## Learner und User Experience messen

von Tobias Hauser

# **Befragungen • Crowdtests • Learning Experience • Nutzertest • Usability • Webanalyse**

**Schlagworte** 

Überblick

Fragt man eine Gruppe E-Learning-Verantwortlicher, welche Faktoren für erfolgreiches Lernen zentral sind, so landet die einfache Benutzbarkeit der Lernplattform und auch der Inhalte in den meisten Fällen auf Platz 1.

Ein Unterschied liegt häufig beim gewählten Begriff: Einige sprechen von Usability, andere von Learner oder Learning Experience. Um diese verschiedenen Begriffe kümmern wir uns in Abschnitt 1. Das grundlegende Ziel bleibt allerdings unabhängig von der Begrifflichkeit immer gleich: Der Lerner steht im Mittelpunkt und sollte es möglichst einfach haben, die Lernplattform und den Lerninhalt zu nutzen. Diese Einfachheit ist übrigens nicht zu verwechseln mit einfachem oder gar leichtem Lerninhalt. Der Fokus liegt darauf, alle Störungen rund um das Lernen selbst zu vermeiden und dem Lerner damit die ideale Learner Experience (abgekürzt LX) zu verschaffen.

Sind sich die meisten E-Learning-Verantwortlichen beim Ziel noch weitgehend einig, so gibt es doch in vielen Organisationen und Unternehmen, die E-Learning betreiben, noch einen weißen Fleck auf der Landkarte: Gute Learner Experience wird zwar angestrebt, aber nicht gemessen. Diesen weißen Fleck füllt dieser Beitrag mit Struktur und gibt einen Überblick über praktisch anwendbare Messmethoden von der Webanalyse über den Test mit Nutzern bis hin zu Crowdtests. Der Fokus liegt dabei auf der praktischen Anwendung, nicht auf den wissenschaftlichen Hintergründen. Aus diesem Grund beleuchtet er auch Methoden wie/b- und multivariaten Tests, die aktuell deutlich häufiger im Bereich E-Commerce zum Einsatz kommen und erst seit kurzem im E-Learning Einzug halten.

**Tobias Hauser** ist Autor, Trainer und Berater mit Schwerpunkt E-Learning-Technologie. Als Autor schreibt er seit über fünfzehn Jahren zu wichtigen Webthemen von Technologie bis Usability. Seit Anfang der 2000er spricht er auf Konferenzen und berät Unternehmen und Organisationen im E-Learning-Umfeld.

Autor

Kontakt: Tobias Hauser, Arrabiata Solutions GmbH, Seidlstraße 25, 80335 München

E-Mail: tobias.hauser@arrabiata.de

Inhalt	1	Was ist eigentlich Learner Expierence?	2
	1.1	Usability vs. User & Learner Experience	3
	1.2	Fokus bei der Optimierung	4
	1.3	Begriffswirrwarr: Learning Experience und LXPs	4
	2	Messmethoden	5
	2.1	Webanalyse	6
	2.1.1	Tools	6
	2.1.2	Nutzerverhalten	7
	2.1.3	Informationen zu Browsern und Endgeräten	8
	2.1.4	Herkunft und geografische Daten	8
	2.1.5	Individuelle Ereignisse und Ziele messen	9
	2.1.6	Suche	10
	2.1.7	Technische Herausforderungen und Alternativen	11
	2.2	Reports aus dem System	12
	2.3	Mit Nutzern testen	13
	2.3.1	Testschwerpunkt	14
	2.3.2	Probanden	15
	2.3.3	Testsituation	16
	2.3.4	Lautes Denken	18
	2.3.5	Eye Tracking	18
	2.3.6	Stressmessung	22
	2.3.7	Emotionserkennung	25
	2.4	Crowdtests	26
	2.5	a/b- und multivariate Tests	26
	3	Optimierungs- und Messbedarf erkennen	29
	3.1	Feedback & Support	29
	3.2	Befragungen	29
	4	Fazit	31

## 1 Was ist eigentlich Learner Expierence?

#### **Definition**

Für Learner Experience gibt es zwar keine offizielle Definition, wir versuchen uns hier dennoch an einem ersten Ansatz:

Learner Experience bezeichnet die gesamte Lernerfahrung eines Lerners auf seiner Learner Journey.

Sehen wir uns die Bestandteile dieser einfachen Definition genauer an:

- *Learner* adressiert den Lerner an sich. Das heißt, der Lerner steht im Fokus dieser Betrachtungsweise.
- Experience bezeichnet die Erfahrung bei der Nutzung einer Lernplattform und des Lerninhalts. Manchmal wird Experience auch mit Erlebnis über-

setzt. Dieser Begriff ist im Lernumfeld etwas missverständlich, da hier Erlebnis oft mit Gamification gleichgesetzt wird. Die Experience fokussiert sich auf Wahrnehmung und Reaktion des Nutzers.

 Die Learner Journey bezeichnet den Weg, den ein Nutzer beim Lernvorgang durch die Lernanwendung nimmt. Dieser Weg kann dabei auch aus mehreren Journeys bestehen und umfasst wirklich alle Vorgänge: Inkludiert sind zum Beispiel auch die Registrierung und Anmeldung auf einer Plattform vor dem Lernen oder der Abruf eines Teilnahmezertifikats oder Badges nach erfolgreichem Kursabschluss.

#### 1.1 Usability vs. User & Learner Experience

Um Learner Experience noch besser einordnen und abgrenzen zu können, lohnt ein genauer Blick auf die älteren – besser definierten – Begriffe Usability und User Experience sowie deren Unterschiede.

*Usability* leitet sich vom Englischen "usable" ab und bedeutet "nutzbar". Der Fokus liegt hier also auf Nutzerfreundlichkeit bzw. Gebrauchstauglichkeit. Das findet sich auch in der Definition in der DIN ISO-Norm 9241-11 wieder:

Usability ist das Ausmaß, in dem ein System durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen. (Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018); Deutsche Fassung EN ISO 9241-11:2018, Ausgabe 2018-11)

Damit fokussiert sich Usability auf die Nutzung eines Systems. Im Gegensatz dazu ist der Begriff *User Experience* (abgekürzt UX) weiter gefasst, und umfasst die gesamte Wahrnehmung und Reaktion des Nutzers vor, während und nach der Nutzung. In der DIN ISO-Norm 9241-210 findet sich hierzu eine interessante Definition:

Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produkts, eines Systems oder einer Dienstleistung resultieren. [...] Dies umfasst alle Emotionen, Vorstellungen, Vorlieben, Wahrnehmungen, physiologischen und psychologischen Reaktionen, Verhaltensweisen und Leistungen, die sich vor, während und nach der Nutzung ergeben. (Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019); Deutsche Fassung EN ISO 9241-210:2019).

Damit wird auch die Beziehung zwischen Usability und User Experience klar: Usability ist ein Teil von User Experience, sogar der zentrale Teil, denn ohne gute Usability während der Nutzung ist keine gute User Experience möglich.

Bleibt die Beziehung zwischen *Learner Expierence* und ihren "User"-Schwestern. Learner Experience gleicht in der Ausprägung der User Experience und bezieht sich damit auf die Wahrnehmungen und Reaktionen der Lerner in allen Phasen der Nutzung eines Lernsystems. Dementsprechend ist die Usability natürlich auch für die Learner Experience die wichtigste Komponente. Gerade für die Messung steht die Usability-Phase während der Nutzung des Lernsystems im Vordergrund.

Ein letzter Begriff aus dem Experience-Universum sei hier auch noch erwähnt: die *Customer Experience* (kurz CX). Sie stammt aus dem E-Commerce-Umfeld und wird deutlich weiter gefasst als die User Experience. Sie umfasst nicht nur einzelne Nutzungsvorgänge, sondern betrachtet alle Interaktionen zwischen Kunde und Marke/Unternehmen an allen Touchpoints. Unter Touchpoints werden alle Berührungspunkte zwischen Kunde und Marke verstanden. Dies umfasst sowohl Online-Anwendungen wie den eigentlichen Online-Shop als auch Offline-Kontaktpunkte wie im Laden in der Fußgängerzone oder eine Plakatwerbung.

**Definition Usability** 

Definition User Experience

Zusammenspiel Usability und User Experience

Zusammenspiel Learner Experience und User Experience

Definition Customer Experience Übertragen auf die Learning Experience würde diese breite Betrachtungsweise alle Lerninteraktionen inklusive Offline-Lernen umfassen. Für die Messung liegt der Fokus allerdings üblicherweise auf der enger gefassten Definition.

#### 1.2 Fokus bei der Optimierung

Im Projektkontext bei einem Optimierungsprojekt einer Lernanwendung lassen sich zwei Ansätze unterscheiden: Der usability-orientierte Ansatz fokussiert sich auf die effiziente Nutzung, der experience-orientierte Ansatz bezieht auch die emotionale Erfahrung und die psychologischen Bedürfnisse mit ein. Die folgende Tabelle gibt einen Überblick.

	Usability-orientierter Ansatz	Experience-orientierter Ansatz
Fokus	Effiziente und unterbre- chungs-/abbruchfreie Nutzung	Emotionales Erfahren und die Erfüllung psychologischer Bedürfnisse
Ziel	Probleme und Stress beim Nutzer vermeiden	Positive (Lern-)Erfahrungen verstehen, erzeugen und erweitern
Im Projekt	Usability testen, Verbesse- rungen umsetzen	Usability ist die Basis – der Denkansatz "Erfahrung" kann aber zu anderen Lösungen führen

Tab. 1: Usability-orientierter und experience-orientierter Ansatz

Bei der Messung überschneiden sich beide Ansätze stark. Bestimmte Messmethoden in Nutzertests haben ihren Fokus zwar auf Usability, können aber mit Stress- und Emotionserkennung ausgeweitet werden und die Nutzererfahrung ebenfalls erfassen.

#### 1.3 Begriffswirrwarr: Learning Experience und LXPs

#### Das Erlebnis steht im Fokus

Ähnlich wie die genaue Definition von Learner Experience so ist auch die Schreibweise nicht normiert. So finden sich sowohl Learner als auch *Learning Experience*. Tatsächlich liefert eine Google-Suche sogar deutlich mehr Treffer für Learning Experience. Das liegt allerdings hauptsächlich darin begründet, dass seit einigen Jahren so genannten LXP-System den Lernsystem-Markt durcheinanderwirbeln. LXP steht dabei für Learning Experience Plattform. Der Name soll andeuten, dass diese Systeme vor allem mit Fokus auf der Learner Experience konzipiert wurden.

Funktional handelt es sich dabei allerdings in vielen Fällen um enge Verwandte der LMS (Lernmanagementsysteme). Manche LXP unterscheiden sich von klassischen LMS durch ihren Fokus auf soziales und kollaboratives Lernen. Sie verwenden keine klassische Kursstruktur und oft können auch die Lerner selbst eigene Lerninhalte erzeugen. Allerdings ist der Begriff LXP im Marketing- und Vertriebsgeschnatter der Software-Anbieter zu willkürlich im Einsatz, um ihn wirklich genau zu definieren.

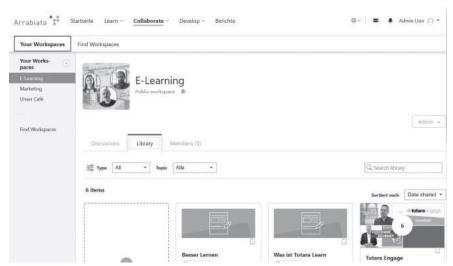


Abb. 1: Ein LXP-System setzt oft den Fokus auf soziale Interaktion und Wissensdatenbank (Totara Engage)

Dieser Beitrag zieht den Begriff Learner Experience aus zwei Gründen dem etwas gebräuchlicheren Learning Experience vor:

- 1) Weil Learner Experience eine klarere Unterscheidung von den LXP-Systemen erlaubt. In unserem Kontext bedeutet Learner Experience nämlich die Messung jeder Lernerfahrung, egal in welcher Art von Lernsystem.
- 2) Weil die verwandten Begriffe User und Customer Experience ebenfalls den Nutzer für die Bezeichnung verwenden.

#### 2 Messmethoden

Leider gibt es nicht die eine optimale Messmethode für die Learner Experience, sondern einen Strauß an Möglichkeiten. Wir unterscheiden grob folgende Methoden:

- Anonyme Auswertungen zum Verhalten der Lerner erhalten Sie ohne Zusatzaufwand aus der klassischen Webanalyse oder auch direkt aus den Lernsystemen. Die Webanalyse liefert dabei eher allgemeine Informationen zu den Lernern und deren Verhalten, während die Reports der Lernsysteme vor allem den Lernerfolg bzw. -misserfolg verzeichnen. Die große Stärke der anonymen Auswertungen ist, dass Sie sie quasi jederzeit einsetzen und ohne zusätzlichen Aufwand "mitlaufen" lassen können.
- Tests mit Lernern, also echten Nutzern, liefern qualitativ die besten Ergebnisse: Sie erhalten direkt Feedback Ihrer Zielgruppe(n) und mit den entsprechenden Testmethoden erhalten Sie einen direkten Einblick in die Lernerfahrung, indem Sie dem Lerner beim Eyetracking durch die Augen schauen und bei der Stressmessung den Puls fühlen. Wo immer ein direkter Kontakt mit dem Nutzer nicht möglich ist, können Tests mit Lernern auch Remote oder als Crowdtests durchgeführt werden.
- Die letzte Gruppe an Messmethoden dient vor allem dazu, die Auswirkung von Änderungen zu bewerten. In diesen sogenannten Splittests werden verschiedenen anonymen Nutzern im Livebetrieb zwei Varianten (a/b-Test) oder auch eine Kombination mehrerer Varianten (multivariate Tests) einer Webseite oder eines Inhalts gezeigt.

#### 2.1 Webanalyse

#### Auf Websites und im E-Commerce ein absoluter Klassiker

Die Webanalyse ist auf Websites und im E-Commerce ein absoluter Klassiker. Mit zentralen Vertretern wie Google Analytics oder Matomo gehören die Reports zur täglichen Pflichtlektüre der Verantwortlichen. Im E-Learning-Bereich ist das interessanterweise noch nicht flächendeckend so. Viele LMS-oder LXP-Administratoren beschränken sich rein auf die Reports aus den Systemen. Dabei ist die Implementierung von einem Webanalyse-Tool in einem solchen System mit sehr wenig technischem Aufwand verbunden.

Und sogar in Web Based Trainings (WBT) ist die Integration grundsätzlich denkbar. Dort ist – je nach Autorensystem – die Integration allerdings mit einigen Herausforderungen verbunden. Für die meisten grundlegenden Daten zu Nutzern reicht allerdings auch schon die Integration im LMS.

#### **2.1.1** Tools

#### Einteilung in vier Kategorien

Für die Webanalyse kommen sehr viele verschiedene Tools in Betracht. Das bekannteste, aber auch umstrittenste, ist Google Analytics. Die bekannteste, datenschutz-freundliche Alternative ist Matomo, das man auf dem eigenen Server betreiben kann. Generell lassen sich die Tools in folgende Kategorien unterteilen:

- **Open Source-basierte Messtools**, die man auf dem eigenen Server betreibt. Der bekannteste Vertreter dieser Gattung ist Matomo (https://matomo.org/), eine Alternative bildet Open Web Analytics (http://www.openwebanalytics.com/). Matomo selbst bietet auch eine kommerzielle Cloud-Version und auch kommerzielle Plugins. Damit ähnelt das Angebot den kommerziellen Tools.
- Die **kommerziellen Tools** bieten sehr viele Funktionen und hosten den Analytics-Server selbst. Hier gibt es deutsche Anbieter mit dem Vorteil deutscher Server, z. B. eTracker (https://www.etracker.com/) oder Econda (https://www.econda.de/). Ansonsten unterscheiden sich die Anbieter im angebotenen Funktionsumfang und auch im Preisbereich. Die Spanne reicht hier von einer dreistelligen Summe bis zu einer fünfstelligen Summe im Jahr, abhängig von der Menge an gebuchten Leistungen und der Menge an Seitenabrufen.
- Google Analytics spielt eine Art Sonderrolle, da es von den meisten Unternehmen und Organisationen kostenfrei genutzt wird und sich Google selbst auch komplett um die Server kümmert. Wer mit der Datenliebe von Google leben kann, stößt hier auf ein unschlagbares Preis-/Leistungsverhältnis. Funktional bieten andere Anbieter allerdings teilweise mehr. Nicht verschwiegen werden soll, dass Google mit Analytics 360 auch eine Premium-Version anbietet. Diese ist allerdings mit einem nur auf Anfrage kommunizierten Preis von ca. 150.000 Dollar pro Jahr sehr teuer.
- Spezielle kommerzielle Analysetools für die Usability-Optimierung wie Hotjar (https://www.hotjar.com/) oder CracyEgg (https://www.crazyegg.com/). Diese Tools sind nicht für klassische Webanalyse-Daten wie Seitenaufrufe gedacht, sondern erfassen zum Beispiel, wohin Nutzer klicken und wie sie durch Seiten scrollen.

In den folgenden Abschnitten erläutern wir die Möglichkeiten der Webanalyse tool-unabhängig. Als Beispiele kommen Google Analytics und Matomo zum Einsatz, alle gezeigten Erkenntnisse lassen sich aber auch mit beliebigen anderen Webanalyse-Tools erzielen.

6 92. Erg.-Lfg. August 2021 Handbuch E-Learning

#### 2.1.2 Nutzerverhalten

Eine Kernaufgabe der Webanalyse ist die grundlegende Messung des Nutzerverhaltens. Aus der Perspektive der Learner Experience handelt es sich hierbei um anonyme und aggregierte Daten. Das heißt, der Nutzer selbst ist nicht bekannt, die Werte sind über einen frei wählbaren Zeitraum der Mittelwert aus allen Lernern, die in dem Zeitraum auf der Plattform aktiv waren.

Keine Nutzerbezogenheit

Hier ein Überblick der wichtigsten Kennzahlen:

Kennzahlen

- Eindeutige Besucher/Unique Visitors sind die Besucher, die das Lernangebot aufrufen. Eindeutig heißt dabei, dass ein Besucher bei mehreren Besuchen nur einmal gezählt wird.
- Besuche pro Besucher zeigt, ob und wie oft einzelne Besucher das Lernangebot besuchen.
- Seitenaufrufe/Page Impressions umfasst die Menge der Aufrufe einzelner Webseiten. Erfasst wird dies sowohl als Gesamtzahl als auch pro einzelner Webseite. Daraus lässt sich z. B. auch ableiten, wie viele Seiten pro Besuch aufgerufen werden.
- Verweildauer umfasst die Zeit, die ein Nutzer auf einer Seite vom ersten bis zum letzten Seitenaufruf verbringt. Je nach Einstellung des Analysetools kann auch das Wahrnehmen von bestimmten Teilen einer länger scrollenden Seite gemessen werden.
- Absprungrate/Bounce Rate misst normalerweise, wieviele Besucher nach dem Besuch von nur einer Seite direkt wieder abspringen. Eine hohe Rate deutet dementsprechend auf viele Abbrüche hin.

Diese Kennwerte werden zumeist auf das gesamte Lernangebot bezogen. Allerdings lassen sich – je nach Tool – diese Werte auch nur für einzelne Seiten oder Seitenbereiche *filtern*. Außerdem können Sie auch die Nutzer selbst filtern und beispielsweise die Werte von Smartphone-Nutzern und Desktop-Nutzern miteinander vergleichen. Die Möglichkeiten sind hier nahezu unendlich.





Abb. 2: Nutzungsdaten wie Verweildauer und Absprungrate (Google Analytics)

Wichtig ist, bei der Webanalyse zu wissen, dass es sich um *keine exakte Wissenschaft* handelt. Die gemessenen Werte können sich je nach verwendetem Tool deutlich unterscheiden. Beispielsweise zählen manche Tools einen Besucher als "wiederkehrend", wenn er innerhalb von 28 Tagen erneut vorbeischaut; andere warten dagegen 30 Tage. Und auch die Art der Implementierung und der Konfiguration kann unterschiedliche Messergebnisse erzeugen.

Unterschiedliche Messmaßstäbe

#### Problematische Bewertungen

Ebenfalls problematisch sind *Pauschalaussagen*, was denn eigentlich gute Werte sind. Ist beispielsweise eine durchschnittliche Verweildauer in einem LMS von 10 Minuten gut? Das hängt tatsächlich vom Kontext ab. Wenn gleichzeitig die Abbruchrate unter 20 % liegt und die Lerneinheiten kurz sind, kann das sehr gut sein. Wenn sich andererseits der Lerner aber während der 10 Minuten hauptsächlich in der Suche bewegt und damit keine Zeit zum Lernen hat, wäre die hohe Verweilbauer nicht positiv zu werten.

# Kontext und Veränderungen beachten

Der Schlüssel zum Erfolg ist dementsprechend, die Werte immer in den *Kontext* zu setzen und *Veränderungen* zu beobachten. Die Webanalyse ist hier also ein erstes Indiz, das uns mit dem Wissen der von uns durchgeführten Veränderungen bei Inhalt und Plattform auf potentielle Probleme oder natürlich auch positive Entwicklungen des Nutzungsverhaltens der Lerner hinweist.

#### 2.1.3 Informationen zu Browsern und Endgeräten

#### Smartphone oder Tablet

Ein wichtiger Punkt sind die Informationen zu den genutzten Endgeräten. Die meisten Tools liefern hier sogar genaue Informationen wie Smartphone- und Tablet-Marken. Auch wenn diese nicht absolut zuverlässig sind, so sind zumindest die eindeutig feststellbaren Verhältniszahlen für die Verteilung zwischen Smartphone bzw. Tablet und Desktop besonders wertvoll. Wenn beispielsweise auf einem noch nicht für die mobile Darstellung optimierten Angebot über 20 % der Lerner mit Smartphones unterwegs sind, dann weiß man, dass die Optimierung für die ideale Learner Experience erforderlich ist.

#### **Browser-Version**

Neben den Endgeräten selbst, verraten die Daten auch die eingesetzten Browser-Versionen und die Auflösungen der Bildschirme. Diese Informationen liefern Ihnen klare Angaben, in welchen Plattformen und unter welchen Bildschirmauflösungen Sie Ihre Lerninhalte unbedingt testen sollten. Die Auswirkung auf die Learner Experience ist klar: Indem Sie die Geräte Ihrer Nutzer besser kennenlernen, können Sie Ihr Angebot auf deren echte Erfahrung besser zuschneiden, indem Sie ihre Tests und Qualitätssicherungsmaßnahmen mit den richtigen Geräten und Browsern durchführen.

#### 2.1.4 Herkunft und geografische Daten

## Bedeutung für internationale Lernplattformen

Gerade bei internationalen Lernplattformen ist es besonders interessant, aus welchen Regionen die Lernenden kommen. Hier liefern die Webanalysetools Länder, Regionen und u. U. sogar Städte. Auch hier sollte man etwas skeptisch sein. Gerade Klassiker wie das Dreiländereck Deutschland/Schweiz/Frankreich sorgen für teilweise mäßige Ergebnisse. Und auch VPN-Nutzer in Unternehmen verraten ihren echten Standort oft nicht.

#### **Browser-Sprachen**

Dennoch lassen sich die Daten aus Learner-Experience-Sicht gut nutzen, um besser zu verstehen, woher die Lerner kommen und wo sie lernen. Hierzu gehört auch die Sprachen der verwendeten Browser, die beispielsweise Hinweise geben, welche Oberflächensprachen gefragt sind.

8



Abb. 3: Auswertung der geografischen Aufrufe (Matomo)

Neben dem Wissen über den Nutzer, gewinnen wir aus den geografischen Daten allerdings auch direkte Hinweise auf die Optimierung der Learner Experience. Das Stichwort ist *hier Performance*. Wenn wir, wie in der obigen Abbildung 4 zu sehen, eine Lernsituation haben mit vielen Lernern in Asien und einigen in afrikanischen Ländern, sollten wir technisch sicherstellen, dass der Zugriff auf die Plattform und die Inhalte in diesen Ländern auch performant und reibungsfrei abläuft. Dabei ist es egal, ob die Lernplattform in der Cloud läuft oder zentral betrieben wird. Eine performante internationale Lösung ist in beiden Szenarien nicht automatisch garantiert.

Bei international ausgerichteten Lernplattformen sollten Sie anschließend an die erste Auswertung in der Webanalyse die internationale Performance noch mit entsprechenden Tools messen. Hier bieten beispielsweise Spezialisten für die internationale Lastverteilung wie Akamei (https://www.akamai.com/de/de/products/performance/cloudtest.jsp) entsprechende Lösungen, die auch das Simulieren von internationalen Zugriffen erlauben.

#### 2.1.5 Individuelle Ereignisse und Ziele messen

Die bisher betrachteten Messwerte gehörten alle zum Standardumfang der Tools. Um sie zu erhalten, war keine Konfiguration notwendig. Allerdings bieten die meisten Webanalysetools auch die Möglichkeit, eigene Messkriterien festzulegen, um besondere Lernsituationen zu testen.

Die meisten Webanalyse-Tools kennen zwei Arten solcher individuellen Messungen:

- 1) Ereignisse für die Messung von Nutzerinteraktionen und
- 2) Ziele.

Die folgende Abbildung zeigt beispielsweise ein sogenanntes **Ereignis** in Google Analytics. Dieses Ereignis musste dazu vorher in der Lernanwendung einmal technisch konfiguriert werden. In diesem Fall handelt es sich um einen Klick auf einen Teaser für ein Lernangebot. Sobald es dann aber einmal definiert ist, wird es kontinuierlich ausgewertet und gemessen.

Performance beachten

Zusätzliche Auswertung

Zwei Messarten

Ereignisse



Abb. 4: Individuell gemessene Ereignisse (Google Analytics)

**Ziele** Ein Ziel kann z. B. die Erreichung der Seite mit dem Teilnehmerzertifikat sein. Die möglichen Einsatzszenarien sind quasi unbegrenzt, hier ein paar Anregungen aus unseren Praxisprojekten:

- Bei Lernvideos können Sie messen, wie viele Nutzer z. B. mindestens 50 % des Videos gesehen haben.
- Bei statischen Lerninhalten messen Sie die Zahl der Downloads.
- Auf Dashboards können Sie messen wie häufig ein Teaser mit den empfohlenen Lerninhalten angeklickt wird.
- An der Grenze von E-Commerce und E-Learning könnten Sie aufeinanderfolgende Ziele definieren, zuerst für den Verkauf eines Lerninhalts, dann für die Durchführung.

#### 2.1.6 Suche

#### **Interne Suche**

Eine der wichtigsten Schnittstellen zwischen Nutzer und Lernplattform und damit zentral für die Learner Experience ist die Lernplattform-interne Suche (bei Google Analytics z. B. Site Search genannt). Hier liefert uns die Webanalyse in den gängigen Tools sowohl allgemeine Informationen zur Suchnutzung als auch eine Auflistung der zentralen Suchbegriffe. Allerdings ist hierfür eine entsprechende technische Konfiguration notwendig, damit das Webanalyse-Tool die interne Suche und den eingesetzten Suchbegriff erkennt.

#### Allgemeine Informationen

10

Die *allgemeinen Informationen* aus der Suche verraten uns einige für die Learner Experience interessante Informationen:

- Der Prozentsatz der Lerner, die die Suche nutzen, sagt uns, ob die Suche gefunden und genutzt wird. In Verbindung steht natürlich, bei wie vielen Besuchen/Sitzungen die Suche verwendet wurde.
- Die Menge an Suchanfragen pro Lerner verrät uns, ob die Lerner bereits in der ersten Suche schon passende Inhalte gefunden haben oder mehrfach suchen mussten.
- Die Ausstiege nach der Suche verraten uns, wieviele Lerner die Plattform verlassen, nachdem Sie nichts gefunden haben. Ergänzt wird dies je nach Tool unter anderem von der Verweilzeit, die Lerner noch nach dem Suchvorgang auf der Plattform verbringen.

Handbuch E-Learning

92. Erg.-Lfg. August 2021



Abb. 5: Kennzahlen für die Suche (Google Analytics)

Aus den *Suchbegriffen* können wir für die Learner Experience ebenfalls viel lernen, denn hier erfahren wir wie und wonach Nutzer wirklich suchen. Auf den ersten Blick handelt es sich dabei meist um lange Listen, die etwas unübersichtlich wirken. Um daraus sinnvolle Ergebnisse abzuleiten, ist es empfehlenswert, aus dem Wust Anwendungsfälle abzuleiten, deren Häufigkeit und den Sucherfolg für den Lerner zu bewerten. Hier einige Beispiel von häufigen Anwendungsfällen:

- Suchen nach allgemeinen Themen (oft Einwortsuchen)
- Suchen nach konkreten Wissensfragen (oft Mehrwortsuchen, manchmal Fragen)
- Suchen nach bestimmten Lernleveln (Grundlagen, Fortgeschritten)
- Suche nach Kursnummern
- Suche nach Trainern/Referenten
- Suche nach Lernformaten
- Fehlerhafte Suchen und falsche Schreibweisen

Wenn Sie die Suchbegriffe für einen Beispielzeitraum nach Ihren Anwendungsfällen kategorisieren, in der Häufigkeit einschätzen und grob qualifizieren, ob der Nutzer für die Anwendungsfälle erfolgreich gesucht hat, dann wissen Sie anschließend genau, wie Sie Ihre Suche für die beste Learner Experience optimieren können.

Auf eine Gefahr sei allerdings noch in Bezug auf die Suchanalyse hingewiesen: die Gefahr der sich selbst erfüllenden Prophezeiung. Dahinter steckt, dass Lerner natürlich vor allem bei der Mehrfachnutzung der Lernplattform schnell merken, wenn ihre Suche bei bestimmten Sucharten überhaupt nicht funktioniert, bei anderen aber gut. In einem solchen Fall passen die Lerner ihr Suchverhalten an ihre "schlechte" Suche an und Sie sehen in den Suchbegriffen nicht mehr das, was die Nutzer eigentlich suchen.

#### 2.1.7 Technische Herausforderungen und Alternativen

Technisch werden Webanalyse-Tools in der HTML-Seite integriert. Oft kursiert hierfür noch der Name Zählpixel, im Grunde handelt es sich dabei aber um ein kleines JavaScript-Codefragment, das vom jeweiligen Tool bereitgestellt wird.

Zum Nachvollziehen des Nutzerwegs verwenden die Webanalyse-Tools traditionell sogenannte Tracking Cookies. Die Herausforderung in diesem Bereich ist klar: Über den Weg des Cookie-Hinweis verliert man hier potentiell viele Sitzungen, die sich nicht messen lassen. Deswegen bieten viele Webanalyse-Systeme auch alternative Wege mit Tracking ohne Cookie und damit ohne Einwilligung. Dies ist mit kleinen technischen Anpassungen verbunden, aber die in diesem Beitrag beschriebenen Messerkenntnisse für die Learner Experience sind auch auf diesem Weg zu erreichen.

Eine weitere Alternative, die auf den ersten Blick ähnlich aussieht wie die Ergebnisse der Webanalyse, ist eigentlich ein Klassiker der Webanalyse, aber Auswertung der Suchbegriffe

**Tracking Cookies** 

Logfile-Analyse

seit vielen Jahren aus der Mode gekommen. Die Rede ist von der so genannten *Logfile-Analyse*. Logfiles werden vom Webserver angelegt. Der Webserver erfasst alle Seitenaufrufe und auch zugehörige Informationen zu den Nutzern. Der bekannteste Klassiker in diesem Bereich ist AWStats (https://awstats.sourceforge.io/). Da nahezu jeder Webserver standardmäßig Logfiles anlegt, ist dies oft eine sofortige Analysemöglichkeit, auch wenn kein Analysetool implementiert ist. Ein Wermutstropfen besteht allerdings meistens darin, dass die Logfiles in der Standardkonfiguration der meisten Webserver oft innerhalb von 30 oder 60 Tagen gelöscht werden. Insgesamt ist die Logfile-Analyse qualitativ und technisch keine optimale Alternative zur Webanalyse und damit nur eine Notfall-Analysemöglichkeit.

#### 2.2 Reports aus dem System

## Daten direkt aus der Quelle

Eine wichtige Orientierung für die Learner Experience sind die Daten der jeweiligen Lernplattform. Sie liefern die Basisinformationen zu Lernerfolgen und -misserfolgen. Ein Beispiel: Wenn Sie in Ihrem Lernmanagementsystem (kurz LMS) viele Web Based Trainings (kurz WBT) einsetzen, erfasst das System nicht nur die erfolgreichen und nicht erfolgreichen Versuche, sondern auch Informationen dazu wie lange die Nutzer durchschnittlich dafür brauchen. Filtert man diese Erkenntnisse und betrachtet man die nicht erfolgreichen Versuche separat, so lässt sich feststellen, in welchen WBTs die größten Problemstellen verborgen sind.



Abb. 6: Ergebnisse von WBTs auf SCORM-Basis (Report aus Totara Learn)

#### Beispiele für Informationen aus dem System

Der Markt an Lernplattformen und LMS ist extrem groß. Ähnlich groß ist die Vielfalt der Reporting- und Auswertungsmöglichkeiten in den Systemen. Deswegen hier einige Beispiele für Informationen, die in den meisten Systemen einfach verfügbar sind:

- Teilnehmerzahlen nach Kategorien, Kursen und unter Umständen sogar einzelnen Lernaktivitäten
- Verhältnis von Teilnehmern zu erfolgreichen Abschlüssen meist auf Kurslevel, je nach System auch auf Ebene von Curriculae
- Durchführungsdauer von Kursen und von einzelnen Lernaktivitäten
- Test- und Abschlussergebnisse mit dem Durchschnitt und der Verteilung der Noten oder Prozentwerte
- Informationen zu den Aktivitäten der Teilnehmer, beispielsweise der Menge an Logins in die Plattform, aber auch einzelne Interaktionen wie Postings im Forum etc. Interessant ist hier vor allem wie viele Lerner sich von der Plattform zu aktivem Mitmachen motivieren lassen.



Abb. 7: Die Teilnehmerzahlen der Kurse identifizieren erfolgreiche und weniger erfolgreiche Lernangebote (Report aus Totara Learn)

Unter dem Stichwort Learning Analytics kommt den Auswertungen in den Lernplattformen in den letzten Jahren gesteigerte Bedeutung zu. Das hat sich zum einen positiv auf die Reporting-Funktionen der bestehenden Plattformen ausgewirkt, zum anderen gibt es neue Lösungen, die an bestehende Plattformen per Schnittstelle andocken und die gelieferten Daten weiterverarbeiten. Ein Beispiel ist Zoola (https://www.zoola.io/), das zum Open-Source-Marktführer Moodle und zum auf Unternehmen ausgerichteten Pendant Totara eine Schnittstelle anbietet.

Neben solchen Spezialisten, kommen in manchen Unternehmen und Organisationen auch allgemeine Business Intelligence Software wie Tableau, Microsoft Power BI etc. zum Einsatz, um aus den Lernplattformen exportierte Lerndaten zu verarbeiten und grafisch aufzubereiten. In allen Fällen gilt: Die Datenbasis selbst verändert sich dabei nicht, sondern basiert nach wie vor auf dem, was die Lernplattform zur Verfügung stellt. Allerdings hilft die bessere Aufbereitung und die Verknüpfung mit anderen Daten – beispielsweise zum Präsenzlernen – dabei, neue Perspektiven auf die Lerner zu eröffnen.

Wo liegen die Grenzen von Systemreports? Das hängt zum einen natürlich von der eingesetzten Lernplattform ab. Hier gibt es sehr unterschiedliche Reporting-Qualitäten sowohl in der Menge der erfassten Daten als auch bei der Ausgabe. Zum anderen handelt es sich bei den Systemdaten um reine Messung von Interaktionen. Wie sich der Nutzer verhält und warum er sich so verhält, verraten die Daten nicht. In der Praxis macht es dementsprechend Sinn, mögliche Problemstellen aus den Systemdaten abzuleiten, z. B. wenig erfolgreiche Kurse, und anschließend genau an diesen Problemstellen mit Nutzern zu testen, um die Wahrnehmung der Nutzer zu verstehen.

#### 2.3 Mit Nutzern testen

Die beste Messmethode für die Learner Experience ist eindeutig, die Lerner direkt zu beobachten. Damit betrachtet man das eigene Lernangebot durch die Augen des Nutzers. Der typische Test mit Nutzern besteht aus folgenden Schritten:

1) Am Anfang steht die Frage nach dem Testschwerpunkt: Sollen einzelne Lerninhalte begutachtet werden, oder komplexe Prozesse in einem Lernmanagementsystem? Auf Basis dieses Schwerpunkts entstehen dann die Aufgaben für den Test, die die Testprobanden durchführen müssen.

Erweiterte Analysemöglichkeiten

Die Grenzen der Systemreports

Die Nähe zum Lerner ist entscheidend

- 2) Danach folgt die Wahl der richtigen Zielgruppe. Idealerweise sind die Probanden für den Test identisch mit den (möglichen) Lernern der Plattform.
- 3) Den dritten Schritt bildet die Planung der Testsituation. Sie umfasst sowohl die räumliche Planung, als auch die Festlegung der relevanten Rollen wie Moderator und Beobachter.
- 4) Mit verschiedenen Messverfahren lassen sich die Testergebnisse noch weiter verbessern. Die Spanne reicht hier von einfachen Verfahren wie lautem Denken bis hin zu Eye Tracking und Emotionserkennung.
- 5) Nach dem Test startet die Auswertung der Ergebnisse. Sie ist direkt abhängig von den eingesetzten Messverfahren und dementsprechend in den jeweiligen Abschnitten der Messverfahren beschrieben.

#### 2.3.1 Testschwerpunkt

#### Der richtige Fokus ist entscheidend

Die Ziele eines Nutzertests können vielfältig sein: Funktioniert die Registrierung eines Lernmanagementsystems? Wie nutzt ein Lerner die neu erstellten Lernpfade? Welche Probleme hat ein Lerner beim Erfassen eines WBTs? Wichtig ist, für jeden Test einen Schwerpunkt festzulegen, um den Test nicht zu umfangreich oder unspezifisch zu machen. Als Ideenlieferanten für diesen Testschwerpunkt können folgende Quellen dienen:

- **die Ergebnisse der Webanalyse:** Sind die Abbruchraten bei der Registrierung für ein LMS sehr hoch, lohnt sich beispielsweise eine Überprüfung des Registrierungsprozesses mit einem Test, um festzustellen, welche Probleme die Nutzer genau haben.
- die Reports aus dem System: Werden bestimmte Lernaktivitäten beispielsweise besonders selten abgeschlossen, kann ein Nutzertest Aufschluss über Probleme in der Steuerung und Erfassung liefern.
- Feedback von Lernenden: Erreichen die Trainingsadministration z. B. häufiger Fragen zur Lernpfadverwaltung im LMS, so kann dies ein sinnvoller Anlass sein, einen Nutzertest zu planen, um Probleme zu identifizieren.
- **Konzeptionsfragen:** In einer Neukonzeption, beispielsweise beim Einführen neuer Lernformen wie videogestütztem Lernen, entstehen oft spezielle Fragestellungen, bei denen ein Nutzertest auch dazu dienen kann, mehrere Ansätze und Varianten miteinander zu vergleichen.

#### Fragebogen und Aufgaben

Abhängig vom Testschwerpunkt entsteht dann der zugehörige *Fragebogen* mit den Aufgaben für die Probanden. Eine gute Aufgabe ist dabei klar formuliert und verständlich, lässt dem Probanden aber auch entsprechenden Spielraum, verschiedene Wege in der Lernanwendung zu nehmen.

Aufgaben können auch –wie das folgende Beispiel zeigt – aufeinander aufbauen:

- 1) Registrieren Sie sich unter xyz.de als neuer Lerner.
- 2) Suchen Sie sich einen passenden Kurs aus, um mehr über Brandschutz zu lernen.
- 3) Führen Sie den gefundenen Kurs zum Thema Brandschutz durch und beantragen Sie ein Zertifikat.

In diesem Fall sollte nur geregelt sein, ob der Proband bei nicht erfolgreichem Abschluss einer Aufgabe dennoch bei der nächsten Aufgabe weitermachen kann. Im Beispiel: Findet der Proband in Aufgabe 2) keinen passenden Kurs, muss ihm der Moderator für Aufgabe 3) einen Kurs vorgeben, damit der Test nicht an dieser Stelle abgebrochen werden muss.

Der häufigste Fehler beim Fragebogen ist, zu viel auf einmal erfahren zu wollen. Ein normaler Test sollte für den "durchschnittlichen" Probanden in einer Zeit von ca. 30 Minuten durchführbar sein, da bei längeren Tests die Konzentration stark leidet. Mit Einführung, Tests und Befragung sollte ein Test erfahrungsgemäß ca. 60 Minuten nicht überschreiten.

Vorsicht bei der Dauer!

Einziges Problem: Es ist nicht ganz leicht, einzuschätzen wie lange eine einzelne Aufgabe oder ein ganzer Fragebogen dauert. Die Dauer kann in der Praxis zwischen verschiedenen Probanden sehr stark schwanken. Die Empfehlung ist allerdings, auf jeden Fall den Fragebogen mit einem unbeteiligten Kollegen oder einem Testprobanden vorab zu testen.

#### 2.3.2 Probanden

Die Hauptperson bei jedem Nutzertest ist der Proband. Der am besten geeignete Proband kommt direkt aus der Zielgruppe Ihrer Lernanwendung. Soweit, so logisch. Allerdings ist es manchmal nicht so einfach, Probanden direkt aus der Zielgruppe zu akquirieren. In solchen Fällen eignen sich auch Nutzer mit ähnlichem Wissensstand und demografischem Hintergrund. Wichtig ist allerdings immer zu beachten, ob es in der jeweiligen Zielgruppe bestimmte Fachbegriffe oder Wissenshintergründe gibt, die zum Durchführen der Lernanwendung notwendig sind.

Wer eignet sich als Proband?

Die Praxiserfahrung zeigt, dass es beim Testen von Prozessen in Lernmanagementsystemen wesentlich weniger stark darauf ankommt, ob die Probanden exakt aus der Zielgruppe kommen, wohingegen das Testen von Lernaktivitäten stark vom Vorwissen des Probanden abhängt. Dennoch, wenn die Chance besteht, Probanden direkt aus der Zielgruppe zu akquirieren, ist das immer die beste Lösung.

Vor einem Fehler gilt es sich auf jeden Fall zu hüten. Niemals sollte ein Proband aus dem Kreis der für das Projekt verantwortlichen Personen rekrutiert werden. Hier fehlt nicht nur die Neutralität, auch das Verständnis für Zusammenhänge und Begrifflichkeiten ist oft zu nah am bereits bekannten System, um valide Testergebnisse liefern zu können.

Zu große Nähe gilt es zu meiden

Wer Lerninhalte oder -systeme für Kunden entwickelt, sollte sogar komplett auf Kollegen als Testpersonen verzichten, denn auch jedes Unternehmen hat seinen eigenen "Sprech". Das heißt, solche Probanden verstehen die Sprache und Logik einer Anwendung oft besser, als externe Kunden, die sie später aber eigentlich nutzen sollen.

Wieviele Probanden sind notwendig?

Sobald die Zielgruppe identifiziert ist, stellt sich die Frage wieviele Nutzer für einen Test notwendig sind. Die Antwort auf diese Frage unterscheidet sich stark zwischen wissenschaftlich orientierten und praktisch orientierten Tests. Der wissenschaftliche Test arbeitet oft mit 20 oder mehr Probanden pro Zielgruppe. Für einen praktischen Test ist eine solche Menge vollkommen unwirtschaftlich. Nielsen und Loranger haben herausgefunden, dass bereits ein Test mit nur 5 Probanden ca. 80 % der Usability-Probleme einer Anwendung aufdeckt.

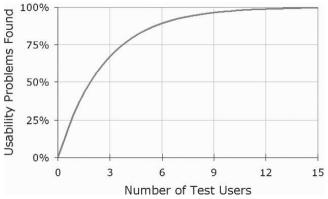


Abb. 8: Die Zahl der notwendigen Probanden nach Nielsen und Loranger (https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/)

# 5-10 Probanden pro Zielgruppe

Berücksichtigt man noch einen kleinen Puffer, da manche Probanden nicht auftauchen oder in seltenen Fällen der Test an sich nicht gut klappt, so ergibt sich eine optimale Zahl zwischen 5 und 10 Probanden pro Zielgruppe. Die Einschränkung "pro Zielgruppe" ist hier wichtig. Ein Beispiel: Beim Testen einer Lernplattform für ein Krankenhaus gibt es mit Pflegepersonal, Ärzten und Verwaltung mindestens drei Zielgruppen, deren Bedürfnisse und Lernverhalten sich u. U. stark unterscheiden. Wollte man alle drei Zielgruppen in einem Nutzertest abwickeln, wären 15 Probanden optimal, 10–12 Probanden mit einer ausgeglichenen Verteilung der Zielgruppen das Minimum für valide Ergebnisse.

#### Probandenakquise

Der Probandenbedarf führt direkt zur Herausforderung der Probandenakquise. Wenn es um unternehmensinternes Lernen geht, ist dies meist eher unproblematisch. Bei einem eingeschränkten Kundenstamm – häufig im B2B-Bereich – lohnt es sich ebenfalls, direkt geeignete Kunden um die Mitarbeit zu bitten. Aller Erfahrung nach sind sehr viele Ansprechpartner hierfür sehr offen. Etwas aufwendiger wird es im B2C-Bereich. Dort sind vor allem spezielle Zielgruppen oft schwer zu akquirieren. Unter Umständen kann hier ein kommerzieller Usability-Dienstleister oder ein Umfrageinstitut mit seiner Kartei unterstützen.

#### Entlohnung für Probanden

Eine vieldiskutierte Frage bei der Probandenakquise ist die angemessene Entlohnung. Eine kleine Aufmerksamkeit hat sich auf jeden Fall bewährt, bei gekauften Probanden gilt es allerdings vorsichtig zu sein. Gerade professionelle "Tester" sind oft zu weit von der normalen Zielgruppe entfernt und spulen nur ein gelerntes Programm ab.

#### 2.3.3 Testsituation

#### Wo wird getestet?

Beim eigentlichen Test stellt sich als erste Frage die Frage nach dem Ort, an dem der Test stattfinden soll. Hier gibt es verschiedene Möglichkeiten, die alle ihre Vor- und Nachteile haben:

- Im Labor: Das klingt medizinisch, kann aber auch ein ganz normales Büro sein, das als Usability-Test-Umgebung eingerichtet ist. Vor allem die aufwendigeren Messverfahren wie Eye Tracking werden meist in einem Büro durchgeführt. Der große Vorteil des Labors ist, dass alles optimal auf die Testsituation vorbereitet ist und z. B. für Beobachter ein eigener Raum zur Verfügung steht. Der größte Nachteil des Labors ist, dass die Probanden erst dort hingelangen müssen.
- **Beim Probanden in seiner natürlichen Umgebung:** Der Vorteil ist, dass sich der Proband hier wohl fühlt und weniger angespannt ist. Der größte Nachteil liegt darin, dass der Testrechner und alle Messverfahren transportiert und bei jedem Probanden einzeln aufgebaut werden müssen. Auch die notwendigen Reisezeiten für das Testteam steigen immens.
- Im Büro, einem Konferenzraum oder in einem Hotelzimmer: Der Vorteil gegenüber dem Labor ist, dass ein beliebiger Ort näher am Probanden gewählt werden kann. Außerdem lassen sich hier Tests mit mehreren Probanden durchführen und der Aufbauaufwand ist geringer als beim Test direkt in der natürlichen Umgebung des Probanden. Ein im B2B-Bereich sehr erfolgreicher Ansatz ist zum Beispiel das Durchführen von Tests im Rahmen von Messen oder Kundentreffen. Die fortgeschrittenen Messverfahren wie Eye Tracking und Stressmessung sind unter Umständen nutzbar.
- Remote-Test per Webmeeting: Der Remote-Test per Webmeeting empfiehlt sich vor allem beim Einsatz von Testpersonen, die aus den unterschiedlichsten Regionen, aus anderen Ländern oder gar anderen Kontinenten stammen, da hier der hohe Reiseaufwand entfällt. Der größte Nachteil liegt darin, dass hier die Moderation und auch die Führung des Probanden schwieriger werden. Außerdem sind die Testbedingungen, z. B.

die eingesetzten Endgeräte, bei den verschiedenen Probanden höchst unterschiedlich. Die fortgeschrittenen Messverfahren sind in diesem Szenario nicht nutzbar.

Ein häufiger Fehler geschieht schon bei der Einladung der Probanden: Oft wird dort bereits verraten, um welchen Testfall es geht. Leider gibt es immer wieder Probanden, die einen Usability-Test eher als Test der eigenen Fähigkeiten begreifen und sich deswegen vorab mit der entsprechenden Anwendung beschäftigen. Dies nimmt dem Probanden den ersten, unvoreingenommen Blick und verschlechtert damit die Testergebnisse erheblich.

Bei Nutzertests unterscheidet man neben dem Probanden noch zwei wichtige Rollen: *Moderator* und *Beobachter*. Der Moderator begleitet den Probanden von der Begrüßung über das Stellen der Aufgaben bis hin zum Stellen der Abschlussfragen und der Verabschiedung. Im Gegensatz zum Moderator hat der Beobachter keinen oder kaum Kontakt zum Beobachter. Seine Aufgabe ist, Reaktionen des Probanden zu beobachten und zu vermerken.

Für die einfachste Art der Testdurchführung reicht eine Person in der Doppelrolle als Beobachter und Moderator generell aus. Allerdings zeigt die Erfahrung, dass die Teilung trotz des erhöhten Personaleinsatzes effizienter ist. Denn der Beobachter spart mit seiner Arbeit oft sehr viel Zeit in der Auswertung.

Die wichtigste Eigenschaft beim Moderator ist Neutralität. Das heißt, der Moderator sollte den Probanden beruhigen, aber auch in Ruhe arbeiten lassen, wenn er bei einzelnen Aufgaben Schwierigkeiten hat. Das kann durchaus so weit gehen, dass der Moderator einen Probanden auch einmal länger bei der Durchführung einer Lernaktivität leiden lassen muss, um stressige Passagen zu identifizieren. Und noch wichtiger, der Moderator sollte niemals Probleme rechtfertigen oder frühzeitig bei Problemen des Probanden eingreifen.

All das erfordert vom Moderator eine emotionale Distanz zum Untersuchungsgegenstand. Die Praxis zeigt, dass es deswegen sehr empfehlenswert ist, nicht als Produktverantwortlicher selbst zu moderieren, sondern sich maximal in die Rolle des Beobachters zu begeben. Wer also sein eigenes Lernsystem testet, sollte idealerweise einen neutralen Kollegen als Moderator gewinnen.

Den wichtigsten Teil eines Nutzertests bilden sicherlich die Aufgaben, die der Proband mit dem Lernsystem erledigt. Dort wird sichtbar, was der Nutzer erlebt und wie er es wahrnimmt. Daneben hat es sich aber bewährt, den Probanden im Umfeld des Tests zu befragen. Generell ist die Fragengestaltung hier sehr frei; bewährt haben sich folgende Ansätze:

- Der Proband sollte, bevor der Fragebogen zum Einsatz kommt, mit einem allgemeinen "Warmup" an die Testsituation herangeführt werden. In diesem Umfeld haben sich allgemeine demografische Fragen bewährt.
- Vor dem Einsatz des Fragebogen können auch allgemeine Erwartungshaltungen an eine Lernanwendung erfragt werden, z.B. wie der Nutzer normalerweise lernt und welche Erwartungen er an ein Lernsystem hat.
- Nach Abschluss des Fragebogen sollte mit normierten Fragen die Zufriedenheit des Probanden ermittelt werden. Normiert deswegen, weil so die Vergleichbarkeit zwischen verschiedenen Probanden gegeben ist.
- Neben den normierten Fragen sollten auch Besonderheiten erfragt werden,
   z. B. besondere Schwierigkeiten bei einer bestimmten Aufgabe.

Um die Auswertung später zu erleichtern, sollte neben den Notizen des Beobachters auch der Screen des Nutzers beim Durchführen der Aufgaben aufgezeichnet werden. Für diese Aufgabe gibt es mehrere Tool-Klassen:

Vorsicht bei der Einladung

Testbegleitung und -durchführung

Neutralität ist entscheidend

Befragung während des Tests

Tools zur Aufzeichnung

- Die meisten Webmeeting-Softwaren erlauben nicht nur das Webmeeting an sich, sondern bieten auch eine integrierte Aufzeichnungsmöglichkeit und können dementsprechend für einen Test zweckentfremdet werden.
- Der Markt bietet eine Vielzahl an professionellen Screen-Recording-Tools wie SnagIt und Camtasia von TechSmith oder Bandicam. Diese Spezialinstrumente kosten zwischen 50 und 300 €. Die Funktionen unterscheiden sich vor allem in der Videoverarbeitung.
- Neben den professionellen, kommerziellen Recording-Tools gibt es auch eine Vielzahl an, meist etwas einfacher aufgebauten, kostenfreien Screen-Recording-Tools. Manche davon sind spezialisiert auf bestimmte Betriebssysteme; es gibt aber auch auf allen Plattformen verbreitete Systeme wie OBS (Open Broadcast Software).
- Ein reiner Usability-Spezialist ist Morae von TechSmith. Hier kann der Beobachter direkt im Video live während der Aufnahme Aufgaben und Problemstellen markieren. Mit einem Preis von über 2.000 € ist das auf Windows beschränkte Tool allerdings recht teuer und rentiert sich normalerweise nur im Labor-Alltag.
- Problematisch ist die Aufzeichnung von Smartphone-Tests, da die Smartphone-Betriebssysteme aus Sicherheitsgründen Screenrecording nicht
  erlauben. Hier müssen u. U. Ersatzmöglichkeiten wie die Aufzeichnung per
  Apple TV genutzt werden.

#### 2.3.4 Lautes Denken

#### Im Kopf des Nutzers

Das Messverfahren "Lautes Denken" kommt ohne Gerätschaften aus. Der Moderator animiert den Probanden dazu, seine Aktionen in der Plattform mit laut gesprochenen Gedanken zu begleiten. Diese gesprochenen Gedanken können aufgenommen oder vom Beobachter mitgeschrieben werden und sind oft sehr aufschlussreich.

Einen Nachteil an diesem Ansatz gibt es allerdings: Sie funktioniert mit extrovertierten Probanden deutlich besser als mit introvertierten Probanden. Unter Umständen muss der Moderator immer wieder eingreifen und motivieren, um den Probanden zum lauten Denken zu animieren.

#### Aufzeichnung des Probanden

Sollten Sie den Probanden in Bild und Ton aufzeichnen wollen, beachten Sie bitte unbedingt den **Datenschutz**. Die schriftliche Einverständniserklärung für die Verwendung im Rahmen des Nutzertests stellt den Mindeststandard dar.

#### 2.3.5 Eye Tracking

## Durch die Augen des Nutzers

Das Messverfahren "Eye Tracking" zeichnet die Blickbewegungen des Probanden auf, folgt also den Augen auf dem Bildschirm und erfasst, welche Bereiche der Nutzer wahrnimmt. Für die Messung von Augenbewegungen gibt es zwei gängige Methoden: Auf der einen Seite so genannte Eye Tracking-Brillen, die der Proband aufsetzt, auf der anderen Seite zwei fest am Monitor installierte Hochgeschwindigkeitskameras, die den Augenbewegungen des Nutzers folgen. Die Eye Tracking-Brillen sind generell deutlich kostengünstiger (ca. 5.000 bis 15.000  $\bigcirc$ ), während die Preise der kamera-basierten High-End-Eye Tracker bei 25.000  $\bigcirc$  beginnen. Dafür bieten sie auch deutliche Vorteile: Der Proband wird nicht durch eine ungewohnte Brille gestört und verhält sich damit deutlich natürlicher. Außerdem ist die Datenerfassung meist genauer und enthält auch Informationen zur Größenänderung der Pupille, die zur Stressmessung herangezogen werden können. Die notwendige Konfigurationszeit ist bei beiden Methoden ungefähr gleich.

#### Mieten statt kaufen

Aufgrund der hohen Kosten ist es meist nicht empfehlenswert, sich als Unternehmen oder Organisation selbst einen Eye Tracker anzuschaffen. Hier bieten

viele Agenturen und Usability-Labore die Möglichkeit den Service, das Labor und das Equipment zu mieten oder/und bei der Durchführung mit geschultem Personal zu unterstützen. Letzteres ist für unerfahrene Tester sinnvoll, da die Konfiguration und Aufzeichnung mit Eye Trackern ein wenig Erfahrung voraussetzt.



Abb. 9: Ein hochwertiger Eye Tracker aus dem Labor

Die häufigste Darstellungsform der Ergebnisse von Eyetracking sind so genannte Heatmaps. Auf diesen Screenshots der Anwendung sind die Bereiche meist mit dem Farbbereich gelb bis rot eingefärbt, die der Nutzer wahrgenommen hat, wobei die roten Bereiche eine besonders intensive Wahrnehmung symbolisieren.

Dabei sind in der Auswertung oft größere Abstände zwischen den einzelnen wahrgenommenen Bereichen zu sehen. Das liegt daran, dass das menschliche Auge sich nicht kontinuierlich bewegt, sondern von Punkt zu Punkt springt. Die Bewegungen heißen dabei Sakkaden. Während dieser (sehr schnellen) Bewegung erfolgt keine Wahrnehmung, diese erfolgt erst bei der nächsten sogenannten Fixation. Von der Länge dieser Fixation hängt es ab, ob Wahrnehmung entsteht. Grob ist dies bei einer Fixationsdauer von 200 bis 250 Mil-

Die Augen springen

lisekunden der Fall. Je länger die Fixation bzw. je mehr Fixationen rund um einen bestimmten Ort stattfinden, desto "heißer" wird der jeweilige Bereich der Heatmap.

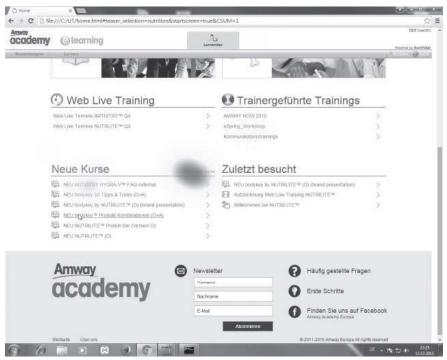


Abb. 10: Heatmap-Vergleich mit zwei Probanden (Proband 1)



Abb. 11: Heatmap-Vergleich mit zwei Probanden (Proband 2)

Die ersten Rückschlüsse für die Auswertung liefert der Vergleich verschiedener Probanden, wie in Abbildungen 10 und 11 zu sehen. Bei der dort gezeigten Aufgabe, sollte der Proband die Hilfe und die FAQ der Seite finden, die sich rechts unten im Footer befinden. Bei Proband 1 ist erkennbar, dass er diesen Bereich kaum wahrgenommen hat, dementsprechend konnte er die Aufgabe nicht erfolgreich abschließen. Proband 2 hat ebenfalls einige Zeit gesucht und ist über die gesamte Seite gesprungen, hat aber am Schluss den richtigen Bereich identifizieren können.

Neben dem Vergleich verschiedener Probanden, macht es auch Sinn, innerhalb einer Seite oder Anwendung so genannte Areas of Interest (kurz AOI) zu definieren. Dabei handelt es sich um Bereiche im Inhalt, deren Wahrnehmung überprüft werden soll. Abbildung 12 zeigt ein Beispiel für ein WBT, mit dem geprüft wird, ob der Nutzer z. B. die Navigation erfasst und den textlichen Inhalt wahrnimmt.

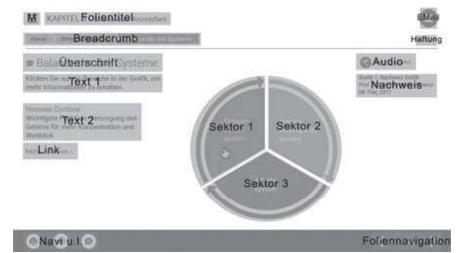


Abb. 12: Unterteilung eines WBT in Bereiche

In der Auswertung liefert die Eye Tracking Software dann für die einzelnen Bereiche (AOIs) folgende Kennzahlen:

- Time to first fixation die Zeit bis zur ersten Fixation des Bereichs.
- Hitrate der Blick war innerhalb des definierten Bereichs.
- Gaze duration Blickdauer in dem jeweiligen Bereich.
- Fixation count die Anzahl der Fixationen des jeweiligen Bereichs.

Diese Auswertung liefert ein klares Indiz, welche Bereiche der Proband wahrgenommen hat. In Verbindung mit den Aufgaben des Fragebogens lassen sich diese Zahlen dann noch ins Verhältnis setzen, und zeigen, ob der Proband bei der Aufgabe erfolgreich war (häufig TCR für Task Completion Rate genannt) und wie lange er für die Aufgabe gebraucht hat (TOT für Time on Task). Wenn ein Fragebogen Aufgaben zur Lernerfolgskontrolle enthält, beispielsweise in einem WBT einen Abschlusstest, so lässt sich für die Messung des Lernerfolgs auch die Wahrnehmung von bestimmten Bereichen ins Verhältnis setzen zum erfolgreichen Abschluss bestimmter Testfragen.

Hierzu ein Beispiel aus der Testpraxis: Beim Test eines WBTs mit eingebundenen Lernvideos stellte sich heraus, dass 7 von 10 Probanden immer die gleichen zwei Testfragen falsch beantwortet haben, bei anderen Fragen aber alle Probanden einen guten Schnitt erzielt haben. Bei der Prüfung der Eye-Tracking-Auswertung hat sich herausgestellt, dass die Information der einen

#### Auswertung

Kennzahlen für die AOIs

Die Verbindung zum Fragebogen Testaufgabe in einem sehr langen Textblock "versteckt" war und von den Probanden nicht wahrgenommen wurde. Bei der anderen Testaufgabe war die Informationen Teil eines animierten Videos, bei dem die Wahrnehmung der Probanden von bunten "Störanimationen" abgelenkt war. Nach Beseitigung dieser zwei Wahrnehmungsstörungen wurde das WBT für Lerner freigegeben und die zwei Testfragen wurden von 98 % der Lerner im LMS korrekt beantwortet.

#### 2.3.6 Stressmessung

#### Von der Belastung zur Überlastung

Das Konzept hinter der Stressmessung basiert auf dem Forschungsgebiet der Mental Workload oder auch kognitiven Workload. Auf wissenschaftlicher Ebene gibt es hier verschiedene Definitionsansätze. Für die Praxis reicht es aus, sich vorzustellen, dass Aufgaben für den Menschen mentale Arbeit bedeuten.

Ein Beispiel stellt das Radfahren dar. Wer Radfahren neu lernt, hat am Anfang eine hohe Workload, alles richtig zu machen und ist von einer komplexen Schaltung schnell überfordert. Wenn er es dann kann, steigt zum einen seine Workload-Effizienz, mit der er diese spezifische Aufgabe erfüllen kann, zum anderen sinkt auch die benötigte kognitive Energie, da die Handlungsweisen automatisiert sind. Aber auch, wenn der Lernende dann Radfahren kann, können externe Störfaktoren seine Mental Workload erhöhen. Angenommen, der Radfahrer muss ganz schnell an sein Ziel gelangen, hat sich aber verfahren und versucht nun, sich von seinem Smartphone durch unbekannte Gegenden leiten zu lassen: In diesem Fall erhöht sich seine Mental Workload und wenn er diese "Last" nicht mehr als fordernd, sondern als überfordernd empfindet, entsteht Stress.

#### Ein bisschen Denken ist doch normal?

Die Schwierigkeit besteht nun darin, Belastungen durch Schwierigkeitsgrade oder Arbeitsmengen (Mental Workload) von schädlichem Stress zu unterscheiden, denn jede Lernaktivität wird und soll Mental Workload erzeugen. In der Praxis haben sich dafür folgende Ansätze bewährt:

- Nicht-lernrelevante Prozesse in Lernsystemen sollten so wenig wie möglich Mental Workload erzeugen. Dazu zählen Registrierungen, Einschreibungen und vor allem die Navigation durch die Plattform.
- Bei lernrelevanten Prozessen gilt, dass das Lernen selbst Mental Workload erzeugen darf, aber nicht in Stress ausarten sollte. Dies ist über den Vergleich verschiedener Aufgaben messbar. In Abbildung 13 erreicht die Kurve einer Probandin bei einer von fünf Aufgaben ein deutlich erhöhtes Niveau: ein klares Anzeichen von Stress.

## EDA-Verlauf Probandin CMED20 während der Registrierung

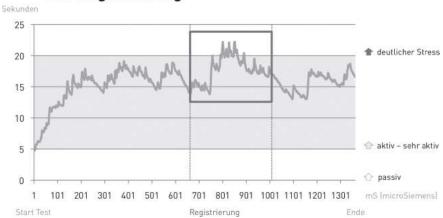


Abb. 13: Verlauf der Stressmessung bei einem Labortest

An dieser Stelle sei eine kurze Stellungnahme gestattet zu den antiquierten Lehr- und Usability-Ansichten, die mancher aus der Schule, der Uni-Zeit oder auch von Berufsausbildungen kennt. In diesem Zusammenhang hört man oft, dass man von jemandem der "XYZ werden will" doch schon ausreichend Intelligenz und Durchhaltevermögen erwarten kann, sich durch unübersichtliches Lehrmaterial zu quälen. Diese Ansicht ist falsch: Wir alle können und sollen als Lernende unsere kognitiven Fähigkeiten einsetzen, aber zielgerichtet für das Lernen und Begreifen, nicht für die Erfassung schlechter Lehrmaterialien oder für lausige Bedienbarkeit in der Lernplattform. Jede unnötige Hürde, die abgebaut wird, erlaubt dem Lerner, seine mentalen Fähigkeiten auf den Lernstoff zu konzentrieren und verbessert damit direkt die Learning Experience.

Die Stressmessung kann mit verschiedenen Geräten erfolgen:

• Die für Usability-Tests praktischste Methode ist die Messung der Hautleitfähigkeit. Dafür kommt ein so genanntes *EDA-Gerät* (Elektrodermale Aktivität) zum Einsatz. Der Proband erhält dazu einen Sensor auf die Haut, meist direkt an der Hand. Einfache Systeme gibt es bereits ab 100 €, die Konfiguration ist einfach, manchmal sogar direkt mit einer Smartphone-App möglich und die Störung für den Probanden ist recht gering.



Abb. 14: Ein EDA-Gerät im Einsatz

• Im wissenschaftlichen Bereich wird oftmals ein *EEG* (Elektroenzephalografie) eingesetzt. Es wird auf dem Kopf befestigt und misst direkt die Hirnströme. Im Gegensatz zum EDA-Gerät ist es damit eine direkte Messung. Neben den hohen Anschaffungskosten sind vor allem das unhandliche Aufsetzen und das Konfigurieren für praktische Usability-Tests zu aufwendig.

Nur die Harten ...

Messmethoden



Abb. 15: Ein EEG-Gerät

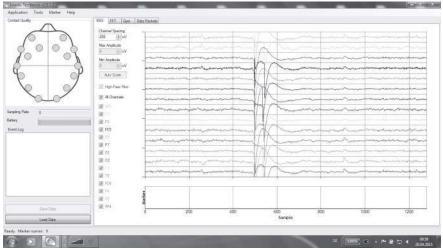


Abb. 16: EEG-Messergebnisse mit einem deutlichen Ausschlag

Mit Hochgeschwindigkeits-Eye-Trackern können auch Pupillengrößenänderungen gemessen werden. Bestimmte Pupillengrößenänderungen sind – basierend auf dem Konzept des Fluchtreflex – ebenfalls für die Stressmessung nutzbar. Dieses Verfahren erfordert keine weiteren Messgeräte außer dem Eye Tracker. Allerdings erfordert die Auswertung der Daten erhebliche Rechenkapazitäten, weswegen die Methode auch eher der Wissenschaft vorbehalten bleibt.

#### 2.3.7 Emotionserkennung

Die Emotionserkennung basiert auf der Gesichtserkennung und -analyse. Voraussetzung sind sehr gute und vor allem gut ausgeleuchtete Kameraaufnahmen des Probanden. Mit AI-Softwarelösungen wie dem FaceReader von Noldus (https://www.noldus.com/) können diese Aufnahmen anschließend ausgewertet und die wahrgenommen Emotionen gewichtet werden.

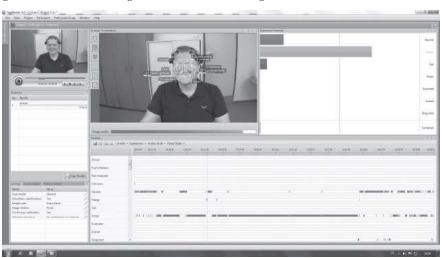


Abb. 17: Die Auswertung der Emotionserkennungssoftware

# Emotionen der Probandin CMED20 während der Registrierung

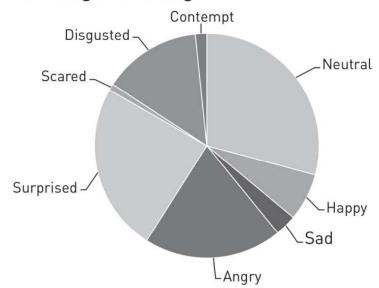


Abb. 18: Beispielhafte Auswertung für eine Emotionserkennung

Gefühle erforschen

#### Keine exakte Wissenschaft, aber eine Tendenz

Die Emotionserkennung ist dabei keine präzise Vorhersage, sondern liefert eher eine Indikation. Überwiegen die negativen Emotionen (grau), so ist der jeweilige Prozess für den Nutzer nicht angenehm. Hier ist ähnlich wie in der Stressmessung vor allem die Veränderung beim Probanden interessant. Sind seine Emotionen auf der Startseite negativer als in einer Lernaktivität, kann das beispielsweise am dort eingesetzten Bildmaterial liegen oder die Form der Ansprache hat ihn verwirrt. Empfehlenswert ist der Einsatz der Emotionserkennung als Ergänzung zu den anderen Methoden immer dann, wenn die emotionale Wirkung eines Lernsystems oder einer Lernaktivität gesondert geprüft werden soll. Liegt der Fokus dagegen hauptsächlich auf der Beseitigung von für die Learner Experience störenden Usability-Problemen, reichen Eye Tracking und Stressmessung aus.

#### 2.4 Crowdtests

## Schnell und mit vielen Nutzern

Die Crowdtests fallen im Grunde auch in die Gattung der Tests mit Nutzern, allerdings entfällt der direkte Kontakt zu den Probanden; diese werden von einer Crowdtest-Plattform wie z. B. Testbirds (https://www.testbirds.de/) oder Rapidusertests (https://rapidusertests.com/) bereitgestellt. Die möglichen Testarten sind dabei umfangreich: Von der einfachen Befragung mit Fragebogen bis zum Remote-Interview mit Beobachtung sind sehr viele Möglichkeiten gegeben.

# Die richtige Zielgruppe?

Generell kann sich jeder bei den Crowdtest-Plattformen auch selbst als Proband zur Verfügung stellen. Je nach Aufwand des Tests gibt es hier dann Entlohnungen für die Probanden von 5 bis  $50\,\mathrm{C}$ , in Einzelfällen auch mehr. Gerade im technik-affinen Umfeld und allgemein mit B2C-Zielgruppen ergibt sich so ein großer Pool an Testern. Je spezieller die Zielgruppen werden, kann es allerdings auch schwierig sein, die richtigen Probanden ausfindig zu machen. Wer sich selbst mal in die Perspektive eines Probanden begeben möchte, kann sich auch selbst bei einer der Plattformen anmelden und die Testsituation ausprobieren.

#### Ersatz oder Ergänzung?

Crowdtests eignen sich aufgrund des verringerten Organisationsaufwands für schnelle und häufige Tests recht gut. So lassen sich beispielsweise verschiedene Varianten eines Kurskatalogs recht einfach miteinander vergleichen. Ein vollständiger Ersatz für Tests mit Nutzern sind sie allerdings nicht, denn der Test mit Nutzern bietet durch die genaue Beobachtung – unabhängig von den eingesetzten Messverfahren – mehr direkten Einblick in das Verhalten und Erleben der Nutzer.

#### 2.5 a/b- und multivariate Tests

#### Die E-Commerce-Methode fürs Lernen adaptieren

Die Technik von a/b-Tests und multivariaten Tests ist schon recht alt und vor allem im E-Commerce-Bereich sehr etabliert. Dort kann schon die Farbe von Schaltflächen einen Unterschied bei den Verkaufszahlen ausmachen, dementsprechend lohnt sich der Einsatz von vergleichenden Tests.

Genau dieser Vergleich hat dem **a/b-Test** auch seinen Namen gegeben. Das Prinzip ist simpel, ein Teil der Nutzer (z. B. 50 %) erhält Variante A einer Seite, der andere Teil Variante B. Anschließend wird gemessen, welche Variante im Verhältnis zu einem bestimmten Ziel, z. B. Finden des Kurskatalogs in der Lernplattform, das bessere Ergebnis erzielt. Diese Variante wird dann behalten und u. U. mit weiteren a/b-Tests verbessert.

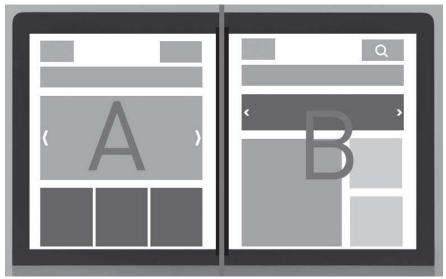


Abb. 19: Der a/b-Test vergleicht zwei Varianten

Der *multivariate Test* ist die komplexere Version des reinen a/b-Tests. In diesem Fall werden mehrere Kriterien festgelegt, beispielsweise Farbe der Schaltflächen, Position der Suche, Größe der Überschrift und in verschiedenen Kombinationen an die Nutzer ausgespielt. Aus dem Klickverhalten der Nutzer errechnet das Testtool dann, welche Kriterienkombination die erfolgreichste ist.

Für klassische a/b- und multivariate Tests haben sich ein paar Best Practices etabliert:

- Tipps & Tricks
- Tests funktionieren nur mit klarem Ziel, das heißt, es sollte eine Kennzahl feststehen, auf die optimiert wird.
- Die Änderungen zwischen den verschiedenen Versionen dürfen nicht zu umfassend sein. Der Vergleich von zwei völlig unterschiedlichen Navigationsleisten funktioniert mit a/b-Tests nur sehr eingeschränkt, das Testen von Positionen, Schaltflächen und Benennungen dagegen sehr gut.
- Alle Varianten sollten von einer ausreichenden Menge von Nutzern über einen längeren Zeitraum getestet werden, um Seiteneffekte sicher auszuschließen. Wie hoch die Nutzerzahl sein muss, hängt stark vom jeweiligen Einsatzgebiet ab. Im B2C-Bereich haben sich beispielsweise Zeiträume von 1 bis 4 Wochen und zumindest mehrere 100 Nutzer bewährt. Wichtig ist auch, Seiteneffekte wie Wochentage oder Uhrzeiten weitgehend auszuschließen, deswegen lautet die Empfehlung, immer zumindest 7 Tage inklusive einem Wochenende zu testen.
- Wichtig ist die Bereitschaft, durchgeführte Änderungen auch wieder zurückzunehmen, wenn sie wider Erwarten nicht funktionieren. In der Praxis sind das sicherlich 5–10 % der Optimierungsmaßnahmen.

Generell können Variantenvergleiche auch direkt in einem LMS oder Autorentool durchgeführt werden, wenn die Messbarkeit über z. B. ein Webanalyse-Tool sichergestellt ist. Alternativ gibt es am Markt eine Reihe an Spezialisten wie Google Optimize, Optimizle (https://www.optimizely.com/) oder den Visual Web Optimizier (https://vwo.com/), und auch Webanalyse-Tools wie Matomo und eTracker bieten a/b-Test-Möglichkeiten.

Tools

Die Spezialisten bieten dabei zwei Arten von Tests, die sich danach unterscheiden, wo die Varianten technisch durchgeführt werden:

• **Direkt im Tool erstellte Tests:** Hierfür lädt das a/b-Tool das Frontend der Webanwendung und erlaubt es, die Varianten über eine Oberfläche zu bearbeiten. Auch das Ausspielen der Variante übernimmt das jeweilige Tool, das heißt, für die Variante wird beispielsweise jeder zweite Nutzer auf eine Website von Google Optimize oder Optimizle weitergeleitet, ohne dass er es merkt. Dazu müssen Sie allerdings selbstverständlich innerhalb der Webanwendung ein entsprechendes Stück Code des jeweiligen Tools einfügen. In Abbildung 20 ist die Bearbeitung der udemy-Startseite in Google Optimize zu sehen, da hier natürlich kein passender Code in der Website ausgeliefert wird, kann dieser kleine Test leider nie live gehen.

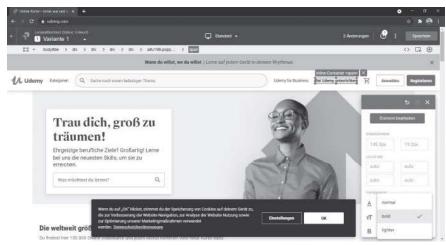


Abb. 20: Eine Variante der udemy-Startseite in Google Optimize

• In der Webanwendung (serverseitig) erstellte Varianten: In diesem Fall leitet das a/b-Tool den Nutzer auf zwei verschiedene URLs, die allerdings von der Webanwendung selbst bereitgestellt werden. Diese Variante ist immer dann zu wählen, wenn es um Unterschiede geht, die serverseitige Logik z. B. des Lernsystems benötigt. Im E-Learning-Bereich ist das sehr häufig der Fall, da hier meist Angaben zum Nutzernotwendig sind oder Varianten direkt in das Autorensystem eingebaut werden müssen.

Der Einsatz im Lernbereich ist neu, aber vielversprechend Im E-Learning-Bereich sind a/b- und multivariate Tests bisher noch wenig im Einsatz, allerdings funktionieren erste Praxistests ausgesprochen gut. Einzig die Implementierung ist meist etwas komplizierter als bei Websites oder im E-Commerce, da oft serverseitig erstellte Varianten notwendig sind. Nachfolgend ein paar Beispiele aus unserer Praxis und weitere Anregungen für mögliche Einsatzbereiche:

- Tests für die optimale Positionierung von Suche, Login und Registrierung.
- Tests, ob sich kleinere Änderungen des Kurskatalogs positiv auf die Buchungszahlen auswirken
- Anpassungen auf der Kursdetailseite, die die Übersichtlichkeit für den Nutzer erhöhen und zu höheren Abschlüssen führen sollen
- Tests für die Schaltflächenpositionierung und -farbgestaltung in der WBT-Navigation

Jenseits dieser, den klassischen a/b-Tests entnommenen Szenarien, können natürlich auch vollständige Varianten von einzelnen Lernformaten ausgespielt werden. Beispielsweise erhält die eine Hälfte der Nutzer ein Erklärvideo mit

menschlichen Protagonisten und die andere Hälfte eines mit gezeichneten Szenarien, anschließend führen alle Nutzer denselben Abschlusstest durch. In diesem Fall liefert das Testergebnis einen Indikator, mit welcher Videovariante das Lernen besser funktioniert hat. Die Auslieferung der verschiedenen Lernformen und die Kopplung der Testergebnisse übersteigt die Fähigkeiten eines klassischen a/b-Testtools, lässt sich aber mit fortgeschrittenen LMS-Systemen durchaus durchführen.

## 3 Optimierungs- und Messbedarf erkennen

Die Optimierung von Learner Experience ist kein einmaliges Projekt, sondern eine kontinuierliche Aufgabe. Dementsprechend sollte die Messung mit den hier vorgestellten Messmethoden zum einen als kontinuierlicher Prozess etabliert werden (z. B. Webanalyse und Reports aus dem System) und zum anderen sollte bei wichtigen Änderungen oder, wenn akuter Optimierungsbedarf erkennbar wird, immer akut gemessen werden.

Der akute Optimierungsbedarf kann zum einen durch das kontinuierliche Monitoring sichtbar werden. Zum anderen hilft es sehr, wenn etablierte Kanäle vorhanden sind, um auf Probleme der Lerner aufmerksam zu werden. Hierfür bieten sich zwei Ansätze an: Die kontinuierliche Beobachtung von Lernerfeedback und die Befragung.

#### 3.1 Feedback & Support

Es klingt selbstverständlich, wird aber leider selten gelebt: Das direkte Feedback der Lerner ist extrem wertvoll. Hier lohnt ein Blick, wo dieses Feedback eigentlich aufläuft. Meist ist das die Trainingsadministration oder der Support im Unternehmen. Wichtig ist, hier einen ständigen Draht zum für die Learner Experience verantwortlichen Team herzustellen (wenn das nicht sowieso schon in einer Hand liegt).

Im direkten Feedback landen meist konkrete Probleme, z. B. mit der Registrierung oder dem Einschreiben in einen Kurs etc. Oft lassen sich hier aber auch allgemeine Probleme ablesen, z. B. dass Nutzer immer wieder die gleichen Probleme haben, um an ein Teilnehmerzertifikat zu kommen. Wichtig ist allerdings, die Informationen zu filtern. Es gibt auch "ewige" Nörgler und selbstverständlich ist es auch wichtig, einzuschätzen wieviele Problemstimmen im Verhältnis zur gesamten Lernerschar vorliegen. Wichtig ist, die goldene Mitte zu finden zwischen völligem Ignorieren von Lernerfeedback (leider nicht so selten, wie man glaubt) und direktem Reagieren auf jedes Feedback, was schnell ins Usability-Chaos führt. Hierzu hilft es auch, wenn das Feedback Problemstellen liefert, diese dann über die Messmethoden aus dem letzten Kapitel zu verifizieren.

#### 3.2 Befragungen

Die zweite wichtige Quelle zum Entdecken von Optimierungsbedarf ist die gezielte Durchführung von Befragungen. Befragungen können auch als Teil eines Nutzertests eingesetzt werden, sind andererseits aber auch eigenständig einsetzbar. Die unabhängigen Befragungen lassen sich grob in zwei Arten unterscheiden: Zum einen die direkte Befragung des Nutzers nach seiner Meinung zur Learner Experience und Usability. Zum anderen die Befragung mit konkreten Fragen zu Detailaspekten der Learner Experience.

Die direkte Frage nach der Learner Experience bzw. Usability einer Lernanwendung hat einen großen Nachteil: Nutzer sind keine Experten darin, auszudrücken, ob sie etwas gut finden und vor allem, weshalb sie es kritisieren.

Auf Lerner hören

Feedback positiv annehmen

Befragungen

**Direkte Befragung** 

Auch Effekte wie sozial erwünschte Antworten können das Ergebnis einer solchen direkten Befragung verfälschen. Dennoch kann vor allem bei regelmäßiger Befragung mit immer dem gleichen Fragebogen hier zumindest Veränderungen identifizieren und rechtzeitig vor Problemen warnen.

Die bekannteste, fragebogen-basierte Messung ist der *System Usability Scale* (SUS) (https://www.usability.gov/how-to-and-tools/methods/system-usability-scale. html). Er errechnet aus 10 Fragen mit jeweils fünf Antwortmöglichkeiten einen Score von 0 bis 100. Allerdings ist der Score nicht als Prozentwert zu sehen und die Orientierung am Durchschnitt von Anwendungen (bei 68) kann sehr missverständlich sein. Wichtig ist, immer zu bedenken, dass die Werte zwischen verschiedenen Anwendungen nicht direkt vergleichbar sind, weil verschiedene Zielgruppen in verschiedenen Kontexten befragt wurden und weil alleine die Art der Anwendungen schon große Unterschiede in der Wahrnehmung erzeugen. Wer diese durchaus großen Einschränkungen akzeptiert, kann mit einer regelmäßigen an den SUS angelehnten Befragung zumindest ein generelles Stimmungsbild seiner Anwendung erhalten.

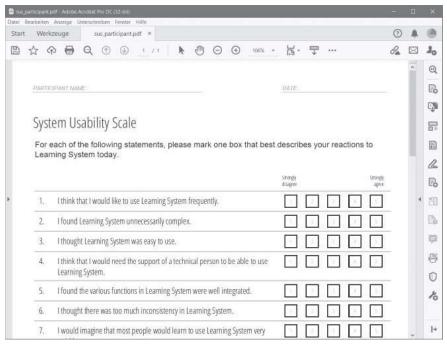


Abb. 21: Ein SUS-Fragebogen

#### **Konkrete Fragen**

In der Praxis funktionieren konkrete Fragen zu einzelnen Aspekten einer Lernanwendung deutlich besser bei der Optimierung der Learning Experience, als allgemein gehaltene Fragebögen wie der SUS. Das Beispiel in Abbildung 22 zeigt einen Ausschnitt eines Befragungsergebnis für ein Kategoriebild eines Lernmanagementsystems. In der Befragung (hier mit nur 6 Probanden) wurden verschiedene Kategoriebilder zur Wahl gestellt und bewertet und schließlich die erfolgreichsten Bilder ausgewählt.

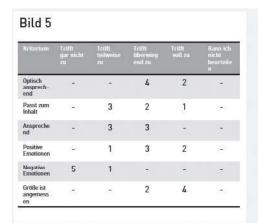




Abb. 22: Befragungsergebnis für ein Kategoriebild eines LMS

Es ist schon schwierig, Nutzer allgemein nach ihrem Usability-Feedback zu fragen und wesentlich weniger aussagekräftig als den Lerner beim Nutzen der Anwendung indirekt (Webanalyse, Reports, a/b-Tests) oder direkt (Nutzertests, Crowdtests) zu beobachten. Noch problematischer wird der Versuch, den Nutzer nach Verbesserungsvorschlägen zu fragen. Auf die Frage, "Was würden sie auf dieser Seite verbessern?", kann ein normaler Lerner nicht vernünftig reagieren, da ihm die Kenntnisse im Web- und Lerndesign fehlen und da er kaum in der Lage ist, alle Aspekte zu berücksichtigen, die hinter der Anwendung stecken.

Ein Beispiel aus einem Nutzertest mit vorangegangener Befragung: In der Befragung wurden alle 10 Probanden gefragt wie wichtig Ihnen die Suche bei einer Lernanwendung ist. Nur 2 Probanden haben sich für besonders wichtig entschieden. Im anschließenden Test, haben 8 Probanden die (schlecht nutzbare) Suche gefunden und eingesetzt. Bei der anschließenden Befragung zu Verbesserungspotentialen wurde die Suche von keinem der Probanden genannt, obwohl 6 von 8 Probanden im Test erkennbare Probleme beim Nutzen der Suche hatten.

## Was nicht funktioniert

#### 4 Fazit

Learner Experience ist messbar, manchmal erfolgt die Messung indirekt und oft steht die Usability als Kern des Lernerlebnisses im Fokus, das ändert aber nichts daran, dass die Messmethoden heute schon vorhanden sind. Warum wird dann bei E-Learning-Anwendungen und -Inhalten vergleichsweise wenig gemessen? Ein Grund ist sicherlich, dass im Gegensatz zum E-Commerce oft die direkte monetäre Auswirkung nicht klar messbar ist. Optimiert man die Customer Experience im E-Commerce, äußert sich das direkt in mehr Verkäufen, während im E-Learning "nur" das Leben des Lerners einfacher wird.

Dankenswerterweise erfolgt hier aber ein Umdenken. Statt immer neuer Inhalte oder mehr Funktionen, wenden sich die E-Learning-Verantwortlichen immer mehr der Frage zu wie sie das Wissen besser und nachhaltiger vermitteln können. Deswegen erscheint es folgerichtig, dass die Messung und Optimierung der Learner Experience in den nächsten Jahren kontinuierlich an Bedeutung gewinnen wird. Grund genug, nicht länger zu zögern und das Messen zu beginnen!

Messen ist möglich!