die Navigation zur Kursansicht des Kurses August 2020 über den Ansehen-Button sowie den Schriftzug umgesetzt (siehe Abbildung 2).

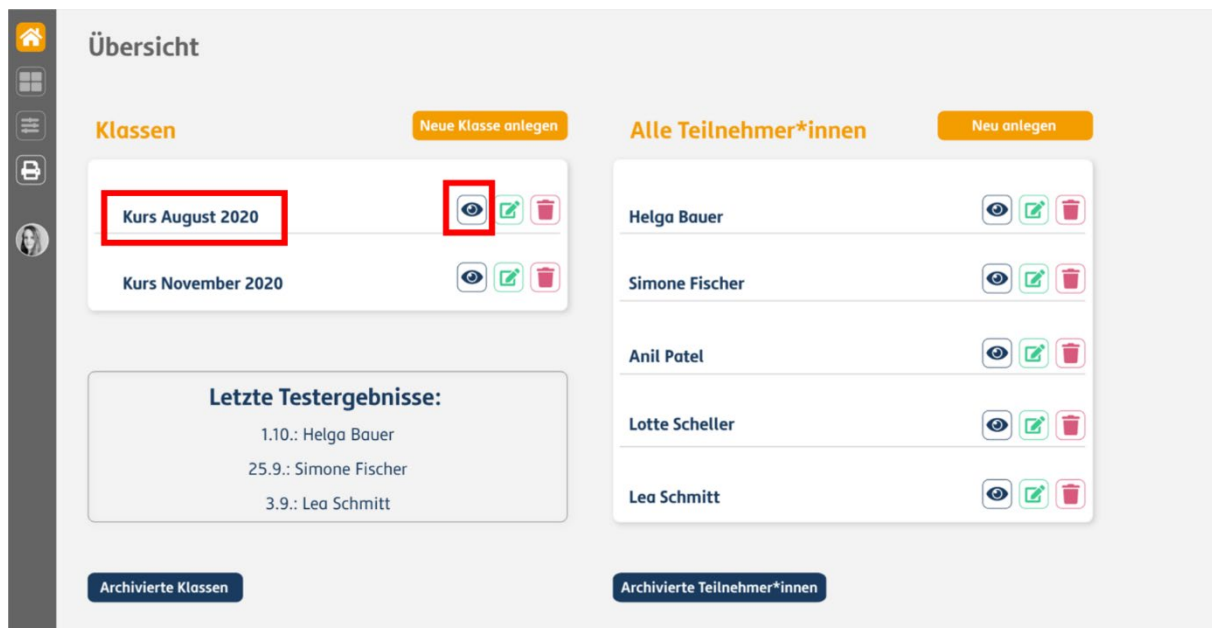


Abbildung 2: Startseite (Gestaltungsentwurf), klickbare Elemente mit roten Kästen hervorgehoben

Für die Untersuchung des interaktiven Gestaltungsentwurfs des Dashboards wird nur der Bereich Schreiben betrachtet, weshalb in den folgenden Gestaltungsentwürfen nur diese Dimension gezeigt wird. Die Darstellung der anderen Dimensionen folgt demselben Aufbau. In der Kursansicht (zweiter Bereich) werden die Ergebnisse aller Teilnehmenden eines Kurses visualisiert (siehe Abbildung 3). Über das Menü auf der linken Seite können die Startseite, die Kursansicht und eine Druckoption ausgewählt werden. Über das Menü auf der rechten Seite können einzelne Lernende eines Kurses ausgewählt werden. In der Mitte der Kursansichtsseite befindet sich die Verlaufsgrafik. In dieser werden die Ergebnisse bezogen auf die einzelnen Alpha-Level (Kompetenzstufen) aller Kursteilnehmenden im zeitlichen Verlauf dargestellt. Die Alpha-Level-Ergebnisse werden nicht als Punkt-, sondern als Prozentwerte dargestellt, und zwar differenziert für jedes einzelne Level. Jeder Balken steht dabei für das Erreichen eines bestimmten Prozentwerts eines Alpha-Levels. Durch Mouseover oder Klick können die Ergebnisse eines Teilnehmenden hervorgehoben werden (siehe Abbildung 4). Unter der Verlaufsgrafik befindet sich eine Listenansicht einzelner getesteter Kompetenzen (siehe Abbildung 5). Dort können die Kursleitenden alle Ergebnisse der Lernenden auf Kann-Beschreibungs-Ebene einsehen und vergleichen. Alle Kann-Beschreibungen sind in didaktisch sinnvollen Gruppen sortiert. Diese sollen Anregungen zur weiteren Förderung liefern. Im Bereich Schreiben decken die Gruppen orthografische Prinzipien ab (Koppel u. a. 2022). Die einzelnen Fortschrittsbalken unter den Kann-Beschreibungen in den Listen zeigen das durchschnittliche Ergebnis des Kurses zur jeweiligen Kompetenz. Ein gefüllter Balken bedeutet, dass alle getesteten Personen die Kompetenz voll erfüllt haben. Neben jeder Kann-Beschreibung befindet sich ein Pfeil, über den die Liste ausgeklappt werden kann, sodass alle Einzelergebnisse der Lernenden sichtbar werden. So können Lehrende einsehen, ob bestimmte Kompetenzen vom gesamten Kurs nicht erfüllt werden oder ob es einzelne Ausreißer in der Gruppe gibt. Auch auf der Kursübersichtsseite wurden alle im Design enthaltenen Buttons mit Hover-Status versehen. Darüber hinaus wurde das Hovern über die Balken der Verlaufsgrafik zum Hervorheben

einzelner Ergebnisse im Prototyp interaktiv umgesetzt (siehe Abbildung 4). Zudem ist es möglich die Detailliste der Kompetenz „Kann Dauerkonsonanten (m, n, l etc.) am Anfang des Wortes verschriftlichen“ aus- und einzuklappen (siehe Abbildung 5).

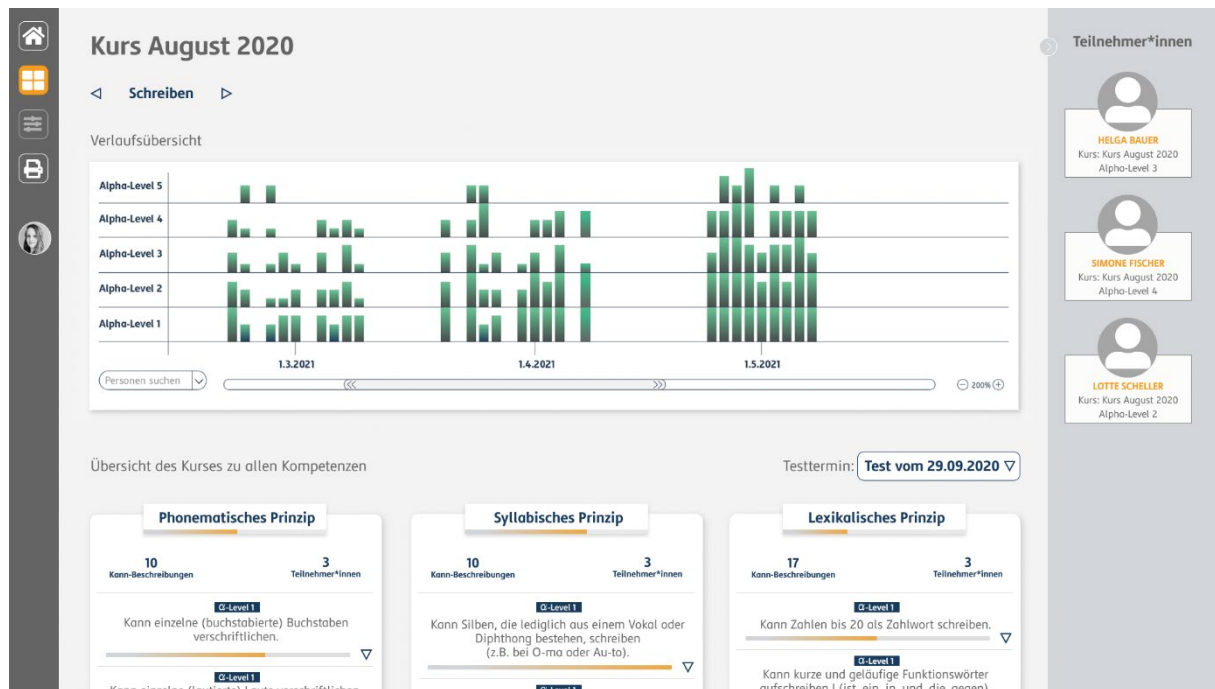


Abbildung 3: Kursansicht im lea.Dashboard (Gestaltungsentwurf)

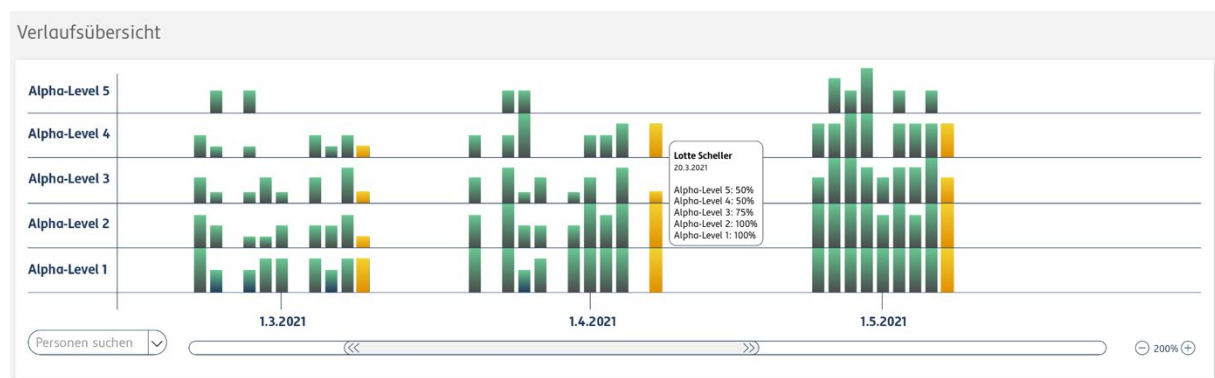


Abbildung 4: Detailansicht Verlaufsgrafik mit Hervorheben einzelner Ergebnisse durch Hover über die Balken

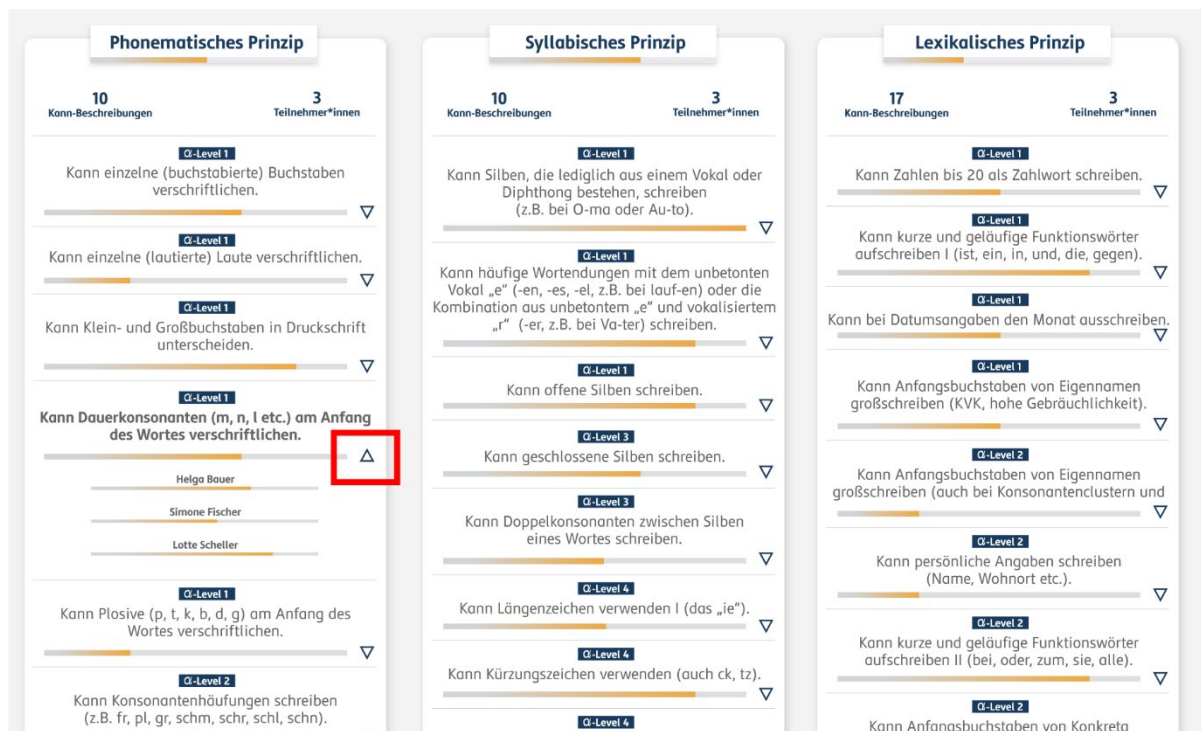


Abbildung 5: Listenansicht, klickbares Element mit roten Kästen hervorgehoben

Der dritte Teil des Dashboards ist die individuelle Ansichtsseite einzelner Lernender (siehe Abbildung 6). Hier sind alle Kann-Beschreibungen, die in einem Testset einer Dimension getestet wurden, in Form von digitalen Karten aufgelistet. Diese Karten beinhalten folgende Parameter: Titel der Kann-Beschreibung, Alpha-Level der Kann-Beschreibung, Testergebnis des Lernenden zu dieser Kann-Beschreibung, den Leistungsvergleich zu vorherigen Ergebnissen und eine detaillierte Erklärung der Kann-Beschreibung auf der „Rückseite“ der Karte. Diese kann durch Klicken auf die Karte aufgerufen werden (Flip-Card-Effekt). Über ein Filtersystem können sich Lehrkräfte ausschließlich Karten mit spezifischen Informationen anzeigen lassen (z. B. alle Karten mit Leistungsverbesserung oder nur die Karten eines bestimmten Alpha-Levels). Im Filtersystem sind alle Parameter der Karten enthalten. So sollen Lehrkräfte befähigt werden, spezifische und individuelle Zusammenhänge der Testergebnisse ihrer Lernenden sichtbar zu machen. Im Prototyp wurden das Scrollen durch den Bereich der Karten, das Klicken und Umdrehen der Karte „Kann einzelne (lautierte) Laute verschriftlichen“ sowie das Auswählen verschiedener Filter interaktiv umgesetzt (siehe Abbildung 6).

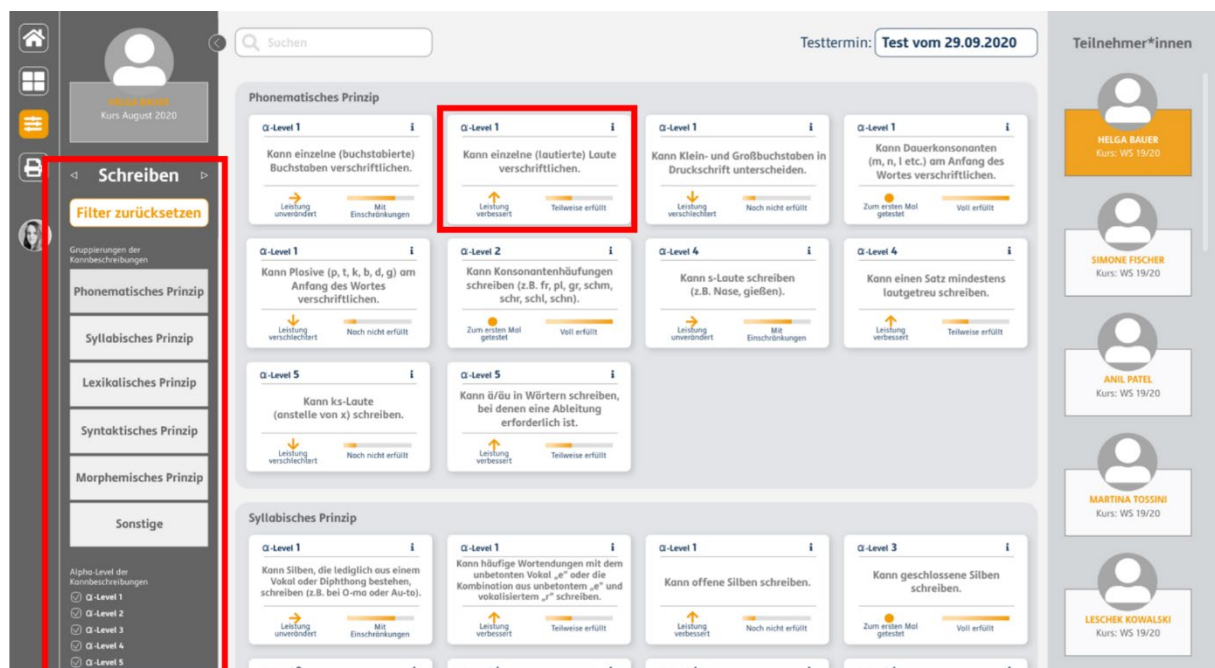


Abbildung 6: individuelle Ansichtsseite einzelnen Lernenden (Gestaltungsentwurf), klickbare Elemente mit roten Kästen hervorgehoben

7. Methodisches Vorgehen

Für den gesamten Entwicklungs- und Evaluationsprozess des lea.Dashboards wurde ein methodisches Vorgehen umgesetzt, welches sich an das *vierstufige Prozessmodell zur lehrerzentrierten Dashboard-Entwicklung* von Ez-Zaouia (2020) orientiert. In der ersten Phase, der *Anforderungsanalyse*¹, wurden die Anforderungen an das zu entwickelnde Dashboard herausgearbeitet und systematisiert (Meyer 2024). In der Phase der *Ideenfindung*² wurden die Datenvisualisierungen im Dashboard gestalterisch entwickelt (Meyer n.d., eingereicht). Um diesen Entwicklungsprozess zu systematisieren und wissenschaftlich zu begleiten, wurde nach der Design-Science-Methodologie (Hevner u. a. 2004) vorgegangen. In der dritten Phase, der *Entwicklung*³, soll der interaktive Prototyp des lea.Dashboards (siehe Kapitel 6) im Rahmen einer Usability-Studie getestet werden. Die Ergebnisse dieser Phase bilden den Schwerpunkt dieses Artikels. Anschließend wird das

¹ Im Original heißt die Phase Situate (Verortung). Sie wurde aufgrund kleinerer Anpassungen umbenannt (vgl. Meyer 2024, S. 527)

² Im Original heißt die Phase Ideate, eigene Übersetzung ins Deutsche

³ Im Original heißt die Phase Develop, eigene Übersetzung ins Deutsche

Dashboard programmiert und in der vierten Phase, der *Evaluation*⁴, im praktischen Einsatz getestet. Die Ergebnisse dieser Phase werden in einem separaten Artikel veröffentlicht.

In dieser Studie wird eine Think-Aloud-Testung mit einem klickbaren Prototyp durchgeführt. Ein interaktiver Prototyp ermöglicht realitätsnahe Interaktionen, mit denen sich Navigation und Interaktionsmuster prüfen lassen (Schmidt, Earnshaw, u. a. 2020). An der Testung nahmen insgesamt zwölf Personen teil. Sechs der teilnehmenden Lehrkräfte waren Berufsschullehrkräfte (vier Frauen und zwei Männer), sechs waren Lehrkräfte aus Alphabetisierungskursen (drei Männer und drei Frauen). Alle Lehrkräfte hatten bereits Berufserfahrung. Das Alter der Teilnehmenden wurde nicht erfasst. Die Lehrkräfte aus den Alphabetisierungskursen hatten sich eigenständig auf einen Aufruf gemeldet, der an alle Volkshochschulen mit Grundbildungskursen in Norddeutschland gesendet wurde. Die Berufsschullehrkräfte wurden über bestehende Kooperationskontakte im Projekt lea.online akquiriert.

Im Rahmen der Think-Aloud-Testung soll die Benutzerfreundlichkeit des Dashboards geprüft werden. Weiterhin sollen Erkenntnisse darüber gewonnen werden, ob und welche Informationen die Lehrenden in den Datenvisualisierungen erkennen. Darüber hinaus sollen Erkenntnisse dazu gewonnen werden, wie die Lehrenden das Dashboard in ihrem beruflichen Alltag nutzen würden. Während der Testung mussten die Teilnehmenden sechs Aufgaben erfüllen, die Schlüsselinteraktionen mit der Software darstellen (siehe Tabelle 1). Diese beziehen sich auf das Aufrufen der Ergebnisse eines gesamten Kurses, das Auswählen einer bestimmten Kompetenz aus der Listenansicht, das Navigieren zur Personenansicht aus der Kursansicht heraus, das Filtern der einzelnen Kompetenzen (die als Karten dargestellt sind), sowie das Umdrehen einer Karte, um zusätzliche Erklärungen darzustellen. Mithilfe dieser Schlüsselinteraktionen wird geprüft, ob die Probanden grundsätzlich in der Lage sind die Software eigenständig zu bedienen und welche Usability-Probleme sich aufzeigen. Der Anmeldeprozess wurde nicht in die Testung integriert, da zum Zeitpunkt der Testung lediglich ein interaktiver Prototyp zur Verfügung stand, der keine textbasierten Nutzereingaben verarbeiten konnte. In der vorliegenden Untersuchung wird zusätzlich die Rückmeldung der Lehrenden zu den Datenvisualisierungen erfasst. Damit soll der Herausforderung begegnet werden, dass Lehrende Schwierigkeiten im Umgang mit Datenvisualisierungen in Dashboards zeigen. Dazu werden die Lehrenden während der Think-Aloud-Testungen bei jeder Datenvisualisierung aufgefordert, zu beschreiben, welche Informationen sie erkennen und wie sie diese interpretieren. Um Einblicke in die Gesamtbewertung des Systems durch die Lehrenden und ihre Nutzungsmöglichkeiten zu gewinnen, werden am Ende der Think-Aloud-Testung zusätzlich Fragen zu diesen Aspekten gestellt, sofern diese nicht explizit von den Proband:innen vorab genannt wurden (siehe Tabelle 1). Mit Frage (a) soll ein holistisches Urteil der Teilnehmenden zum lea.Dashboard eingeholt werden. Frage (b) bezieht sich ebenfalls auf den Gesamteindruck der Anwendung und fokussiert auf Design, Interaktionsabläufe und Erlernbarkeit. Sie begründet sich in den Usability-Heuristiken „Aesthetic and Minimalist Design“ und „Flexibility and Efficiency of Use“ (Nielsen 2024). Weiterhin wird die Frage implizit aus der Usability-Definition der DIN abgeleitet, da Nutzende nur dann effektiv, effizient und zufriedenstellend mit einem neuen System interagieren können, wenn dieses gut erlernbar ist (International Organization for

⁴ Im Original heißt die Phase Evaluate, eigene Übersetzung ins Deutsche

Standardization 2018). Frage (c) zur Nutzung des Filtersystems begründet sich durch Nielsens siebte Heuristik "Flexibility and Efficiency of Use" (Nielsen 2024), da das Filtersystem als zentrales Interaktionselement auf der Personenseite die Effizienz und Nutzbarkeit der Auswertungsergebnisse verbessern soll. Die Fragen (d) zum Mehrwert der diagnostischen Auswertungen und (e) zu den Einsatzmöglichkeiten im beruflichen Alltag gehen über klassische Usability-Kriterien hinaus. Sie leiten sich einerseits aus den Forschungsfragen und andererseits aus dem multidimensionalen Ansatz des Learning Experience Designs (LXD) (Janke 2022) ab. Demnach sollen neben der technischen Dimension (Usability und User Experience) auch die pädagogische Dimension (konkret: der Mehrwert und der Nutzen der diagnostischen Auswertungen (Frage (d)) und die sozio-kulturelle Dimension (konkret: die Einsatzmöglichkeiten im beruflichen Alltag, Frage (e)) berücksichtigt werden.

Aufgaben	Mögliche Äußerungen	Mögliche Interaktionen im Prototyp
Auf der Startseite		
	Äußerungen zur Erscheinung der gesamten Seite	Hovern über alle Buttons möglich (Schaltflächen verändern ihre Erscheinung)
Bitte schauen Sie sich die Ergebnisse des gesamten Kurses {{Kursname}} an. (Aufgabe 1)	Äußerungen zu Listen der Kurse und Lernenden	Navigation über Klick auf den Vorschau-Button oder den Titel des Kurses möglich.
Auf der Kursseite		
	Äußerungen zur Erscheinung der gesamten Seite	Hovern über alle Buttons möglich (Schaltflächen verändern ihre Erscheinung)
(Bitte schauen Sie sich die Verlaufsgrafik und die Informationen, welche in dieser bereitgestellt werden an und beschreiben diese)	Äußerungen zur Darstellung der Verlaufsgrafik	Hovern über die einzelnen Balken der Grafik
Bitte schauen Sie sich die Ergebnisse aller Kursteilnehmenden zur Kann-Beschreibung {{Name der KB}} an. (Aufgabe 2)	Äußerungen zur Darstellung der Listenansicht	Scrollen durch die Listen, Öffnen der Detailliste durch Klick auf den Pfeil, den Namen der Kann-Beschreibung oder den gefüllten Balken möglich.
Bitte rufen Sie die Einzelergebnisse der Teilnehmenden {{Name der	Äußerungen zum rechten Seitenmenü mit den einzelnen	Durch Klick auf die entsprechende Person im rechten Seitenmenü möglich

Teilnehmenden}} auf. (Aufgabe 3)	Teilnehmenden eines Kurses	
Auf der Personenseite		
	Äußerungen zur Erscheinung der gesamten Seite	Hovern über alle Buttons möglich (Schaltflächen verändern ihre Erscheinung)
Bitte filtern Sie alle Kompetenzen des Lexikalischen Prinzip, die den Alpha-Leveln 1-4 entsprechen und zum ersten Mal getestet wurden. (Aufgabe 4)	Äußerungen zu den Filtern und der Darstellung der Karten	Durch Klick auf die Titel der Filter oder Radio-Buttons möglich
Bitte schauen Sie sich alle Informationen/Erklärungen zur Kompetenz {{Name der Kompetenz}} an. (Aufgabe 5)	Äußerungen zu der Darstellung der Karten und den Informationen auf den Karten	Umdrehen der Karte durch Klick die Karte möglich.
Bitte setzen Sie alle Filter zurück. (Aufgabe 6)	Äußerungen zu den Filtern	Durch Klick auf den Button Filter zurücksetzen möglich.
Anschließende Fragen: (a) Wie bewerten Sie den Gesamteindruck? (b) Wie bewerten Sie das Design, die Interaktionsabläufen sowie die Erlernbarkeit des Tools? (c) Wie bewerten Sie die Nutzung des Filtersystems? (d) Inwieweit sehen Sie einen Mehrwert und Nutzen in den diagnostischen Auswertungen? (e) Welche Einsatzmöglichkeiten im beruflichen Alltag können Sie sich vorstellen?		

Tabelle 1: Ablauf, Aufgaben und Fragen des Think-Aloud-Tests

Aufgrund der geltenden Kontaktregelungen während der Corona-Pandemie wurden die Think-Aloud-Testungen online mit dem Videokonferenz-Tool Zoom durchgeführt. Die zwölf Tests wurden als Bildschirmaufzeichnung aufgenommen und anschließend transkribiert. Zu Beginn der Testung wurde allen Teilnehmenden die „Think Aloud“-Methode erklärt und ihnen wurde erzählt, dass es sich um ein Dashboard zur Auswertung der Testergebnisse des otu.lea-Testes handelt. Lehrende, die otu.lea nicht kannten, wurden kurz darüber informiert, was mit diesem Test erfasst werden kann. Zusätzlich wurde angekündigt, dass am Ende der Testung weitere Leitfragen gestellt werden. Wenn Teilnehmende ihre Gedanken nur wenig aktiv äußerten, forderten die Interviewer sie erneut dazu auf. Alle Probanden nahmen zum ersten Mal an einer Usability-Testung teil und waren zum Teil etwas nervös.

In der Usability-Forschung existieren keine einheitlichen wissenschaftlichen Standards zur Auswertung von Think-Aloud-Protokollen. Deshalb werden unterschiedliche Verfahren zur Auswertung der Gesprächsprotokolle genutzt (Doi 2021; Hornbæk 2010). Zur Auswertung der Gesprächsprotokolle wurde ein deduktiv-induktives Vorgehen entsprechend der inhaltlich-strukturierender Inhaltsanalyse (Kuckartz 2018) gewählt, das um die Berücksichtigung etablierten Heuristiken aus der Usability-Forschung ergänzt wurde. Die Transkripte wurden mithilfe der Software MAXQDA kodiert. Für die Erarbeitung des Kategoriensystems wurden nach einer ersten Sichtung des Materials die Usability-Heuristiken von Nielsen (Nielsen 2024) sowie die Forschungsfragen zugrunde gelegt. Aufgrund der eingeschränkten Funktionalität des Prototyps wurden nur die Heuristiken integriert, die sinnvoll abbildbar waren. Da der Prototyp keine Möglichkeiten zum Abbruch oder Verlassen eines Vorgangs sowie keine Fehlermeldungen enthielt, wurden die Heuristiken „User Control and Freedom“, „Error Prevention“ und „Help Users Recognize, Diagnose, and Recover from Errors“ nicht erfasst. In Bezug auf Fehlervermeidung liefert die Auswertung der einzelnen Aufgaben Hinweise auf problematische Bereiche im Design. Da der Prototyp noch keine zusätzlichen Unterstützungsbereiche oder Tutorials enthielt, wurde auch die Heuristik „Help and Documentation“ nicht integriert. Aufbauend auf den deduktiv abgeleiteten Kategorien, wurde das Kategoriensystem während der Kodierung induktiv überarbeitet und erweitert. Das entwickelte Kodiersystem wurde im Sinne einer konsensuellen Validierung mit dem Projektteam diskutiert und argumentativ abgesichert. Anschließend schloss die Autorin die erste Kodierungsphase ab. Im Rahmen des Verfassens dieses Artikels wurde eine erneute Sichtung des Materials durchgeführt, um dessen Stabilität zu überprüfen (Intra-Coder-Übereinstimmung). Dabei wurden die Kodierungen der Oberkategorien **Design**, **Usability** und **Datenauswertung und -interpretation** bestätigt. In der Oberkategorie **Einsatzszenarien** wurden kleinere Anpassungen vorgenommen. Die Bewertung der Aufgaben, welche die Lehrenden mit dem Prototyp zu erfüllen hatten, erfolgte auf Basis der Sichtung der Bildschirmaufzeichnungen. Die Operationalisierung der Aufgaben-Erfüllung erfolgte induktiv. Hierzu hat die Autorin die Bildschirmaufzeichnungen in einem iterativen Vorgehen mehrfach gesichtet. Die Ergebnisse wurden im Anschluss im Projektteam vorgestellt und diskutiert. Parallel zum Verfassen des Artikels erfolgte eine erneute Kontrolle der ausgewerteten Bildschirmaufzeichnungen. Dabei wurden alle Aufgaben, die als „Durchführungen mit Nachfragen“, „fehlerhafte Durchführungen“ oder „Durchführungen mit technischen Problemen“ bewertet wurden, erneut geprüft. Es wurde lediglich eine neue Zuordnung von „Durchführung mit Nachfragen oder Umwegen“ zu „Durchführung ohne Fehler oder Umwege“ bei Aufgabe 5 vorgenommen.

8. Ergebnisse

Im Folgenden werden die Ergebnisse der Auswertung gegliedert nach den Forschungsfragen beschrieben.

FF1: Welche Usability-Probleme treten bei der Nutzung des Dashboards auf und wie bewerten die Nutzer:innen die Usability des Dashboards?

In Tabelle 2 ist eine Auflistung der zu erledigenden Aufgaben und der Ergebnisse zu finden, die angibt, in welcher Form die Proband:innen die Aufgaben erfüllt haben. Die Zahl 1 steht

für eine Durchführung ohne Fehler oder Umwege, die Zahl 2 für eine Durchführung mit Nachfragen oder Umwegen und die Zahl 3 für eine fehlerhafte Durchführung oder eine Durchführung nur mit Hilfe. Freie Tabellenzellen bedeuten, dass keine Durchführung aufgrund technischer Probleme möglich war. Zur Aufgabe 1 gab es die Instruktion „Bitte schauen sie sich die Ergebnisse des gesamten Kurses {{Kursname}} an.“ Diese Aufgabe konnte neunmal fehlerfrei absolviert werden. Dreimal erfolgte die Durchführung mit Nachfragen oder Umwegen. Die Teilnehmenden vergewisserten sich, ob sie wirklich den Vorschau-Button klicken sollen, obwohl sie diesen eigentlich selbst auswählen wollten. Die Nervosität einiger Personen zu Beginn der Testung kann das zusätzliche Nachfragen erklären. Daher wird die Interaktion insgesamt als problemlos betrachtet. Bei Aufgabe 2 sollten sich die Lehrenden die Ergebnisse aller Kursteilnehmenden zu einer bestimmten Kann-Beschreibung in der Listenansicht anschauen. Dies wurde zehnmal fehlerfrei und zweimal mit Umwegen oder Nachfragen durchgeführt. Zwei Berufsschullehrende fanden die Kann-Beschreibung nicht ohne Hilfe, führten die Interaktion zum Aufklappen der Einzelergebnisse dann aber direkt korrekt aus. Somit wird auch diese Interaktion als problemlos gewertet. Die mögliche Unübersichtlichkeit der Listen wird für die weitere Auswertung dennoch berücksichtigt. Bei Aufgabe 3 mussten die Lehrenden eine bestimmte Einzelperson aus dem seitlichen Menü auf der rechten Seite auswählen. Dies wurde achtmal fehlerfrei durchgeführt. Dreimal verdeckte ein Fenster des Videokonferenzprogramms die Stelle des seitlichen Menüs, aus der die Person ausgewählt werden musste. Daher fanden die Teilnehmenden die Person zuerst nicht. Ein:e Lehrende:r wählte die Interaktion korrekt aus, vergewisserte sich jedoch vorab, ob dies korrekt sei. Dies wird als Durchführung mit Nachfragen oder Umwegen gewertet. Insgesamt wird auch diese Aufgabe als problemlos gewertet.

Die Testteilnehmenden hatten vermehrt Probleme bei der Benutzung des Filtersystems. So wurde die Aufgabe 4 zum Filtern einzelner Kompetenzen auf der Personenansicht nur fünfmal korrekt durchgeführt. Viermal erfolgte die Nutzung der Filter fehlerhaft und dreimal war nur durch Umwege oder Nachfragen eine Nutzung möglich. Mehrere Personen waren sich unsicher, ob sie einen Filter auswählen oder nur zu den entsprechenden Karten scrollen sollen. Zudem wurde teilweise nicht erkannt, dass die Liste der Filter scrollbar ist oder ein Filter wurde trotz Betrachten der Filterliste nicht gesehen. Eine weitere Person scrollte durch die Liste der Filter und fand dann den auszuwählenden Filter anschließend nicht mehr wieder. Das Zurücksetzen der Filter (Aufgabe 6) war ebenfalls dreimal nur durch Umwege oder Nachfragen möglich, die Bedienung war jedoch nicht fehlerhaft. Entweder vergewisserten sich die Teilnehmenden, ob der Button „Zurücksetzen“ der korrekte Weg zum Zurücksetzen der Filter ist, oder sie wollten alle Filter einzeln zurücksetzen, was als Umweg gewertet wurde. Interessant ist hier zu erwähnen, dass die Proband:innen das Filtersystem gleichzeitig als besonders hilfreich und, nachdem anfängliche Verständnisprobleme behoben waren, als einfach zu bedienen beschrieben haben. Technische Probleme, die dazu führten, dass die Aufgabe nicht eigenständig von Lehrenden durchgeführt werden konnte, gab es beim Zurücksetzen der Filter dreimal. Bei einer Lehrkraft aus einem Alphabetisierungskurs gab es zudem Darstellungsfehler des Prototyps, deren Ursache nicht geklärt werden konnte. Daher konnten die letzten beiden Aufgaben (5 und 6) nicht mehr durchgeführt werden. Aufgabe 5 wurde elfmal fehlerfrei umgesetzt. Einmal gab es wie oben beschrieben technische Schwierigkeiten, sodass auch diese Aufgabe als fehlerfrei gewertet werden kann. Insgesamt kamen die Berufsschullehrkräfte etwas besser mit dem Dashboard zurecht. Sie

führten 27 Aufgaben fehlerfrei durch, hatten fünf Durchführungen mit Umwegen oder Nachfragen und eine fehlerhafte Durchführung. Die Lehrenden aus Alphabetisierungskursen hatten 22 fehlerfreie Aufgabendurchführungen, sieben Durchführungen mit Nachfragen oder Umwegen und drei fehlerhafte Durchführungen.

	Aufgabe 1	Aufgabe 2	Aufgabe 3	Aufgabe 4	Aufgabe 5	Aufgabe 6
BSL1	1	1	1	1	1	2
BSL2	2	2	1	2	1	1
BSL3	1	2	1	1	1	1
BSL4	1	1		3	1	1
BSL5	1	1	1	1	1	
BSL6	1	1	1	1	1	
AL1	1	1	1	2	1	2
AL2	1	1	1	3		
AL3	2	1	1	1	1	1
AL4	1	1		2	1	2
AL5	2	1	2	3	1	1
AL6	1	1		3	1	1

Tabelle 2: Ergebnisübersicht der einzelnen Aufgaben

1 = Durchführung ohne Fehler oder Umwege, 2 = Durchführung mit Nachfragen oder Umwegen, 3 = Durchführung fehlerhaft oder nur mit Hilfe möglich, leer = Keine Durchführung aufgrund von technischen Problemen

Anschließend wurden die Gesprächsprotokolle ausgewertet. Die Kategorie Usability wurde auf Grundlage der Usability-Heuristiken nach Nielsen (Nielsen 2024) gebildet. Im ersten Schritt der Erarbeitung des Codebuches, wurden die gewählten Heuristiken mit aufgenommen und im Verlauf der Kodierung induktiv angepasst. Eine genaue Definition der jeweiligen Codes ist in der folgenden Tabelle zu finden (siehe Tabelle 3).

Kategorie	Code	Beschreibung	Häufigkeit
Usability	Selbstbeschreibungsfähigkeit	Aussagen von Nutzende, ob sie in der Lage sind die zur Bedienung des Systems relevanten Informationen zu finden und zu verstehen sowie erkennen können, ob und wann das System Aktionen ausführt. Dieser Aspekt wird in der ersten Heuristik von Nielsen „Visibility of system status“ formuliert	26

		(Nielsen 2024).	
	Erwartungskonformität	Aussagen von Nutzenden, wie konsistent der Einsatz von z.B. Symbolen beschrieben wird und ob das Verhalten der Software den Erwartungen entspricht. Dieser Aspekt wird in der vierten Heuristik von Nielsen „Consistency and standards“ formuliert (Nielsen 2024).	22
	Effiziente Benutzung	Der Code wurde deduktiv-induktiv erstellt. Er umfasst Aussagen der Nutzenden welche Aspekte und Interaktionspfade sie in der Anwendung als effizient beschreiben. In der siebten Heuristik „Flexibility and efficiency of use“ (Nielsen 2024) wird beschrieben, dass Nutzer:innen verschiedene Wege (z.B. für erfahrende und unerfahrene Nutzer:innen) zur Interaktion mit der Software angeboten werden sollen. Die Nutzenden äußerten nur Teilaspekte dieser Kategorie, weshalb sie mit abgewandelter Beschreibung mitaufgenommen wurde.	8
	Unpassendes Wording	Unpassendes Wording umfasst Aussagen der Nutzenden zur missverständlichen Verwendung von Vokabular in der Software. Die Kategorie wurde induktiv, aufbauend auf der zweiten Heuristik „Match between system and the real world“ (Nielsen 2024), weiterentwickelt.	7
	Einarbeitung	Der Code Einarbeitung basiert auf der Heuristik „Recognition rather than recall“ (Nielsen 2024) und beinhaltet Aussagen der Lehrenden zur Intuitivität und Aufwand der Einarbeitung in das System.	34
Design	Übersichtlichkeit des Designs	Der Code „Übersichtlichkeit des Designs“ basiert (wie „Attraktivität des Designs“) auf der achten Usability Heuristik „Aesthetic and minimalist design“ (Nielsen 2024). Um zwischen positiven Aussagen zum Design und der expliziten Beschreibung von Übersichtlichkeit zu differenzieren, wurde die Heuristik in zwei Kategorien aufgeteilt. Der Code umfasst alle Aussagen, wie	29

		übersichtlich und klar das Design ist.	
	Attraktivität des Designs	Der Code „Attraktivität des Designs“ basiert (wie „Übersichtlichkeit des Designs“) auf der achten Usability Heuristik „Aesthetic and minimalist design“ (Nielsen 2024). Um zwischen positiven Aussagen zum Design und der expliziten Beschreibung von Übersichtlichkeit und Minimalismus zu differenzieren, wurde die Heuristik in zwei Kategorien aufgeteilt. Der Code umfasst alle Aussagen, die dazu getroffen wurden, wie attraktiv die Nutzenden das Design finden.	11
	Verbesserungsvorschläge Design	Der Code „Verbesserungsvorschläge Design“ ergab sich induktiv aus der Durchsicht des Materials.	12

Tabelle 3: Kategorien und Codes zu FF1

Selbstbeschreibungsfähigkeit: Während der Usability-Testung wurden durch die Gedankenäußerungen der Nutzer:innen insbesondere auf der Startseite, welche eine Übersicht aller Kurse und Lernenden anzeigt, Probleme mit der Selbstbeschreibungsfähigkeit des Systems deutlich. Die Nutzer:innen erkannten zwar den Aufbau von Klassen und Teilnehmer:innen auf der Startseite, verstanden aber nicht warum Teilnehmende zusätzlich zu den Klassen einzeln aufgelistet sind und welcher Zusammenhang zwischen diesen besteht (siehe dazu auch Abbildung 2). Um die Verständlichkeit zu verbessern, sollte das Design hier angepasst werden. Weiterhin traten Verständnisprobleme bei den Listenansichten (siehe dazu auch Abbildung 5) der gruppierten Kann-Beschreibungen auf der Kursübersichtsseite auf. In diesen Listen waren auch die Informationen zum durchschnittlichen Alpha-Level der dargestellten Kann-Beschreibungen enthalten. Dies wurde von den Teilnehmer:innen entweder nicht verstanden oder als überflüssig bewertet.

Erwartungskonformität: Die Symbole auf den Buttons (z. B. Bearbeiten, Ansehen, Drucken) in unterschiedlichen Bereichen des User Interfaces wurden von den Testteilnehmenden als konsistent empfunden. Obwohl Fehler bei der Nutzung des Filtersystems auftraten (s.o. Auswertung der Aufgaben), wurde dieses von zwei Nutzenden als besonders intuitiv beschrieben, da es sie an gängige Online-Tools erinnerte.

Effiziente Benutzung: Aussagen zur effizienten Benutzung wurden nur von Lehrenden aus Alphabetisierungskursen getätigt. Diese umfassen den direkten Überblick über alle Lernenden bei der Verlaufsgrafik sowie die Möglichkeit, per Mouseover weitere Informationen sichtbar zu machen. Weiterhin wurde erwähnt, dass die Auflistung der Testergebnisse nach zeitlichem Verlauf praktisch ist. Das direkte Hin- und Herklicken zwischen Kurs- und Einzelansicht, zusätzlich zur Auswahl über die Startseite, wird als unkompliziert und effizient wahrgenommen.

Unpassendes Wording: In Bezug auf das im Dashboard verwendete Fachvokabular wurden Verständnisprobleme geäußert. Dies betraf insbesondere die Bezeichnung der

Gruppen der einzelnen Kann-Beschreibungen im Bereich Schreiben (z. B. Syllabisches Prinzip) sowie die Formulierung einzelner Kann-Beschreibungen (z. B. Kann Silben, die lediglich aus einem Vokal oder Diphthong bestehen, schreiben). Diese Anmerkungen wurden ausschließlich von Berufsschullehrenden gemacht.

Einarbeitung: Zehn der zwölf Personen beschrieben das Dashboard als gut erlernbar sowie als einfach und intuitiv zu bedienen. Gleichzeitig wurde von einigen Personen erwähnt, dass die Fülle an Informationen, die im Dashboard zur Verfügung gestellt werden, eine gewisse Einarbeitungsphase erfordert.

Übersichtlichkeit des Designs: Acht Teilnehmende (sechs Berufsschullehrende und zwei Lehrende aus Alphabetisierungskursen) beschrieben das Dashboard als „übersichtlich“, „klar unterteilt“ und „eindeutig“. Gleichzeitig wurde von sieben Teilnehmenden (vier Berufsschullehrende und drei Lehrende aus Alphabetisierungskursen) angemerkt, dass die Fülle an Informationen, die dargestellt werden, eine gewisse Eingewöhnungszeit erfordert. Vier Lehrkräfte beschrieben beide Aspekte, was auf eine ambivalente Wahrnehmung hindeutet. Mehrere Teilnehmende äußern, dass sie es nachvollziehen können, dass so viele Informationen dargestellt werden. Sie halten dies zum Teil für unumgänglich. Dennoch empfinden sie ein gewisses Gefühl von der Fülle der Informationen erschlagen zu werden.

Attraktivität des Designs: Drei Berufsschullehrende und drei Lehrende aus Alphabetisierungskursen heben attraktive Aspekte des Designs hervor. Positiv bewertet werden das zurückhaltende Design und die übersichtliche Struktur. Ebenfalls positiv bewertet wird die Schriftgröße, die Farbgebung (Orange und Grautöne) sowie der Gesamteindruck.

Verbesserungsvorschläge Design: Eine Lehrkraft aus einem Alphabetisierungskurs merkte an, dass sie bei der Darstellung der Ergebnisse der Lernenden (voll erfüllt, beinahe erfüllt, teilweise erfüllt, noch nicht erfüllt) eine farbliche Abstufung von Grün (erfüllt) nach Rot (nicht erfüllt) bevorzugen würde. Weiter wurde angeregt, das Design etwas farbenfroher zu gestalten und auf die Farbe Orange zu verzichten und durch Grün zu ersetzen. Ein:e Teilnehmende:r hatte Probleme mit der Helligkeit der Schrift und schlug vor eine noch dunklere Schriftfarbe zu verwenden. Zur Verbesserung der Übersichtlichkeit wurde angeregt, die Listen der Kann-Beschreibungen auf der Kursübersichtsseite einzuklappen und sie für die Nutzenden manuell ausklappbar zu machen.

Zusammenfassung

Die grundlegenden einfachen Interaktionen konnten von beiden Zielgruppen – den Berufsschullehrenden sowie den Lehrenden aus Alphabetisierungskursen – nahezu fehlerfrei durchgeführt werden. Probleme gab es hingegen bei der Nutzung des Filtersystems. Die Lehrkräfte beschrieben das Filtersystem nach kurzer Hilfestellung oder Einarbeitung jedoch als besonders intuitiv und nützlich, weshalb von einer grundlegenden Überarbeitung des Filtersystems abgesehen wird. Dennoch sollte die Übersichtlichkeit der Filter in einem weiteren Designzyklus überarbeitet werden. Negativ wurde zudem die hohe Informationsdichte der Anwendung bewertet. Dies sollte bei einer nachfolgenden Überarbeitung des Designs unbedingt berücksichtigt werden. Ein Ansatz könnte darin bestehen, die Listen der Kann-Beschreibungen in der Kursansicht standardmäßig einzuklappen. Zudem kann die Anzeige des durchschnittlichen Alpha-Levels der in den Listen enthaltenen Kann-Beschreibungen entfernt werden, da sie von den Nutzenden entweder nicht verstanden oder als überflüssig beschrieben wurde. Einige Nutzende hatten

Probleme mit der im Dashboard verwendeten Fachsprache. Für eine zielgruppenspezifische Weiterentwicklung des Dashboards sollte das Wording überarbeitet werden. Sollte dies aus sprachwissenschaftlichen Gründen abgelehnt werden, könnte eine umfassende Handreichung mit Erklärungen für fachfremde Nutzende zur Verfügung gestellt werden. Zudem könnten Direkthilfen in der Software unterstützen. Es sollten außerdem kurze Tutorials oder kontextsensitive Hilfen in Erwägung gezogen werden, um die Einarbeitungszeit so kurz wie möglich zu halten. Von einer farblichen Anpassung der Ergebnisse mit einem Farbverlauf von Grün nach Rot wird abgesehen. Es wurde von dieser Farbgebung bewusst Abstand genommen, da sie deutlich mit „gut“ und „schlecht“ konnotiert ist. Dies könnte sich negativ auf die Lernende auswirken, wenn sie die Testergebnisse gemeinsam mit der Lehrkraft betrachten (Meyer n.d., eingereicht). Weitere Rückmeldungen zum Design waren heterogen. Einerseits wurde das Design als ansprechend empfunden. Andererseits wurde angeregt, das Design etwas farbenfroher zu gestalten und eine dunklere Schriftfarbe zu verwenden. Bei einer nachfolgenden Überarbeitung des Designs sollten diese Detailaspekte berücksichtigt werden.

FF2: Inwieweit sind die Lehrkräfte in der Lage, die im Dashboard dargestellten diagnostischen Informationen korrekt zu beschreiben und zu interpretieren?

Die Kategorie Datenauswertung und -interpretation wurde auf Basis der Forschungsfrage „Inwieweit sind die Lehrkräfte in der Lage, die im Dashboard dargestellten diagnostischen Informationen korrekt zu beschreiben und zu interpretieren?“ gebildet. Die Codes sind nach mehrschrittiger Durchsicht des Materials induktiv entstanden (s. Tabelle 4).

Kategorie	Code	Beschreibung	Häufigkeit
Datenauswertung und -interpretation	Analytisch-beschreibende Auswertung der Datenvisualisierungen	Aussagen der Teilnehmende, in denen sie 1:1 beschreiben, was sie in den Datenvisualisierungen sehen (z.B. der Balken ist halb gefüllt) sowie Aussagen, in denen die Teilnehmenden das Gesehene analysieren (z.B. Lotte Scheller hat das besser gemacht als Simone Fischer)	36
	Intuitives Ableiten von Förder- und Handlungsmöglichkeiten	Aussagen der Teilnehmenden, in denen sie die Datenvisualisierungen interpretieren und weitere Schritte im Umgang mit den Ergebnissen beschreiben (z.B. dann kann ich Lotte Scheller noch weiteres Material zu XY zur Verfügung stellen oder dann können	8

		Lotte Scheller und Simone Fischer gemeinsam an XY arbeiten)	
	Falsche oder fehlerhafte Beschreibung, Analyse oder Interpretation	Aussagen der Teilnehmenden, in denen sie fehlerhafte Beschreibungen/ Analysen/ Interpretationen zu den Datenvisualisierungen nennen.	15

Tabelle 4: Oberkategorien und Kategorien zu FF2

Analytisch-beschreibende Auswertung der Datenvisualisierungen: Alle Testpersonen haben zu den Auswertungen und Datenvisualisierungen analytisch beschreibende Aussagen gemacht. Die Aussagen umfassen das Erfassen von Testergebnissen auf Alpha-Level-Ebene einzelner Lernender in der Verlaufsgrafik sowie von Ergebnissen, bezogen auf einzelne Kann-Beschreibungen, in den gruppierten Listen auf der Kursübersichtsseite. Dabei gelingt es den Lehrkräften, zwischen den Einzelergebnissen der Lernenden und den Gruppenergebnissen des Kurses zu unterscheiden. Zudem können die Lehrkräfte den Zusammenhang zwischen den Ergebnissen der Kursübersichtsseite und der Einzelansicht von Lernenden herstellen. Es wird erkannt, dass die Fortschrittsbalken auf den Karten und in den Listen anzeigen, ob eine Kompetenz erfüllt ist. Ebenso wird in der Verlaufsgrafik erkannt, dass Ergebnisse verschiedener Lernender über einen zeitlichen Verlauf hinweg visualisiert werden. In der Einzelansicht der Lernenden gelingt es den Lehrenden, die Leistungen bezogen auf die einzelnen Kann-Beschreibungen zu beschreiben. Durch Anwenden der Filter wird erkannt, ob Kompetenzen sich verbessert oder verschlechtert haben. Die Ergebnisse zeigen, dass die Testteilnehmenden grundsätzlich in der Lage sind basale Informationen und Zusammenhänge der Datenvisualisierungen in der Verlaufsgrafik und den Listenansichten auf der Kursübersichtsseite sowie Detailinformationen in der Einzelansicht der Lernenden zu erkennen.

Intuitives Ableiten von Förder- und Handlungsmöglichkeiten: Sechs der zwölf Lehrkräfte – zwei Berufsschullehrende und vier Lehrende aus Alphabetisierungskursen – machten korrekte Angaben zu Interpretationen der Auswertungen. Es wurden Aussagen dazu getroffen, wie die Lehrkräfte die Ergebnisse im Dashboard nutzen können, um ihren Unterricht zu planen und im Kursverlauf adaptiv anzupassen. Auf Grundlage der Auswertungen im Dashboard können Aufgaben einzeln oder gruppenspezifisch gestellt sowie individuell differenziert werden. Das Dashboard bietet außerdem Unterstützung dabei, Lernenden fundiert Rückmeldung über deren Stärken, Schwächen und Lernverläufe zu geben. Die interpretierenden Aussagen zu den Datenvisualisierungen wurden mehrheitlich von den Lehrkräften aus den Alphabetisierungskursen getätigt. Dies könnte durch ihre stärkeren Praxiserfahrungen mit Menschen mit geringer Literalität begründet sein.

Falsche oder fehlerhafte Beschreibung, Analyse oder Interpretation: Vier Berufsschullehrkräfte machten fehlerhafte Beschreibungen und Interpretationen zu den Datenvisualisierungen. Dies lässt sich auf Verständnisprobleme der Testteilnehmenden bezüglich des Aufbaus der Kompetenzmodelle zurückführen. So konnten die Berufsschullehrkräfte beispielsweise nicht unterscheiden, dass das Gesamtergebnis

einer Lernenden bzw. eines Lernenden in die Alpha-Levels 1 bis 5 eingestuft wird und alle Kann-Beschreibungen entsprechend ihrer Anforderungen ebenfalls den Levels eins bis fünf zugeordnet sind. Weiterhin nahmen die Berufsschullehrenden fälschlicherweise an, dass eine Kann-Beschreibung je nach Leistung der Lernenden auf verschiedenen Alpha-Levels erfüllt sein kann. In diesem Zusammenhang führte auch die Angabe des durchschnittlichen Alpha-Levels der gruppierten Kann-Beschreibungen in der Listenansicht auf der Kursübersichtsseite zu Verwirrungen. Zwei Lehrende aus Alphabetisierungskursen äußerten Unverständnis über die Alpha-Level-Aufteilungen innerhalb der gruppierten Kann-Beschreibungen. Dies ließ ebenfalls auf fehlerhafte mentale Modelle der Kompetenzdiagnostik schließen.

Zusammenfassung

Alle Teilnehmenden konnten rein beschreibende Aussagen zu den diagnostischen Informationen bzw. Datenvisualisierungen treffen. Förderpädagogische Schlüsse zur Förderung von Einzelpersonen oder Gruppen, wurden hingegen mehrheitlich von Lehrenden aus Alphabetisierungskursen gezogen. Dies ist vermutlich auf die mangelnde Erfahrung der Berufsschullehrenden bezüglich der Förderung von Menschen mit geringer Literalität zurückzuführen. Darüber hinaus machten die Berufsschullehrenden deutlich mehr fehlerhafte Aussagen zu den diagnostischen Informationen. Dies kann auf fehlendes Fachwissen zu den Kompetenzmodellen und der Diagnostik zurückgeführt werden. Für den Einsatz an Berufsschulen sollten demnach spezielle Schulungsangebote zum förderpädagogischen Einsatz des lea.Dashboards entwickeln werden. Einige Lehrende aus Alphabetisierungskursen interpretierten die dargestellten Informationen ebenfalls fehlerhaft. Eine klarere Darstellung der Struktur der Kompetenzmodelle (und Gruppierungen) sowie der Testsets verbunden mit Informationen zum Kompetenzmodell könnte Verwirrungen, Fehler und Unsicherheiten auch bei geübten Lehrkräften aus Alphabetisierungskursen vermeiden.

FF3: Welche Nutzungsmöglichkeiten sehen die Lehrkräfte für ihre persönliche berufliche Praxis?

Die Kategorie Einsatzszenarien begründet sich in der Forschungsfrage „Entsprechen die Anwendungsszenarien des Dashboards den Möglichkeiten, Erwartungen und Bedarfen der Lehrenden?“. Aufgrund der offen formulierten Forschungsfrage wurden die Codes während der Kodierungsdurchläufe induktiv ermittelt (siehe Tabelle 5).

Kategorie	Code	Beschreibung	Häufigkeit
Usability	Nützlichkeit	Der Code Nützlichkeit wurde induktiv durch Sichtung des Materials erarbeitet und fasst von den Lehrkräften genannte, besonders nützliche Aspekte sowie Nutzungshürden zusammen.	56
Einsatzszenarien	Unterstützung des eigenen Präsenzunterrichts sowie Zusammenarbeit	Aussagen von Nutzenden, die sich direkt auf die Arbeit im Unterricht sowie die Unterrichtsvorbereitung	12

	mit Kolleg:innen	beziehen oder die Zusammenarbeit mit Kolleg:innen umfassen.	
	Rückmeldung/Feedback an Lernende	Aussagen von den Nutzenden, wie sie die Dashboard-Ergebnisse für Feedback an die Lernenden nutzen wollen.	5
	Einsatz über den Präsenzunterricht hinaus	Aussagen von den Nutzenden über den Einsatz des Dashboards, die sich auf die Arbeit über den Unterricht hinaus beziehen. Dies umfasst auch alle weiteren Aspekte, die nicht in den Bereichen Unterrichtsvorbereitung und Präsenzunterricht, Feedback an Lernende oder die Zusammenarbeit mit Kolleg:innen enthalten sind.	4

Tabelle 5: Oberkategorien und Kategorien zu FF3

Nützlichkeit: Die Berufsschullehrkräfte beschrieben insbesondere das Messen und Darstellen des Lernfortschritts der Schüler:innen, die Übersetzungen der Kann-Beschreibungen im Dashboard sowie das Filtersystem als nützlich. Als hilfreich beschrieben wurden außerdem die Unterteilung der Kann-Beschreibungen in Prinzipien sowie der differenzierte Überblick über die Stärken und Schwächen der Lernenden. Eine Berufsschullehrkraft beschrieb in der Bewertung des Gesamtkonzepts des Dashboards, dass er oder sie durch das Ausprobieren neugierig geworden sei, die Lese- und Schreibkompetenzen seiner oder ihrer Schüler:innen zu testen. Gleichzeitig befürchteten die Berufsschullehrenden, dass ihre meist nicht vorhandene Vorbildung im Bereich der Lese-, Schreib- und Sprachförderung zu einer zeitintensiven Einarbeitungsphase führt. Dabei wurden die Komplexität der Kompetenzmodelle sowie deren starker Grad an Differenzierung als besonders herausfordernde Faktoren beschrieben. In Verbindung mit dem häufig bestehenden Zeitmangel im Schulalltag und der in der Regel hohen Anzahl von Schüler:innen pro Klasse und Lehrkraft stellt dies eine deutliche Nutzungshürde dar. Als mögliche Ansatzpunkte, um dem entgegenzuwirken, wurden die Möglichkeit an Schulungen teilzunehmen, eine Anpassung des Lehrplans hin zur Integration von Lese-, Schreib- und Sprachförderungsthemen sowie eine Verankerung der Thematik in der Seminausbildung genannt.

Die Erläuterungen zu den Kann-Beschreibungen (auf der Rückseite der Karten) werden von zwei Lehrenden aus Alphabetisierungskursen hingegen als überflüssig beschrieben. Zwei andere hingegen bewerten sie als nützlich. Zudem könnten die Erklärungen im Dashboard eine Unterstützung für unerfahrene Lehrkräfte sein. Vier der Lehrenden aus Alphabetisierungskursen beschreiben den Nutzen des Dashboards darin, Zeit und kognitive

Ressourcen zu sparen. Die Diagnostik und das Auswerten der Testergebnisse werden bislang manuell durchgeführt oder aus Zeitmangel gar nicht umgesetzt.

Unterstützung des eigenen Präsenzunterrichts sowie Zusammenarbeit mit

Kolleg:innen: Eine Berufsschullehrkraft beschreibt, wie sie mithilfe der regelmäßigen Testergebnisse im Dashboard auch das Gelingen ihres eigenen Unterrichts reflektieren könnte. Lehrende aus Alphabetisierungskursen sehen Potenziale darin, einen guten kompetenzdiagnostischen Überblick über größere Kurse zu erhalten und neue Lernende in Kursen gleich zu Beginn des Kurses testen zu können. In Bezug auf die Unterrichtsvorbereitung wurden zwei zentrale Aspekte genannt: eine deutliche Zeitersparnis bei der Erstellung individueller Unterrichtsmaterialien und verbesserte Möglichkeiten, auf die individuellen Bedürfnisse der Lernenden einzugehen. Hervorgehoben wurde außerdem das Feststellen von Lernfortschritten und die Möglichkeit, im Unterricht differenziert auf einzelne Lernende eingehen zu können. Zudem wurde das zielgerichtete Raussuchen von Aufgaben und das Erstellen von Arbeitsblättern genannt. Wenn Kurse an Kolleg:innen übergeben werden müssen, könnte ein gemeinsamer Zugang zu den Testergebnissen im Dashboard die Übergabe erleichtern.

Rückmeldung/Feedback an Lernende: Eine Lehrkraft aus einem Alphabetisierungskurs sowie zwei Berufsschullehrer:innen beschreiben, dass das Dashboard die Möglichkeit bietet, mit Lernenden ins Gespräch zu kommen und herauszufinden, welche Lernwiderstände bestehen. Passgenaue Aufgaben für die Lernenden können außerdem die Motivation der Lernenden fördern. Anhand der differenzierten Ergebnisse des Dashboards könne den Lernenden außerdem genau erklärt werden, welche Kompetenzen sie bereits gut erreicht haben und welche nicht. Dies könnte zu einem höheren Verständnis und einer besseren Akzeptanz von schlechten Testergebnissen bei den Lernenden führen.

Einsatz über den Präsenzunterricht hinaus: In dieser Kategorie wurden ausschließlich Angaben von Lehrenden aus Alphabetisierungskursen gemacht. Sie sehen Potenziale darin, durch das Dashboard tiefergehend mit den Alpha-Levels arbeiten zu können. Dies ist aufgrund des hohen Zeitaufwandes sonst nur schwer möglich. Als weitere Einsatzmöglichkeit wird vorgeschlagen, die Ergebnisse des Dashboards als Grundlage für Zeugnisse zu nutzen oder als Ersatz bzw. Zusatz zu einem neben dem Unterricht geführten Lernheft. Zudem wurde der Wunsch geäußert, das Dashboard direkt mit digitalen Materialien zu verknüpfen, die den Lehrkräften auf Basis der Testergebnisse vorgeschlagen werden.

Zusammenfassung

Die von den Lehrenden beschriebenen Einsatzszenarien ließen sich in drei Unterkategorien unterteilen: „Unterstützung des eigenen Präsenzunterrichts und Zusammenarbeit mit Kolleg:innen“, „Rückmeldung/Feedback an Lernende“ sowie „Einsatz über den Präsenzunterricht hinaus“. Die Entwicklungsziele des Dashboards, Lernfortschritte einzelner Lernender sowie Gruppenergebnisse für die Lehrkräfte differenziert vergleichbar und nutzbar zu machen, werden durch diese Szenarien bestätigt und erweitert. Die Ansätze den eigenen Unterricht adaptiv an die Testergebnisse der Lernenden anzupassen und das Dashboard zur Unterstützung bei der Rückmeldung von Feedback an die Lernenden zu nutzen, werden von beiden Zielgruppen aufgegriffen. Die Lehrenden aus Alphabetisierungskursen sehen außerdem durch den Einsatz des Dashboards große Zeitersparnis bei der Diagnostik und der Erstellung spezifischer Unterrichtsmaterialien. Im Kontext der Berufsschule wurde beschrieben, dass Alphabetisierung und Grundbildung weder Teil der Lehramtsausbildung

sind, noch dass hierfür Zeit in den Lehrplänen eingeräumt wird. Die resultierenden Kompetenzdefizite der Berufsschullehrkräfte, kombiniert mit einer zeitlichen Knappheit im Berufsalltag, sprechen somit zunächst einmal gegen den Einsatz einer komplexen Kompetenzdiagnostik. Gleichzeitig sehen die Berufsschullehrenden aber die Notwendigkeit der Diagnostik und Förderung von gering Literalisierten und beschreiben durchaus die Motivation, das Dashboard einzusetzen. Zur Unterstützung der Berufsschullehrkräfte sollten Schulungsmaterialien und -angebote entwickelt und evaluiert werden, die auf die spezifischen Bedarfe der Lehrkräfte als fachfremde Lehrende abgestimmt sind. Zudem sollte zukünftig eine Verknüpfung von diagnostischen Ergebnissen und konkreten digitalen Lehr-/Lernmaterialien entwickelt werden. Dies würde die Lehrenden weiter entlasten.

9. Limitationen

Im Rahmen dieser Arbeit wurde eine Think-Aloud-Testung durchgeführt. Das methodische Vorgehen eignete sich, um Usability-Probleme sichtbar zu machen und Erkenntnisse zur Interpretation der Lehrenden bezüglich der im Dashboard dargestellten Daten zu liefern. Dennoch sollten einige Limitationen kritisch diskutiert werden. Die Stichprobengröße von $N=12$ kann für formative Usability-Studien als hinreichend betrachtet werden (Faulkner 2003), lässt aber keine statistische Generalisierung auf die Grundgesamtheit der Nutzenden zu. Für weiterführende Erhebungen zu Nutzungsweisen oder dem Umgang mit Datenvisualisierungen, sollte die Stichprobe – auch aufgrund der Heterogenität der Zielgruppe – vergrößert werden. Zudem unterliegt die Datenerhebung von Think-Aloud-Testungen potentiellen Verzerrungen. Zwar konnte Park u. a. (2020) nicht nachweisen, dass lautes Denken komplexere Lernleistungen beeinträchtigt, die Autor:innen zeigten jedoch, dass die Methode mit einer höheren kognitiven Belastung einhergeht und das Erinnerungsvermögen negativ beeinflusst. Teilnehmenden an Testungen, bei denen während der Erprobung gesprochen wird (Concurrent Think-Aloud (CTA)), konnten im Vergleich zu retrospektiven Think-Aloud-Tests (RTA), weniger Erklärungen, Probleme und Gestaltungsempfehlungen verbalisieren (Hertzum 2024). Es ist also nicht auszuschließen, dass insbesondere beim Umgang mit Datenvisualisierungen Handlungen oder Aussagen der Teilnehmenden durch eine hohe kognitive Belastung beeinflusst wurden. Die Testpersonen konnten den Test zwar an ihren eigenen Computern ausführen, dennoch fand dieser in einer simulierten Situation statt, was das Verhalten der Lehrkräfte beeinflussen kann (Hertzum und Holmegaard 2013). Soziale Erwünschtheit könnte außerdem dazu geführt haben, dass negative Emotionen oder Frustration in den Aussagen unterrepräsentiert sind. Zudem müssen die Äußerungen der Lehrenden zu ihren Nutzungsabsichten weiter überprüft werden, da eine Think-Aloud-Testung keine Erkenntnisse zu langfristigen Nutzungsweisen und -effekten liefert. Die qualitative Inhaltsanalyse der Protokolle wurde von einer einzelnen Person durchgeführt. Auf die Berechnung einer Inter-Rater-Reliabilität musste aus Ressourcengründen verzichtet werden, was die Objektivität der Kategorienbildung formal einschränkt. Um diesem Bias entgegenzuwirken, wurde ein achtsamer und transparenter Auswertungsprozess angestrebt (Syed und Nelson 2015). Dieser beinhaltete die Prüfung der Intra-Rater-Reliabilität sowie die Erstellung eines detaillierten Kodierleitfadens. Zudem wurden Teilergebnisse in mehreren Iterationen im Team der Forschungsgruppe diskutiert.

Dennoch können die Kodierungen in gewissem Maße von der subjektiv-interpretativen Haltung der Autorin beeinflusst sein.

10. Diskussion und Ausblick

Über die hier dargestellten Ergebnisse hinaus können weitere Erkenntnisse zur Gestaltung komplexer Dashboards für Lehrende gewonnen werden. Während acht Teilnehmende das Dashboard als „übersichtlich“, „klar unterteilt“ und „eindeutig“ bewerteten, empfanden sieben Teilnehmende die dargestellte Informationsmenge als überwältigend, wobei vier Teilnehmende beide Äußerungen tätigten. Zwar konnten sie die Informationsfülle nachvollziehen und hielten sie teilweise für unumgänglich, dennoch fühlten sie sich davon „erschlagen“. Dieses paradoxe Verhältnis unterstreicht ein grundlegendes Designdilemma bei Lehrenden-Dashboards. Einerseits sollen Dashboards umfassende, differenzierte Informationen bereitstellen, andererseits dürfen sie Nutzende nicht kognitiv überfordern. Mit Blick auf die Gestaltungsgrundlagen für Datenvisualisierungen wird gefordert, dass eine hohe Datendichte angestrebt werden sollte (Tufté 1998). Im Sinne der kognitiven Belastung (Sweller 1988) lässt sich jedoch argumentieren, dass eine hohe Informationsdichte zwar den Germane Load, also das lernförderliche Verarbeiten relevanter Informationen, erhöht, gleichzeitig aber auch den Extraneous Load steigert, wenn Zusammenhänge der Informationen erst erschlossen werden müssen (Hollender u. a. 2010). Diese divergierenden Anwendungserwartungen und Bedarfe an die Gestaltung des User Interfaces könnten durch verschiedene Nutzungsmodi adressiert werden. Eine mögliche Lösung wären Konfigurationen nach den Stufen „Einsteiger“, „Pro“ oder selbst konfigurierbare Einstellungen.

Alle Teilnehmenden konnten oberflächliche Informationen, wie beispielsweise Gesamtergebnisse, beschreiben. Dies deutet darauf hin, dass die Visualisierung keine groben Designfehler aufwies. Förderpädagogische Schlussfolgerungen und konkrete Handlungsoptionen leiteten jedoch überwiegend nur Lehrende aus Alphabetisierungskursen ab. Zudem machten Berufsschullehrende deutlich mehr fehlerhafte Aussagen in Bezug auf die Dateninterpretation. Während Pozdniakov u. a. (2024) zeigen, dass eine geringe Visualization Literacy (VL) die Interpretationstiefe von Datenvisualisierungen beeinträchtigt, scheinen im vorliegenden Fall weniger die allgemeine VL als vielmehr das domänenspezifische kompetenzdiagnostische Fachwissen ausschlaggebend für die Fehlinterpretationen zu sein. Lehrende benötigen das passende pädagogische Fachwissen, um Informationen aus Dashboards zu interpretieren und in Handlungen umzusetzen (Molenaar und Knoop-van Campen 2019). Dies deckt sich mit der Einschätzung von Van Kessel u. a. (2025), die zu dem Schluss kommen, dass Lehrende eine Kombination verschiedener Kompetenzen, insbesondere Fachwissen und Datenkompetenz, für einen pädagogisch nachhaltigen Umgang mit Dashboard-Informationen benötigen. Die sichere Bedienung der Benutzeroberfläche bedeutet nicht automatisch, dass die förderdiagnostischen Informationen verstanden wurden und in das pädagogische Handeln der Lehrenden einfließen. Insbesondere in Bezug auf die Zielgruppe der Berufsschullehrenden wird weiteres Optimierungspotential hinsichtlich der Verständlichkeit der dargestellten Informationen sowie die Notwendigkeit nach spezifischen

Unterstützungsangeboten deutlich. Um dieser Herausforderung zu begegnen, identifizieren Pozdniakov u. a. (2023) zwei Ansätze. Einerseits müsse die Datenkompetenz von Lehrenden verbessert werden. Andererseits sollten Dashboards erklärende Designmerkmale, beispielsweise Data-Storytelling-Elemente zur strukturierten, kontexteingebundenen und handlungsorientierten Darstellung von Daten enthalten, um Lehrenden den Zugang zu den Informationen zu erleichtern. Die Ergebnisse der vorliegenden Think-Aloud-Testung bestätigen insgesamt, dass die Gestaltung von Teacher Dashboards nicht allein als technisches Designproblem verstanden werden kann. Im Sinne des Learning Experience Design (Jahnke u. a. 2022) sollte neben der technischen Gestaltungsebene von User Interface und Datenvisualisierungen auch die pädagogische und die sozio-kulturelle Ebene adressiert werden. Zukünftig sollten zusätzliche Tutorials oder optionale Hilfsinformationen mit förderdiagnostischem Schwerpunkt innerhalb oder außerhalb des lea.Dashboards entwickelt werden. In anschließenden Untersuchungen müsste getestet werden, welchen Effekt z.B. In-App-Tutorials auf die Arbeit der Lehrenden hätten.

Literatur

Borgards, Lena; Onur, Karademir; Strauß, Sebastian; u. a.: 2024. Achieving Tailored Feedback by Means of a Teacher Dashboard? Insights into Teachers' Feedback Practices. In: Mello, Rafael Ferreira; Rummel, Nikol; Jivet, Ioana; Pishtari, Gerti; Ruipérez Valiente, José A. (Eds.): *Technology Enhanced Learning for Inclusive and Equitable Quality Education. Lecture Notes in Computer Science*, Bd. 15160. Springer Nature, Cham, 2024. https://doi.org/10.1007/978-3-031-72312-4_8 (last check 2026-01-27)

Breiter, Andreas; Light, Daniel: Data for School Improvement: Factors for designing effective information systems to support decision-making in schools. In: *IEEE Educational Technology & Society*, 9, 2006, pp. 206–217. https://www.j-ets.net/collection/published-issues/9_3 (last check 2026-01-27)

Doi, Toshihisa: Usability Textual Data Analysis: A Formulaic Coding Think-Aloud Protocol Method for Usability Evaluation. In: *Applied Sciences*, 11 (15), 2021, p. 7047. <https://doi.org/10.3390/app11157047> (last check 2026-01-27)

Ehmig, Simone; Heymann, Lukas; Jester, Melanie; Leck, Johanna (Eds.): *Motivation und Verbindlichkeit bei gering literalisierten Erwachsenen*. Wbv Media, Bielefeld, 2025. 10.3278/9783763976911 (last check 2026-01-27)

Ez-Zaouia, Mohamed: Teacher-Centered Dashboards Design Process. In: *2nd International Workshop on eXplainable Learning Analytics, Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge LAK20, Frankfurt, Germany, März 2020*. <https://hal.science/hal-02516815> (last check 2026-01-27)

Ez-Zaouia, Mohamed; Lavoué, Elise: EMODA: a tutor oriented multimodal and contextual emotional dashboard. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, New York, NY, USA, LAK '17, 2017*, pp. 429–38. <https://doi.org/10.1145/3027385.3027434> (last check 2026-01-27)

Faulkner, Laura: Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. In: *Behavior Research Methods, Instruments, & Computers*, 35 (3), 2003, pp. 379–383. <https://doi.org/10.3758/BF03195514> (last check 2026-01-27)

Few, Stephen: *Information Dashboard Design: The Effective Visual Communication of Data*. 1. Aufl., O'Reilly & Associates, Sebastopol, 2006.

Goodwin, Kim: *Designing for the Digital Age: How to Create Human-Centered Products and Services*. John Wiley & Sons, Indianapolis, 2011.

Grosche, Michael; Huber, Christian: Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. In: *Zeitschrift für Heilpädagogik*, 8 (12), 2012, pp. 312–222.

Grotlüschen, Anke; Buddeberg, Klaus; Solga, Heike: Leben mit geringer Literalität – ein Paradigmenwechsel. In: Grotlüschen, Anke; Buddeberg, Klaus (Eds.): *LEO 2018: Leben mit geringer Literalität*. Wbv Publikation, Bielefeld, 2020. <https://doi.org/10.3278/6004740w> (last check 2026-02-18)

Grotlüschen, Anke; Buddeberg, Klaus; Dutz, Gregor; Heilmann, Lisanne; Stammer, Christopher: Hauptergebnisse und Einordnung zur LEO-Studie 2018 – Leben mit geringer Literalität. In: Grotlüschen, Anke; Buddeberg, Klaus (Eds.): *LEO 2018: Leben mit geringer Literalität*. Wbv Publikation, Bielefeld, 2020. <https://doi.org/10.3278/6004740w> (last check 2026-02-18)

Grotlüschen, Anke; Dessing, Yvonne; Heinemann, Alisha M. B.; Schepers, Claudia: Schreiben Alpha-Levels. Waxmann, Münster, 2010. <https://www.ew.uni-hamburg.de/einrichtungen/ew3/erwachsenenbildung-und-lebenslanges-lernen/files/kompetenzmodell-schreiben.pdf> (last check 2026-01-27)

Grotlüschen, Anke; Riekman, Wibke (Hrsg.): *Funktionaler Analphabetismus in Deutschland: Ergebnisse der ersten leo. - Level-One Studie*. Alphabetisierung und Grundbildung 10. Waxmann, Münster, 2012. https://www.waxmann.com/shop/download?tx_p2waxmann_download%5Baction%5D=download&tx_p2waxmann_download%5Bbuchnr%5D=2775&tx_p2waxmann_download%5Bcontroller%5D=Zeitschrift&cHash=77115f02aa8e5d6b2b3b5a01c5ad6de4 (last check 2026-01-27)

Heine, Jörn-Henrik; Heinle, Martina; Hahnel, Carolin; Lewalter, Doris; Becker-Mrotzek, Michael: Lesekompetenz in PISA 2022 Ergebnisse, Veränderungen und Perspektiven. In: Lewalter, Doris; Diedrich, Jennifer; Goldhammer, Frank; Köller, Olaf; Kristina, Reiss (Eds.): *PISA 2022. Analyse der Bildungsergebnisse in Deutschland*. Waxmann, Münster, 2023. <https://doi.org/https://doi.org/10.25656/01:28666> (last check 2026-02-18)

Hertzum, Morten: Concurrent or Retrospective Thinking Aloud in Usability Tests: A Meta-Analytic Review. In: *ACM Trans. Comput.-Hum. Interact.* ACM, New York, NY, USA, 31 (3), 2024, pp. 1-29. <https://doi.org/10.1145/3665327> (last check 2026-01-27)

Hertzum, Morten; Due Holmegaard, Kristin: Thinking Aloud in the Presence of Interruptions and Time Constraints. In: *International Journal of Human-Computer Interaction*, 29 (5), 2013, pp. 351–364. <https://doi.org/10.1080/10447318.2012.711705> (last check 2026-01-27)

Hevner, Alan R.; March, Salvatore; u. a.: Design Science in Information Systems Research. In: *Management Information Systems Quarterly*, 28 (März), 2004, pp. 75-105.

<https://doi.org/10.2307/25148625> (last check 2026-01-27)

Hollender, Nina; Hofmann, Cristian; Deneke, Michael; Schmitz, Bernhard: Integrating Cognitive Load Theory and Concepts of Human–Computer Interaction. In: *Computers in Human Behavior*, 26 (6), 2010, pp. 1278–88. <https://doi.org/10.1016/j.chb.2010.05.031> (last check 2026-01-27)

Horn, Heike; Lux, Thomas; Ortmanns, Verena; Bachem, Andreas: *Volkshochschul-Statistik: 62. Folge, Berichtsjahr 2023. Bd. 18.* Wbv Publikation, Bielefeld, 2024.

<https://doi.org/10.3278/9783763977949> (last check 2026-01-27)

Hornbæk, Kasper: Dogmas in the Assessment of Usability Evaluation Methods. In: *Behaviour & Information Technology*, 29 (1), 2010, pp. 97–111.

<https://doi.org/10.1080/01449290801939400> (last check 2026-01-27)

International Organization for Standardization. 2018. ISO 9241-11:2018(en) Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.

<https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en> (last check 2026-01-27)

International Organization for Standardization. 2019. ISO 9241-210:2019: Ergonomics of human-system interaction Part 210: Human-centred design for interactive systems.

<https://www.iso.org/standard/77520.html> (last check 2026-01-27)

Isaias, Pedro; Backx Noronha Viana, Adriana: On the Design of a Teachers' Dashboard: Requirements and Insights. In: *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences. Lecture Notes in Computer Science 12205, LNISA, HCII 2020.* Springer International Publishing, Cham, 2020.

https://doi.org/10.1007/978-3-030-50513-4_19 (last check 2026-01-27)

Jahnke, Isa; Schmidt, Matthew; Earnshaw, Yvonne; Tawfik, Andrew: Theoretical Considerations of Learning Experience Design. In: *Theories to Influence the Future of Learning Design and Technology*, 2022.

https://edtechbooks.org/theory_comp_2021/toward_theory_of_LXD_jahnke_earnshaw_schmidt_tawfik (last check 2026-01-27)

Jivet, Ioana; Scheffel, Maren; Drachsler, Hendrik; Specht, Marcus: Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. In: Lavoué, Élise; Drachsler, Hendrik; Verbert, Katrien; Broisin, Julien; Pérez-Sanagustín, Mar (Eds.): *Data Driven Approaches in Digital Education, EC-TEL 2017, Lecture Notes in Computer Science 10474.* Springer International Publishing, Cham, 2017. https://doi.org/10.1007/978-3-319-66610-5_7 (last check 2026-01-27)

Kaliisa, Rogers; Dolonen, Jan Arild: CADA: A Teacher-Facing Learning Analytics Dashboard to Foster Teachers' Awareness of Students' Participation and Discourse Patterns in Online Discussions. In: *Technology, Knowledge and Learning*, 28 (3), 2023, pp. 937–958.

<https://doi.org/10.1007/s10758-022-09598-7> (last check 2026-01-27)

Kaliisa, Rogers; Misiejuk, Kamila; López-Pernas, Sonsoles; Khalil, Mohammad; Saqr, Mohammed: Have Learning Analytics Dashboards Lived Up to the Hype? A Systematic Review of Impact on Students' Achievement, Motivation, Participation and Attitude. In:

Proceedings of the 14th Learning Analytics and Knowledge Conference, 2024, März 18, pp. 295–304. <https://doi.org/10.1145/3636555.3636884> (last check 2026-01-27)

Karademir, Onur; Ahmad, Atezaz; Schneider, Jan; Di Mitri, Daniele; Jivet, Ioana; Drachsler, Hendrik: Designing the Learning Analytics Cockpit - A Dashboard that Enables Interventions. In: De la Prieta, Fernando; Gennari, Rosella; Temperini, Marco (Eds.): In Methodologies and Intelligent Systems for Technology Enhanced Learning, 11th International Conference, 2022, pp. 95-104. Springer International Publishing, Cham, 2022. https://link.springer.com/chapter/10.1007/978-3-030-86618-1_10 (last check 2026-01-27)

Klemm, Ulrich; Müller, Holger; Walther, Anne: Grundbildungszentren für Alphabetisierung - Zum bundesweiten Entwicklungsstand ihrer Basisarbeit. In: *forum erwachsenenbildung*, Nr. (56)4, 2023, pp. 47–49. <https://www.waxmann.com/artikelART105523> (last check 2026-01-27)

Koppel, Ilka: Entwicklung einer Online-Diagnostik für die Alphabetisierung. Springer Fachmedien, Wiesbaden, 2017. <https://doi.org/10.1007/978-3-658-15769-2> (last check 2026-01-27)

Koppel, Ilka; Wolf, Karsten: otu.lea: eine niedrighschwellige Online-Diagnostik für funktionale AnalphabetInnen in der Kursarbeit. *Alfa-Forum*, 2014, Januar, pp. 38–41.

Koppel, Ilka; Wolf, Karsten D.; Kley, Susanne; Meyer, Imke A. M.: Digitale Förderdiagnostik in der Basisbildung. Unterstützung für den Kursalltag mit lea.online. *Magazin erwachsenenbildung.at*. Nr. 47, 2022, pp. 83–95. <https://doi.org/10.25656/01:26037> (last check 2026-01-27)

Kretschmann, Rudolf; Wieken, Petra: Lesen Alpha-Levels. Waxmann, Münster, 2010a. <https://www.ew.uni-hamburg.de/einrichtungen/ew3/erwachsenenbildung-und-lebenslanges-lernen/files/kompetenzmodell-lesen.pdf> (last check 2026-01-27)

Kretschmann, Rudolf; Petra Wieken: Sprachempfinden Alpha-Levels. Waxmann, Münster, 2010b. <https://www.ew.uni-hamburg.de/einrichtungen/ew3/erwachsenenbildung-und-lebenslanges-lernen/files/kompetenzmodell-sprachempfinden.pdf> (last check 2026-01-27)

Kuckartz, Udo: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 4., Überarbeitete Aufl. Grundagentexte Methoden. Beltz, Weinheim, 2018.

Liu, Ying-Dong; Morard, Simon; Adinda, Dina; Sanchez, Eric; Trestini, Marc: A Systematic Review: criteria and dimensions of learning experience. In: *European Conference on e-Learning*, 22 (1), 2023, pp. 174–82. <https://doi.org/10.34190/ecel.22.1.1927> (last check 2026-01-27)

Liu, Yuchen; Pozdniakov, Stanislav; Martinez-Maldonado, Roberto: The effects of visualisation literacy and data storytelling dashboards on teachers' cognitive load. In: *Australasian Journal of Educational Technology*, 40(1), 2024, Online-Vorab-Publikation, März 14. <https://doi.org/10.14742/ajet.8988> (last check 2026-01-27)

Mandinach, Ellen B.; Schildkamp, Kim: Misconceptions about Data-Based Decision Making in Education: An Exploration of the Literature. In: *Studies in Educational Evaluation*, 69 (Juni), 2021, 100842. <https://doi.org/10.1016/j.stueduc.2020.100842> (last check 2026-01-27)

Meyer, Imke A. M.: Zielgruppenorientierte Anforderungsanalyse eines Dashboards zur datengestützten Entscheidungsfindung für Lehrende im Grundbildungsbereich. In:

MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, 2024, Dezember 18, pp. 512–556. <https://doi.org/10.21240/mpaed/00/2024.12.18.x> (last check 2026-01-27)

Molenaar, I.; Knoop-van Campen, C. A. N.: How Teachers Make Dashboard Information Actionable. In: *IEEE Transactions on Learning Technologies*, 12 (3), 2019, pp.: 347–355. <https://doi.org/10.1109/TLT.2018.2851585> (last check 2026-01-27)

Molich, Rolf; Nielsen, Jakob: Improving a Human-Computer Dialogue. In: *Communications of the ACM*, 33 (3), 1990, pp. 338–348. <https://doi.org/10.1145/77481.77486> (last check 2026-01-27)

Nguyen, Ngoc Buu Cat.: Evaluating a Teaching Analytics Dashboard in Adult Education: Lessons Learned. In: *Technology Enhanced Learning for Inclusive and Equitable Quality Education*, herausgegeben von Rafael Ferreira Mello, Nikol Rummel, Ioana Jivet, Gerti Pishtari, und José A. Ruipérez Valiente. *Lecture Notes in Computer Science*, Bd. 15160, Springer Nature Switzerland, 2024. https://doi.org/10.1007/978-3-031-72312-4_33 (last check 2026-02-06)

Nielsen, Jakob: *Usability Engineering*. 3. Nachdr. Morgan Kaufmann, Amsterdam, 2010.

Nielsen, Jakob: 10 Usability Heuristics for User Interface Design. Nngroup, Januar 30, 2024. <https://www.nngroup.com/articles/ten-usability-heuristics/> (last check 2026-01-28)

Nielsen, Jakob; K. Landauer, Thomas: A mathematical model of the finding of usability problems. In: *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (New York, NY, USA), CHI '93, 1993, pp. 206–213. <https://doi.org/10.1145/169059.169166> (last check 2026-01-28)

Nokelainen, Petri: An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. In: *Educational Technology & Society*, 9 (April), 2006, pp. 178–197. <https://www.jstor.org/stable/jeductechsoci.9.2.178> (last check 2026-01-28)

Park, Babette; Korbach, Andreas; Brünken, Roland: Does thinking-aloud affect learning, visual information processing and cognitive load when learning with seductive details as expected from self-regulation perspective? In: *Computers in Human Behavior*, 111 (Oktober), 2020, 106411. <https://doi.org/10.1016/j.chb.2020.106411> (last check 2026-01-28)

Pozdniakov, Stanislav; Martinez-Maldonado, Roberto; Tsai, Yi-Shan, u. a.: The Question-Driven Dashboard: How Can We Design Analytics Interfaces Aligned to Teachers' Inquiry? In: *LAK22: 12th International Learning Analytics and Knowledge Conference*, März 21, 2022, pp. 175–185. <https://doi.org/10.1145/3506860.3506885> (last check 2026-01-28)

Pozdniakov, Stanislav; Martinez-Maldonado, Roberto; Tsai, Yi-Shan; Echeverria, Vanessa; Srivastava, Namrata; Gasevic, Dragan: How Do Teachers Use Dashboards Enhanced with Data Storytelling Elements According to their Data Visualisation Literacy Skills? In: *LAK23: 13th International Learning Analytics and Knowledge Conference*, New York, NY, USA, LAK2023, 2023, pp. 89–99. <https://doi.org/10.1145/3576050.3576063> (last check 2026-01-28)

Pozdniakov, Stanislav; Martinez-Maldonado, Roberto; Tsai, Yi-Shan; Echeverria, Vanessa; Swiecki, Zachari; Gašević, Dragan: Investigating the Effect of Visualization Literacy and

Guidance on Teachers' Dashboard Interpretation. In: *Journal of Learning Analytics*, 12 (1), 2024, pp. 367–390. <https://doi.org/10.18608/jla.2024.8471> (last check 2026-01-28)

Rubin, Jeffrey; Chisnell, Dana: *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. 2nd ed. With. Wiley, Indianapolis, 2008.

Safsouf, Yassine; Mansouri, Khalifa; Poirier, Franck: Experimental Design of Learning Analysis Dashboards for Teachers and Learners. In: *Proceedings of the Eighth ACM Conference on Learning @ Scale*, 2021, Juni 8, pp. 347–350.

<https://doi.org/10.1145/3430895.3460990> (last check 2026-01-28)

Schepers, Claudia; Kley, Susanne; Kosmalla, Lena, u. a.: *Handreichung zum Arbeiten mit den Anwendungen otu.lea, lea.Dashboard und lea.App*. Universität Bremen, 2022.

https://blogs.uni-bremen.de/leasonline/files/Handreichung_Orange_LeaOnline.pdf (last check 2026-01-28)

Schildkamp, Kim: Data-Based Decision-Making for School Improvement: Research Insights and Gaps. In: *Educational Research*, 61 (3), 2019, pp. 257–273.

<https://doi.org/10.1080/00131881.2019.1625716> (last check 2026-01-28)

Schildkamp, Kim; Handelzalts, Adam; Poortman, Cindy L.; u. a.: *The Data TeamTM Procedure: A Systematic Approach to School Improvement*. Springer Texts in Education. Springer International Publishing, Cham, 2018.

<https://doi.org/10.1007/978-3-319-58853-7> (last check 2026-01-28)

Schmidt, Matthew; Earnshaw, Yvonne; Tawfik, Andrew; Jahnke, Isa: *Methods of User Centered Design and Evaluation for Learning Designers*. In: *Learner and User Experience Research*, 1. Aufl. EdTech Books, 2020.

<https://doi.org/10.59668/36> (last check 2026-01-28)

Schmidt, Matthew; Tawfik, Andrew; Jahnke, Isa; Earnshaw, Yvonne: Introduction to the Edited Volume. In: *Learner and User Experience Research*, 1. Aufl. EdTech Books, 2020.

<https://doi.org/10.59668/36> (last check 2026-01-28)

Schneider, Rebecca; Boemmel, Quirin: Kompetenzstufenbesetzung im Fach Deutsch. In: Stanat, Petra; Schipolowski, Stefan; Schneider, Rebecca; Weirich, Sebastian; Henschel, Sofie; Sachse, Karoline A. (Eds.): *IQB-Bildungstrend 2022. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich*. Waxmann, Münster, 2023.

<https://doi.org/10.31244/9783830997771> (last check 2026-02-18)

Souvignier, Elmar; Förster, Natalie: Data-based decision making. In: Quenzel, Gudrun; Hurrelmann, Klaus; Groß Ophoff, Jana; Weber, Christoph: *Handbuch Bildungsarmut*, herausgegeben. Springer Fachmedien, Wiesbaden, 2025.

https://doi.org/10.1007/978-3-658-44698-7_39-1 (last check 2026-01-28)

Sweller, John: Cognitive Load During Problem Solving: Effects on Learning. In: *Cognitive Science*, 12 (2), 1988, pp. 257–285.

https://doi.org/10.1207/s15516709cog1202_4 (last check 2026-01-28)

Syed, Moin; Nelson, Sarah C.: Guidelines for Establishing Reliability When Coding Narrative Data. In: *Emerging Adulthood*, 3 (6), 2015, pp. 375–387.

<https://doi.org/10.1177/2167696815587648> (last check 2026-01-28)

Tufte, Edward R.: *The Visual Display of Quantitative Information*. 16. Aufl. Graphics Press, Cheshire, 1998.

Van Kessel, Manel; Molenaar, Inge; A. N. Knoop-van Campen, Carolien; De Jonge, Mario; Saab, Nadira: Primary School Teacher Perspectives on Effective Dashboard Use in the Classroom: Skills, Knowledge, and Contextual Conditions. In: *Journal of Learning Analytics*, 12 (2), 2025, pp. 279–292. <https://doi.org/10.18608/jla.2025.8493> (last check 2026-01-28)

Van Leeuwen, Anouschka; Strauß, Sebastian; Rummel, Nikol: Participatory design of teacher dashboards: navigating the tension between teacher input and theories on teacher professional vision. In: *Frontiers in Artificial Intelligence*, 6 (Mai), 2023, 1039739. <https://doi.org/10.3389/frai.2023.1039739> (last check 2026-01-28)

Van Leeuwen, Anouschka; Teasley, Stephanie D.; Friend Wise, Alyssa: Teacher and Student Facing Learning Analytics. In: Lang, Charles; Siemens, George; Wise, Alyssa; et al. (Eds.): *The Handbook of Learning Analytics*. 2. Aufl. SOLAR, 2022. <https://doi.org/10.18608/hla22.013> (not available 2026-01-28)

Vermeeren, Arnold P. O. S.; Lai-Chong Law, Effie; Roto, Virpi; Obrist, Marianna; Hoonhout, Jettie; Väänänen-Vainio-Mattila, Kaisa: User Experience Evaluation Methods: Current State and Development Needs. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, 2010, Oktober 16, pp. 521–530. <https://doi.org/10.1145/1868914.1868973> (last check 2026-01-28)

Wolf, Karsten D.; Koppel, Ilka; Schwedes, Kai: Potenziale von Rich E-Assessment für die Förderdiagnostik funktionaler Analphabeten. In: Grotlüschen, Anke; Kretschmann, Rudolf; Quante-Brandt, Eva; Wolf, Karsten D. (Eds.): *Literalitätsentwicklung von Arbeitskräften*. Waxmann, Münster, 2011.