

Summative Usability Evaluation der online Lernplattform Moodle

ISONORM 9241/110-FRAGEBOGEN

3. Semester Informationsdesign

Wintersemester 2018/2019

Human Computer Interaction Methods

Johannes Ströbele (js349) und Ingo Gorickic (ig035)

Inhaltsverzeichnis

Inhaltsverzeichnis	2
Abbildungsverzeichnis	3
Tabellenverzeichnis	4
Abstract	5
1. Einführung in Fragebogenevaluation	6
1.1. Usability Evaluation und online Lernplattformen.....	6
1.2. Summativen Usability Fragebögen und Moodle.....	9
1.3. Überblick	10
2. Grundlegendes zu Fragebögen	11
2.1. Entstehung von Fragebögen anhand des ISONORM 9241/110.....	11
3. Methode	22
3.1. Zieldefinition	22
3.2. Methodenauswahl.....	23
3.3. Datenerhebung	23
3.4. Datenauswertung- und -interpretation.....	24
4. Ergebnisse	25
4.1. Allgemein.....	26
4.2. Grundsätze der Dialoggestaltung	27
5. Diskussion	32
5.1. Allgemein.....	32
5.2. Grundsätze der Dialoggestaltung	33
6. Fazit	37
6.1. Theoretisches Fazit	37
6.2. Praktisches Fazit	38
Literatur	40

Abbildungsverzeichnis

Abbildung 1: Einordnung in Gestaltungsprozess	12
Abbildung 2: Mittelwerte der 7 Dialogprinzipien	26
Abbildung 3: Summen für die 7 Dialogprinzipien	27
Abbildung 4: Aufgabenangemessenheit	28
Abbildung 5: Selbstbeschreibungsfähigkeit.....	28
Abbildung 6: Erwartungskonformität.....	29
Abbildung 7: Lernförderlichkeit	30
Abbildung 8: Steuerbarkeit.....	30
Abbildung 9: Fehlertoleranz.....	31
Abbildung 10: Individualisierbarkeit	32

Tabellenverzeichnis

Tabelle 1: Leitkriterien der Usability	6
Tabelle 2: Überblick über Evaluationsmethoden.....	8
Tabelle 5: Grundsätze der Dialoggestaltung (DIN EN ISO 9241-110, 2006)	14
Tabelle 3: Hauptgütekriterien	16
Tabelle 4: Übersicht über Fragebogen	18
Tabelle 6: Auswertungstabelle	24
Tabelle 7: Bezeichnungen für Mittelwerte und Summen.....	25

Abstract

Heute nutzen zahlreiche Hochschulen online Lernplattformen. Diese bieten unzählige Funktionen, mit denen Studierende ihre hochschulnahen Aufgaben lösen. So auch die Lernplattform Moodle an der Hochschule der Medien. Einige Artikel zeigen Usability-Probleme auf. Jedoch sind diese nur fragmenthaft oder beziehen sich auf andere Versionen. Daher evaluiert diese Studie die Usability von Moodle im Allgemeinen für die Studierenden der Hochschule der Medien. Zuerst erfolgte die Erklärung von Usability-Evaluationsmethoden, der Fragebogenerstellung und mehrere Fragebogenarten. Darauf basierend wurde der ISONORM 9241/110 ausgewählt und damit die Studie durchgeführt.

Die neutrale Gesamtbeurteilung der Usability zeigt, dass Studierenden Moodle als nützlich, aber ausbaufähig ansehen. Als positiv wurde Aufgabenangemessenheit, Lernförderlichkeit und Erwartungskonformität wahrgenommen. Jedoch die Individualisierbarkeit, Selbstbeschreibungsfähigkeit, Steuerbarkeit und Fehlertoleranz als kritisch. Schließlich sollten in diesen Bereichen konkrete Verbesserungsvorschläge mithilfe von Interviews gesucht werden, um damit die übergreifende Usability von Moodle zu verbessern.

Keywords: Usability, summative Evaluation, Fragebogenerstellung, Fragebogen, ISONORM, E-Learning, internetbasiertes Lernen, Learning Management System, Lernplattform, Moodle

1. Einführung in Fragebogenevaluation

Dieser Abschnitt ist in drei Teile gegliedert. Der erste Teil geht auf die allgemeine Relevanz von Usability für online Lernplattformen ein und legt die Forschungsfrage dar. Danach stellt der Artikel die Bedeutung von summativen Fragebögen für die Evaluation von Moodle dar und gibt eine kurze Einführung über Fragebögen. Der letzte Teil gibt einen kurzen Überblick über den Artikel.

1.1. Usability Evaluation und online Lernplattformen

Heutzutage verwenden viele Schulen und nahezu alle Hochschulen in Deutschland online Lernplattformen (Martin u. a., 2008). Diese bieten Lehrenden und Lernenden zahlreiche Vorteile. Dozenten können schnell und einfach Lernmaterial erstellen, den Lernfortschritt der Lernenden verfolgen und Lernmaterial zur Verfügung stellen. Andererseits können Schulkinder und Studierende jederzeit und überall an Veranstaltungen teilnehmen, mit anderen Nutzern kommunizieren und auf Kursmaterial zugreifen (Kakasevski, Mihajlov, Arsenovski, & Chungurski, 2008; Martin u. a., 2008).

Mehrere Studien zeigen, dass viele Lehrende und Lernende unzufrieden mit Lernplattformen und im speziellen mit Moodle sind (Kakasevski u. a., 2008; Melton, 2006; Zemsky & Massy, 2004). Unzureichende Funktionen, eine schlechte pädagogische Führung, ineffektive Bewertungsprozesse und zahlreiche Usability-Fehler werden als Gründe dafür genannt (Kakasevski u. a., 2008; Melton, 2006; Zemsky & Massy, 2004).

Dabei stellt sich die Frage, was Usability konkret bedeutet. In der DIN EN ISO 9241-11(2015) ist Usability, die übersetzt Gebrauchstauglichkeit genannt wird, folgendermaßen definiert. Usability ist ein „Ausmaß, *in dem ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um festgelegte Ziele effektiv, effizient und zufriedenstellend zu erreichen*“. In diesem Artikel ist das Produkt das interaktive System Moodle. Effektivität, Effizienz und Zufriedenheit sind wie folgt definiert.

Tabelle 1: Leitkriterien der Usability

Leitkriterien	Definition
Effektivität	„Die Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen“ (DIN EN ISO 9241/11, 2015).
Effizienz	„Der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem Benutzer ein bestimmtes Ziel erreichen“ (DIN EN ISO 9241/11, 2015).

Zufriedenheit	„Freiheit von Beeinträchtigung und positive Einstellung gegenüber der Nutzung des Produkts“ (DIN EN ISO 9241/11, 2015).
---------------	---

Im Speziellen weisen Kakasevski, Mihajlov, Arsenovski und Chungurski (2008) auf zahlreiche Probleme für Studierende hin, die durch unzureichende Usability entstehen. 80 % der Studierende haben Schwierigkeiten mit dem online Chat. Weiter finden sich viele in der komplexen Software vor allem zu Beginn nicht zurecht. Zuletzt finden viele Studierende, dass die Startseite nur einen schlechten Überblick über die Software bietet. Zuletzt ist interessant, dass Usability einen positiven Einfluss auf das Lernverhalten haben kann (Melton, 2006; Nokelainen, 2006; Van Nuland & Rogers, 2016).

Die zeigt, dass ein enormes Potenzial in der Verbesserung der online Lernplattform Moodle anhand von Usability-Kriterien erreicht werden könnte. Deshalb ist das Ziel dieses Artikels die Usability von Moodle zu verbessern. Da die Ersteller dieses Artikels an der Hochschule der Medien studieren und alle Studierenden dieser Hochschule Moodle verwenden, soll Moodle für diese untersucht werden. Daher heißt die Forschungsfrage für diesen Artikel:

Wie ist die Usability der online Lernplattform Moodle für die Studierenden der Hochschule der Medien?

Dazu wird nachfolgend die Usability evaluiert, um damit Verbesserungspotential zu identifizieren. Der Begriff Usability Evaluation beschreibt „eine systematische und möglichst objektive Bewertung eines geplanten, laufenden oder abgeschlossenen Projekts“ bezüglich seiner Gebrauchstauglichkeit (Sarodnick & Brau, 2015, S. 23).

Dabei kann die Usability nach zahlreichen Evaluationskriterien bewertet werden. Zum Beispiel Zufriedenheit, Nachhaltigkeit, Selbstbeschreibungsfähigkeit, Fehlertoleranz oder Effektivität (Bösser, 2019; Prümper & Anft, 1993; Sarodnick & Brau, 2015, S. 24). Dieser Artikel wird die Usability anhand der Theorie der Dialogprinzipien evaluieren (DIN EN ISO 9241/110, 2006).

Um die Usability von Software zu evaluieren, gibt es zahlreiche Methoden (Günther Gediga, Hamborg, & Düntsch, 2002; Heinsen, 2003, S. 4; Sarodnick & Brau, 2015, S. 119). Diese können anhand verschiedener Kriterien kategorisiert werden. Zum einen können diese basierend auf der Art der erhobenen Daten eingeteilt werden. Subjektive Daten spiegeln die Meinung der Nutzer durch Selbsteinschätzung wider. Objektive Daten werden aus Beobachtungen oder Messungen gewonnen (Hix & Hartson, 1993; Sarodnick & Brau, 2015, S. 120).

Zum anderen können diese in empirische und analytische Methoden unterschieden werden. Empirische Methoden zielen darauf ab, Information über die Usability durch Beobachtungen und Befragungen von Nutzern zu erhalten. Solche Methoden sind Usability-Tests und Fragebogenverfahren (Sarodnick & Brau, 2015, S. 119–120). Andererseits versetzen sich, bei analytischen Methoden, Usability Experten in die Nutzer und beurteilen die Software anhand von Richtlinien. Heuristische Evaluation und Cognitive Walkthrough sind Beispiele dafür (Sarodnick & Brau, 2015, S. 120).

Methoden können auch basierend auf dem Zeitpunkt der Evaluation eingeteilt werden, siehe *Tabelle 1: Leitkriterien der Usability*. Formative Methoden werden verwendet, um die Usability schon während des Entwicklungsprozesses zu verbessern. Bei diesen Methoden steht die Suche nach konkreten Verbesserungsmöglichkeiten der Usability im Vordergrund (Sarodnick & Brau, 2015, S. 120, 201).

Tabelle 2: Überblick über Evaluationsmethoden

Summative Evaluationsmethode	Formative Evaluationsmethode
Fragebögen	Fragebögen
Usability Tests	Usability Tests
Expertenleitfäden	Cognitive-Walkthrough-Verfahren

Bekannte Beispiele sind Card Sorting, Usability Test, Cognitive-Walkthrough-Verfahren und Fragebögen. Mit diesen Verfahren werden meist qualitative Daten gewonnen. Dies sind meist schriftliche oder audio-visuelle Informationen über die subjektive Sicht der Teilnehmer, wie ihre inneren Zustände. Die Offenheit des Vorgehens ermöglicht es, oft neue, bisher unbekannte Sachverhalte zu entdecken. (Sarodnick & Brau, 2015, S. 120, 201).

Andererseits werden summative Evaluationsmethoden eingesetzt, um die Usability von fertigen Systemen zu überprüfen. Im Gegensatz zu formativen Methoden, zielen diese nicht auf die Identifikation von praktischen Verbesserungsmöglichkeiten, sondern auf eine globale Bewertung des Systems ab. Fragebögen, Usability Tests und Expertenleitfäden sind Beispiele für summative Evaluation. Aus diesen können numerischen Daten gewonnen werden. Diese dienen zumeist, um Hypothesen aus qualitative Arbeiten statistisch zu überprüfen (Sarodnick & Brau, 2015, S. 120, 201).

Fraglich ist nun, mit welcher Methode man die Usability von Moodle auswertet. Da es sich bei der Plattform um ein abgeschlossenes Produkt handelt, empfehlen sich

hierbei summative Evaluationsmethoden (Günther Gediga u. a., 2002, S. 3; Sarodnick & Brau, 2015, S. 120, 201).

1.2. Summativen Usability Fragebögen und Moodle

Nun stellt sich die Frage, welche der summative Evaluationsmethode sich für eine globale Untersuchung der Lernplattform eignet. Bei summativen Usability Tests wird die Nutzerinteraktion bei der Lösung von Aufgaben aufgezeichnet. Diese Methode eignet sich konkrete Probleme von realen Nutzern zu erfassen. Hingegen zielen Fragebögen auf eine globale Bewertung ab und decken Problembereichen auf. Bei Expertenleitfäden wird der prozentuale Erfüllungsgrad der Gestaltungsrichtlinien erfasst. Zum Beispiel die Normen Gerechtigkeit des Systems anhand der DIN EN ISO 9241/110 (DIN EN ISO 9241/110, 2006; Sarodnick & Brau, 2015, S. 201).

Das Ziel dieser Studie ist eine allgemeine Bewertung und Identifikation von Problembereichen in der Usability dieser Lernplattform. Deshalb wird Moodle nachfolgend mit einem summativen Fragebogen evaluiert. Ein Fragebogen ist eine Sammlung von Items, welche Fragen oder Aussagen sein können (Sarodnick & Brau, 2015, S. 120).

Younes, Stewart und Kyriakidou (2013) geben an, dass ein Fragebogen eine Methode ist, mit der Menschen anhand des gleichen Fragenkatalogs in einer vorher festgelegten Weise befragt werden (Breakwell, Smith, & Wright, 2012). Ein anderer Artikel (Kabungaidze, Mahlatshana, & Ngirande, 2013) zitiert, dass ein Fragebogen ein Dokument ist, dass Items enthält, um Information für eine Untersuchung zu erwerben (Babbie & Mouton, 2001).

Insgesamt wurde festgestellt, dass sich Fragebögen dazu eignen, Systeme zu bewerten oder mehrere Systemversionen miteinander zu vergleichen. Dabei gibt der Nutzer subjektive Urteile aus seiner persönlichen Erfahrung über das System ab. Dies sind meist Selbstauskünfte über innere Zustände und Eindrücke während der Nutzung des jeweiligen Systems. Dabei relativiert sich die subjektive Bewertung, je mehr Teilnehmer den Fragebogen beantworten (Heinsen, 2003, S. 172; Sarodnick & Brau, 2015, S. 181–182).

Das Ziel von Usability-Fragebögen ist dabei die Untersuchung, ob ein interaktives System bestimmten Kriterien der Usability entspricht (Babbie & Mouton, 2001). Zum Beispiel wird untersucht, ob Moodle die Dialogprinzipien der DIN EN ISO 9241 Teil 110 erfüllt (DIN EN ISO 9241/110, 2006). Dabei sind die Items für die einzelnen Prinzipien in sogenannten Subskalen zusammengefasst.

Auch unterscheiden sich die Fragebögen in dem Grad ihrer Standardisierung. Bei voll standardisierten Fragebögen sind Item Formulierungen, Darbietungsreihenfolge und Antwortformat festgelegt. Die Teilnehmer drücken dabei ihre Meinung durch

Einschätzungs- oder Ratingskalen aus. Bei teilstandardisierten Fragebögen sind die Item Formulierungen und die Darbietungsreihenfolge gleich, jedoch können Antworten frei formuliert werden. Dies geschieht häufig in Form von frei textlichen Anmerkungen am Ende einer Subskala des Fragebogens. Zuletzt nicht standardisierte Tests, bei denen die Fragen gestaltbar sind und der Ablauf flexibel festgelegt wird (Heinsen, 2003, S. 172).

Evaluationsfragebögen haben mehrere Vorteile und Nachteile (Heinsen, 2003, S. 183–184; Kabungaidze u. a., 2013; Sarodnick & Brau, 2015, S. 184–185). Positiv ist die hohe Objektivität, da der Einfluss durch Dritte in der Regel niedrig ist. Durch die vorherige Formulierung der Items, kann gewährleistet werden, dass die Teilnehmer nicht durch die Fragestellung beeinflusst werden. Weiter können die Teilnehmer online Fragebögen selbständig, sowie unabhängig von Zeit und Ort ausfüllen. Dadurch kann in kurzer Zeit eine große Stichprobe gezogen werden. Dies ermöglicht empirische Untersuchungen mit dieser Methode.

Andererseits hängt die Beantwortung stark von der Selbstwahrnehmung, der Erinnerung und der Aufmerksamkeit der Teilnehmer ab. Dadurch kommt es zu Verfälschungen und Fehlern. Weiter sind ungewissenhafte Antworten möglich, die zum Beispiel aufgrund von Zeitdruck entstehen. Zuletzt können Antworten durch eine unterschiedliche Interpretation der Fragen verfälscht werden (Heinsen, 2003, S. 183–184; Kabungaidze u. a., 2013; Sarodnick & Brau, 2015, S. 184–185).

1.3. Überblick

Nun wird der Aufbau des Artikels kurz dargestellt. Im Kapitel 2. *Grundlegendes zur Evaluation mit Fragebögen* wird der Fragebogen-Konstruktionsprozess anhand des ISONORM 9241/110 und verschiedene summative Fragebogenarten vorgestellt.

Danach wird in 3. *Methode* das Verfahren zur Beantwortung der Forschungsfrage erklärt. Dafür wird zuerst das Ziel der Untersuchung konkretisiert und die Auswahl des ISONORM 9241/110 erklärt. Anschließend wird das Vorgehen bei der Untersuchungsplanung, der Datenerhebung und der Interpretation dargestellt.

Das Kapitel 4. *Ergebnisse* stellt die allgemeinen Ergebnisse und dann für jedes der sieben Dialogprinzipien vor. Dabei ist die Gesamtbeurteilung von Moodle neutral. Aufgabenangemessenheit, Lernförderlichkeit und Erwartungskonformität wurden eher positiv wahrgenommen. Hingegen die Individualisierbarkeit, Steuerbarkeit Selbstbeschreibungsfähigkeit und Fehlertoleranz als eher negativ.

Die möglichen Gründe für die Ergebnisse sind in 5. *Diskussion* dargestellt. Dafür wurde eine breite Recherche, die Erfahrungen der Verfasser dieser Studie und eine Diskussion mit Studierenden der Hochschule der Medien verwendet. Zuletzt erfolgt in 6. *Fazit* die Schlussfolgerung für die Theorie und Praxis basierend auf der Studie.

Eine Gewichtung der Prinzipien hätte zu einer genaueren Bewertung führen können. Insgesamt sind die Verfasser jedoch mit der Studie zufrieden.

2. Grundlegendes zu Fragebögen

2.1. Entstehung von Fragebögen anhand des ISONORM 9241/110

Zahlreiche wissenschaftliche Artikel und Bücher beschäftigen sich mit dem Prozess zur Fragebogenkonstruktion (Bühner, 2011, S. 201; Burmester, Görner, Vossen, Zolleis, & Zouboulides, 1997; Jonkisz, Moosbrugger, & Brandt, 2012)

(Bühner, 2011; Burmester u. a., 1997). Zum Beispiel schlägt Bühner (2011) vor, zuerst einen Entwurf des Fragebogens zu erstellen, dann diesen empirisch zu überprüfen und abschließend einer Normierung beziehungsweise Cut-Off-Ermittlung durchzuführen. Jedoch berücksichtigt dieser beispielsweise nicht die Festlegung des Forschungsziels und der Zielgruppe.

Da durch die Recherche kein ganzheitlicher Prozess zur Fragebogenerstellung auffindbar war, schlägt dieser Artikel nachfolgend einen neuen Ablauf vor. Die Reihenfolge und der Inhalt der Abschnitte wurde basierend auf den zuvor genannten Artikeln (Bühner, 2011; Burmester u. a., 1997). Dabei wird jeder Abschnitt zuerst allgemein für Fragebögen und dann beispielhaft am ISONORM 9241/10 erklärt (Prümper, 1997).

- 1) Evaluationsziel
- 2) Relevanz Fragebogenerstellung
- 3) Relevanzüberprüfung Modellerstellung
- 4) Fragebogenentwurf
- 5) Empirische Evaluation
- 6) Prüflaufplan

2.1.1. Evaluationsziel

Bevor man sich Gedanken über die Erstellung eines Fragebogens macht, sollte man sich das Evaluationsziel konkret definieren (Bortz & Döring, 2006, S. 269; Sarodnick & Brau, 2015, S. 182–183). Ohne eine Zieldefinition, kann zum einen nicht gewissenhaft entschieden werden, was genau für ein Fragebogen erstellt werden sollte. Zum anderen ob ein bereits existierender Fragebogen zur Beantwortung der Forschungsfrage ausreicht. Deshalb wird nun kurz darauf eingegangen, welche Ziele vorab festgesetzt werden müssen.

Als erstes sollte geklärt werden, welcher Forschungsgegenstand untersucht werden sollte. Im Falle des ISONORM 9241/10 ist dies eine Software (Prümper & Anft, 1993). Anschließend muss festgelegt werden, wonach dieser

Forschungsgegenstand untersucht werden soll. Für den ISONORM ist dies eine Evaluation der Software nach den Grundsätzen der Dialoggestaltung (DIN EN ISO 9241/10, 1996).

Dabei wurde der Fragebogen für die Usability-Evaluation während der Evaluationsphase im menschenzentrierten Gestaltungsprozess entwickelt (DIN EN ISO 9241/210, 2011). Somit ist nur eine Untersuchung von einer abgeschlossenen beziehungsweise funktionsfähigen Software möglich. Die Einordnung ist auf *Abbildung 123: Einordnung in Gestaltungsprozess* zu sehen.

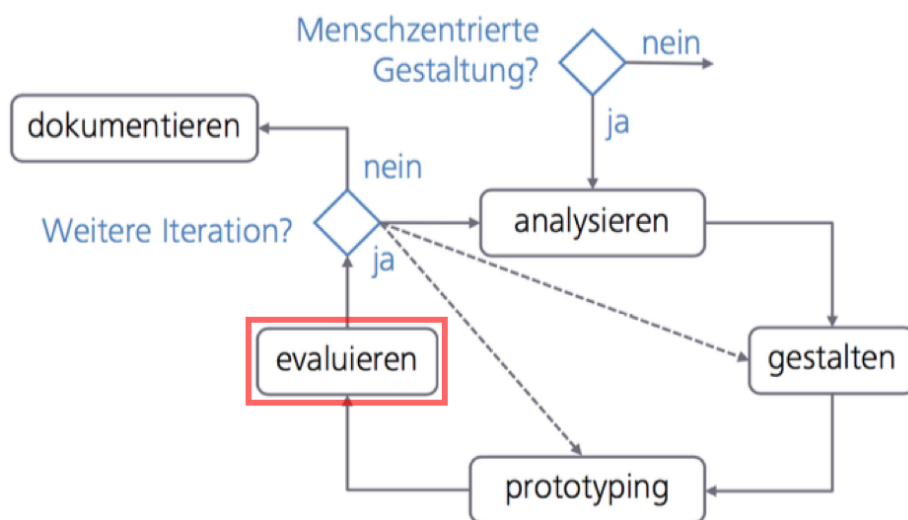


Abbildung 1: Einordnung in Gestaltungsprozess

Weiter muss die Zielgruppe festgelegt werden. Der ISONORM richtet sich an alle Nutzer von Software (Prümper & Anft, 1993). Zuletzt sollten diese Ziele in ein Forschungsziel zusammengefasst werden. Dieses wurde für den ISONORM nur indirekt formuliert. Das Forschungsziel könnte wie folgt lauten: *Inwiefern erfüllt die Software die Grundsätze der Dialoggestaltung für die Nutzer?*

2.1.2. Relevanz Fragebogenerstellung

Auf Basis des Evaluationsziel kann nun überprüft werden, ob ein neuer Fragebogen erstellt werden soll oder ein bereits existierender verwendet werden kann. Zunächst liegt es nahe „einfach irgendeinen passenderscheinenden Fragebogen aus dem Internet zu verwenden“. Viele Forscher zeigen, dass solch ein Vorgehen in der Praxis üblich ist (Kallus, 2016, S. 5; Sarodnick & Brau, 2015, S. 182–183; Vredenburg, Mao, Smith, & Carey, 2002). Jedoch dabei meist wissenschaftliche Gütekriterien, wie Validität, Reliabilität und Objektivität, nicht einhalten (Gray & Salzman, 1998; Molich, Ede, Kaasgaard, & Karyukin, 2004; Molich u. a., 1999).

Andererseits könnte man „schnell selbst einen Fragebogen ohne Leitfaden erstellen“. Entgehen solch einer Vorstellung, erfordert dies jedoch viel Zeit und eine stark systematische Herangehensweise. Nur so kann die Güte gewährleistet werden und somit zuverlässigen Daten gewonnen und Aussagen über die Forschungsfrage getroffen werden (Heinsen, 2003, S. 173; Sarodnick & Brau, 2015, S. 182).

Um diesem Aufwand aus dem Weg zu entgehen, kann ein passender Fragebogen gesucht werden. Dabei sollte auch nach Artikel recherchiert werden, die die Erfüllung der Güte bestätigen. Im Zweifelsfall kann der Suchende selbst den Fragebogen auf seine Güte überprüfen. Falls Anpassungen notwendig sind, sollten diese unbedingt mit dem Ersteller des Fragebogens abgeklärt werden, da diese die Güte beeinträchtigen könnten (Bortz & Döring, 2006, S. 269; Sarodnick & Brau, 2015, S. 182–183).

Falls kein passender Fragebogen existieren würde, könnte ein neuer konstruiert werden. Der ISONORM 9241/10 wurde entwickelt, weil es zu seiner Veröffentlichung, 1993, keinen Fragebogen gab, der Software nach ihrer Usability evaluierte und gleichzeitig wissenschaftlichen Gütekriterien gerecht wurde. Vor 1993 nutzten deshalb viele Firmen deshalb rein wissenschaftliche Fragebögen (Kallus, 2016, S. 5; Sarodnick & Brau, 2015, S. 182–183; Vredenburg u. a., 2002). Diese waren praxisfremden und verfehlten das Evaluationsziel der Firmen. Auch war deren Durchführung kostenintensiv und zeitaufwändig (Vredenburg u. a., 2002).

Die “Discount Usability Methods” von Nielsen (1989) waren eine der ersten Methoden, die sowohl wissenschaftlichen als auch praktischen Anforderungen genügten. Ab diesem Zeitpunkt begannen Forscher weiterer solcher Evaluationsmethoden zu entwickeln. Einer dieser Fragebogen war die Langfassung des ISONORM 9241/10 von Prümper und Anft (1993).

Damit dieser den praktischen Ansprüchen genügt, wurde dieser mit Experten direkt aus der Praxis entwickelt (Prümper & Anft, 1993). Weiter erfolgte einer umfangreiche wissenschaftlichen Überprüfung der Objektivität, Reliabilität und Validität, auf die im *Abschnitt 2.1.6* näher eingegangen wird (Figl, 2010; Prümper, 1997, 1999). Dies gewährleistet eine hohe Aussagekraft des Fragebogens und damit der zu untersuchenden Forschungsfrage.

2.1.3. Relevanz Modellerstellung

Für die Erstellung eines neuen Fragebogens, benötigt man ein qualitatives Modell beziehungsweise Konstrukt. Nur basierend auf den einzelnen Faktoren dieses Modells, können valide Items nachfolgend formuliert werden. Daher sollte in der Literatur nach solch einem Konstrukt gesucht werden.

Falls keines passenden qualitativen Modells bereits existiert, kann ein neues in einer separaten qualitativen Studie erstellt werden. Dabei erfolgt die Erstellung des Modells meistens in folgenden Schritten. Zuerst muss das Forschungsziel für die qualitative Studie definiert werden.

Danach muss mithilfe qualitativer Methode, wie zum Beispiel Interviews, möglichst viele, unterschiedliche Faktoren identifiziert werden. Diese werden anschließend aussortiert und kategorisiert. Diese kategorisierten Faktoren bilden dann das qualitative Modell. Basierend auf diesen Faktoren werden zumeist auch konkrete Hypothese erstellt, die später durch den Fragebogen empirisch evaluiert werden.

Im Falle des ISO NORM 9241/10 wurde ein existierendes Modell verwendet. Dieses zielt darauf ab, Software ergonomisch zu gestalten. Dafür wurde das qualitative Modell Grundsätze der Dialoggestaltung mithilfe einer Studie erstellt (DIN EN ISO 9241/10, 1996). Die kategorisierten Faktoren sind dabei die einzelnen Grundsätze beziehungsweise Prinzipien der Dialoggestaltung. Diese sind mit einer Erklärung in der *Tabelle 5: Grundsätze der Dialoggestaltung* abgebildet.

Tabelle 3: Grundsätze der Dialoggestaltung (DIN EN ISO 9241-110, 2006)

Dialogprinzip	Definition
Aufgabenangemessenheit	„Ein Dialog ist aufgabenangemessen, wenn er den Benutzer unterstützt, seine Arbeitsaufgabe effektiv und effizient zu erledigen.“
Selbstbeschreibungsfähigkeit	„Ein Dialog ist selbstbeschreibungsfähig, wenn jeder einzelne Dialogschritt durch Rückmeldung des Dialogsystems unmittelbar verständlich ist oder dem Nutzer auf Anfrage erklärt wird.“
Erwartungskonformität	„Ein Dialog ist erwartungskonform, wenn er konsistent ist und den Merkmalen des Benutzers entspricht, z.B. seinen Kenntnissen aus dem Arbeitsgebiet, seiner Ausbildung und seiner Erfahrung sowie den allgemein anerkannten Konventionen.“
Lernförderlichkeit	„Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen des Dialogsystems unterstützt und anleitet.“
Steuerbarkeit	„Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist.“
Fehlertoleranz	„Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben

	entweder mit keinem oder minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann.“
Individualisierbarkeit	„Ein Dialog ist individualisierbar, wenn das Dialogsystem Anpassungen an die Erfordernisse der Arbeitsaufgabe sowie an die individuellen Fähigkeiten und Vorlieben des Benutzers zulässt.“

2.1.4. Fragebogenentwurf

Basierend auf den Faktoren des qualitativen Modells, sollten nun im ersten Schritt die Items formuliert werden. Diese messen somit die einzelnen Faktoren. Weiter sollten diese auf sprachliche Konsistenz, Einfachheit und Verständlichkeit überprüft werden. Die Anzahl und deren Reihenfolge muss optimiert werden. Zum Beispiel sollte jeder Faktore mit der gleichen Anzahl von Items gemessen werden (Burmester u. a., 1997; Kallus, 2016).

Darüber hinaus sollte eine passende Antwortart gewählt werden. Zum einen eine passende Skala. Zum Beispiel eine Nominalskala für das Geschlecht. Zum anderen ob die Items oder Frage offen oder geschlossen beantwortbar sind. Bekannte offene Aussagen sind Kommentar und Feedback. Geschlossene Fragen können Ratingskala oder Richtig-Falsch-Antworten sein (Bühner, 2011, S. 201; Jonkisz u. a., 2012).

Danach sollten Instruktionen, Ankern und Vorgabemodalitäten festgelegt werden. Also Hilfstexte für schwierig verständliche Items oder Hinweise, wie ein Fragebogen beantwortet werden sollte. Zuletzt sollten Items formuliert werden, um zu überprüfen, ob die Stichprobe der Population ähnelt. Falls dies nicht zutrifft, können gewonnen Ergebnisse stark verzerrt werden (Bühner, 2011, S. 203; Jonkisz u. a., 2012; Kallus, 2016, S. 13).

Für die Langfassung des ISONORM 9241/10 wurden jeweils fünf Items pro Dialogprinzipien formuliert (DIN EN ISO 9241/10, 1996). Somit umfasste der finalisierte Fragebogen 35 Items. Dabei wurde darauf geachtet, dass die Faktoren sehr gut abgebildet werden und trotzdem verständlich sind (Prümper & Anft, 1993). Weiter wurde ein sieben-stufiges, bipolares Frageformat für die Beantwortung gewählt, welches immer eine positive und negative Aussage gegenüberstellt.

2.1.5. Empirische Evaluation

Nun wird auf die Untersuchung der Güte von Fragebögen eingegangen. Nach Bortz und Döring (Bortz & Döring, 2006, S. 212; Bühner, 2011, S. 33–43) wird die Qualität von Fragebögen anhand Haupt- und Nebengütekriterien gemessen. Validität,

Reliabilität und Objektivität sind die Hauptgütekriterien, welche in *Tabelle 3: Hauptgütekriterien* erläutert sind.

Tabelle 4: Hauptgütekriterien

Kriterium	Erklärung
Objektivität	Die Ergebnisse der Methode muss unabhängig von den Personen sein, die das Verfahren durchführen, auswerten und interpretieren (Endmann, Ackermann, Thielsch, & Hess, 2014).
Reliabilität	Das Verfahren muss das Konstrukt genau messen (Endmann u. a., 2014).
Validität	Die Methode muss das Konstrukt messen, welches es vorgibt zu untersuchen (Endmann u. a., 2014).

Die Objektivität fordert, dass die Methode unabhängig von den Personen ist, die das Verfahren durchführen, auswerten und interpretieren. Dies ist zum Beispiel durch die Berücksichtigung von Richtlinien bei der Fragebogenerstellung, Auswertung und Interpretation möglich (Bühner, 2011, S. 33–43; Endmann u. a., 2014; Heinsen, 2003, S. 173; Sarodnick & Brau, 2015, S. 183). Die Reliabilität stellt die Messgenauigkeit der Methode sicher. Oft wird dies durch die Messung der Korrelation der Items untereinander oder die Korrelation der Daten an verschiedenen Messzeitpunkten herausgefunden (Endmann u. a., 2014).

Für die Validität muss die Methode messen, was sie messen soll. Hier wird also nicht die Genauigkeit, sondern speziell das Konstrukt überprüft. Eine Möglichkeit die Validität zu testen besteht darin die Ergebnisse von verschiedenen Verfahren vergleichen. Wenn beide Verfahren auf dem gleichen Konstrukt basieren, sollten diese auch zu ähnlichen Ergebnissen kommen. Weiter ist eine Evaluation durch Experten dieses Gebietes möglich oder eine Überprüfung der angenommenen Faktoren mittels einer konfirmatorischen Faktorenanalyse (Endmann u. a., 2014).

Mehrere Studien aufzeigen, dass die Validität, Reliabilität und Objekt für den ISONORM 9241/10 erfüllt sind (Figl & Mensch, 2009; Prümper, 1997, 1999; Prümper & Anft, 1993). Die Güte der Objektivität wurde durch die Berücksichtigung von Richtlinien bei der Fragebogenerstellung zur Minimierung von suggestiven Aussagen erfüllt. Auch wurden Auswertungsvorschriften vorgegeben. Die weitere Objektivität ist rein von der Beeinflussung durch die Studienleiter, welche den ISONORM nutzen, möglich.

Darüber hinaus wurde die Reliabilität sichergestellt (Figl & Mensch, 2009; Prümper, 1999). Dabei wurde die Korrelation der Daten, welche an verschiedenen

Messzeitpunkten gewonnen wurden, verglichen. Diese wiesen bei beiden Studien eine große Ähnlichkeit auf, wodurch von einer hohen Messgenauigkeit ausgegangen werden kann.

Weiter wurde durch Experten überprüft, dass der Fragebogen das misst, was er missen soll. Nämlich die Gebrauchstauglichkeit basierend auf den Grundsätzen für die Dialoggestaltung (Figl & Mensch, 2009; Prümper, 1997, 1999). Darüber hinaus kommen der ISONORM und der IsoMetrics, die auf dem gleichen Konstrukt basieren, zu ähnlichen Ergebnissen, wie Figl und Mensch zeigen (Figl & Mensch, 2009). Somit kann auch von einer hohen Validität ausgegangen werden.

Allgemein Nebengütekriterien

Andererseits werden Normierung, Vergleichbarkeit, Ökonomie, Zumutbarkeit, Unverfälschbarkeit und Nützlichkeit als Nebengütekriterien anführt (Bortz & Döring, 2006, S. 212; Bühner, 2011, S. 43). Die Nominierung besagt, dass Untersuchungsergebnisse nur im Vergleich mit einer Referenzstichprobe interpretierbar sind. Zum Beispiel besagt ein IQ von 17 Punkten ohne weitere Informationen nicht aus, ob die Leistung der Person niedrig oder hoch ist.

Erst der Vergleich mit der Referenzstichprobe ermöglicht eine Interpretation des Ergebnisses. Bei der Vergleichbarkeit wird der Fragebogen mit anderen inhaltsähnlichen Verfahren verglichen. Oft wird dieses Kriterium schon während der Überprüfung der Reliabilität und Validität abgehandelt (Bortz & Döring, 2006, S. 212; Bühner, 2011, S. 43).

Das Kriterium der Ökonomie impliziert, dass ein Fragebogen zu hohen Wissensgewinn führen und gleichzeitig kosten- sowie zeitgünstig sein sollte. Eine akzeptable psychische und körperliche Belastung des Teilnehmers während des Fragebogenverfahren, beschreibt die Zumutbarkeit.

Das Kriterium der Unverfälschbarkeit besagt, dass Nutzende die Testwerte nicht gezielt manipulieren können. Zum Beispiel sollte mit Kontrollfragen vermieden werden, dass Teilnehmer sozial erwünschte Antworten geben. Zuletzt sagt die Güte der Nützlichkeit aus, dass die Durchführung des Fragebogens eine praktische Relevanz hat, also nützlich ist (Bortz & Döring, 2006, S. 212; Bühner, 2011, S. 43).

Nur teilweise konnte unsere Recherche die Einhaltung dieser Nebengütekriterien für den ISONORM 9241/10 beweisen. Dieser Fragebogen erfüllt ökonomische Ansprüche, da er kosten- und zeitgünstig ist (Prümper & Anft, 1993). Weiter zeigt dieser Artikel auch, dass dieser in enger Kooperation mit Experten aus der Praxis entwickelt wurde und somit nützlich ist.

2.1.6. Prüfleitfadens

Ein Prüfleitfaden ist optional, erhöht aber die Objektivität des Fragebogenverfahrens (Bortz & Döring, 2006, S. 212; Bühner, 2011, S. 43; Burmester u. a., 1997). Dabei sollte aufgezeigt werden, dass zuerst die Evaluationsziel festgelegt werden müssen. Dann muss dargelegt werden, wie und unter welchen Bedingungen die Daten erhoben werden sollen. Weiter müssen Richtlinien für die Auswertung der Daten vorgegeben werden, um weiter die Objektivität des Verfahrens zu gewährleisten. Zuletzt sollte auch ein Leitfaden für die Interpretation vorgegeben werden, damit die Prüfer die gewonnen Daten nahezu identisch interpretieren und sich nicht widersprechen.

Dieser Prüfleitfaden wurde nicht für die Primärliteratur des ISONORM 9241/10 erstellt (Prümper & Anft, 1993). Jedoch existiert ein Handbuch, welches genau die einzelnen Schritte beschreibt, um das Fragebogenverfahren für den ISONORM 9241/10 ohne hohen Zeitaufwand und gleichzeitig ordentlich durchzuführen (Burmester u. a., 1997)

Heutzutage existieren unzählige summative Fragebögen zur Usability-Evaluation. Zum Beispiel der ISONORM 9241/110 (Prümper, 2008), der IsoMetrics (Günther Gediga & Hamborg, 1999), der SUMI (Bösser, 2019), der QUIS (Chin, Diehl, & Norman, 1988), der SUS (Brooke, 1996) und der PSSUQ (Lewis, 1995). Fünf dieser Fragebögen werden nachfolgend erläutert. Dabei unterscheiden sich diese nicht nur im Aufbau, sondern auch im unterliegenden Modell und Einsatzzweck. Die Eigenschaften sind in *Tabelle 4: Übersicht über Fragebogen* kurz zusammengefasst.

Tabelle 5: Übersicht über Fragebogen

Fragebogen	Modell	Eigenschaften	Einsatzzwecke
ISONORM 9241/110 (Prümper, 2008)	7 Dialogprinzipien (DIN EN ISO 9241/110, 2006)	21 Items (7 Konstrukte)	Evaluation komplexer Software
IsoMetrics (Günther Gediga, Hamborg, & Düntsch, 1999)	7 Dialogprinzipien (DIN EN ISO 9241/110, 2006)	75 Items (7 Konstrukte)	Evaluation komplexer Software
SUMI (Bösser, 2019)	1) Effizienz, Affekt, Hilfe & Unterstützung, Kontrollierbarkeit und Erlernbarkeit 2) Gebrauchstauglichkeit	1) 50 Items (5 Konstrukte) 2) 25 Items (1 Konstrukt)	Software-Evaluation

QUIS (Chin u. a., 1988)	Subjektive Zufriedenheit	90 Items (20 Konstrukte)	Software-Evaluation
SUS Brooke, 1996)	Usability	10 Items (1 Konstrukt)	Einfache Software-Evaluation

2.1.7. ISONORM

Der ISONORM testet, inwiefern Systeme die Grundsätze der Dialoggestaltung erfüllen (Prümper, 2008; Prümper & Anft, 1993). Diese sieben Grundsätze beziehungsweise Dialogprinzipien evaluieren tendenziell die Usability. Insgesamt existieren vier Versionen. Die ISONORM 9241/10 Langfassung (Prümper, 1997) und Kurzfassung (Pataki, Sachse, Prümper, & Thüning, 2006). Sowie die ISONORM 9241/110 Langfassung (Prümper, 2008) und Kurzfassung (Prümper, 2008).

Die älteren Fragebögen des ISONORM 9241/10 stützt sich dabei auf die DIN EN ISO 9241/10 (1996). Die aktualisierte Fragebögen des ISONORM 9241/110 basieren auf dem DIN EN ISO 9241/110 (2006).

Die Kurzfassung des ISONORM 9241/110 wird eher verwendet, um abgeschlossene Softwaresystemen zu bewerten (Prümper, 2008). Die Usability-Evaluation in der Entwicklungsphase oder für Prototypen ist mithilfe der Langfassung möglich (Endmann u. a., 2014; Prümper, 2008). Mehrere Forscher weisen darauf hin, dass sich dieser Fragebogen nicht nur für die Evaluation von Business-Software, Mobile Apps und Webseiten, sondern auch Hardware eignet (Endmann u. a., 2014).

Die hohe Aussagekraft des Fragebogens wurde, wie schon zuvor erklärt, durch die mehrmalige Analyse der wissenschaftlichen Güte bewiesen (Figl, 2010; Prümper, 1997, 1999). Die Langfassung operationalisiert diese Grundsätze mit 35 Items sowie der Möglichkeit zur Gewichtung der Antworten. Jedes der sieben Dialogprinzipien wird dabei mit jeweils fünf Items abgefragt. Die Kurzfassung besteht aus 21 Items, bietet aber keine Möglichkeit zur Gewichtung. In dieser Version wird jedes Dialogprinzip mit drei Items operationalisiert (Prümper & Anft, 1993).

Anhand der siebenstufigen Rating Skala können Teilnehmer wählen, wie sie der gestellten Aussage gegenüberstehen. Diese deckt die Werte von negative -3 bis positive +3 ab. Um eine zielführende Erhebung zu gewährleisten, wird vom Teilnehmer viel Erfahrung vorausgesetzt, da einige Items nur nach langen Auseinandersetzung mit der Software zu bewerten sind (Prümper & Anft, 1993). Die Kurzfassung des ISONORM 9241/110 ist in diesem Artikel zu finden (Prümper, 2008).

2.1.8. IsoMetrics

Die IsoMetrics Fragebögen operationalisieren, ebenso wie ISONORM, die Grundsätze der Dialogprinzipien (DIN EN ISO 9241/10, 1996; Günther Gediga u. a., 1999). Die Usability wird dabei mit 75 Items gemessen. Die Fragebögen können, ähnlich wie bei den ISONORM-Fragebögen, für die Evaluation von komplexerer Software und E-Commerce Webseiten verwendet werden (G. Gediga, Hamborg, & Willumeit, 1998; Günther Gediga u. a., 1999).

Die summative und formative Version unterscheidet sich nicht in der Anzahl der Items. Der formative Fragebogen lässt eine Priorisierung der Items zu und bietet die Möglichkeit zusätzlich Anmerkungen einzufügen. Demnach können durch die Langversion nicht nur Problembereiche, sondern auch konkrete Möglichkeiten zur Optimierung der Usability identifiziert werden (Günther Gediga u. a., 1999).

Der IsoMetrics hat insgesamt eine hohe Güte. Durch die gründliche Erstellung des Fragebogens ohne Suggestivfragen wurde die Objektivität des Fragebogens gewährleistet. Darüber hinaus kann diese jedoch durch die Personen, die die Studie durchführen, negativ beeinflusst werden. Die Reliabilität und Validität wurde durch Untersuchungen in verschiedenen Studien bewiesen (Figl & Mensch, 2009; G. Gediga u. a., 1998; Günther Gediga u. a., 1999).

Die Stärken von IsoMetrics liegen daran, dass er sowohl summativ, als auch formativ genutzt werden kann. Somit eignet er sich für eine Evaluation zu jedem Zeitpunkt des Gestaltungsprozesses. Auch ist dieser in Deutsch und Englisch verfügbar. Der Nachteil des Fragebogens liegt jedoch in seiner Länge. Die Auswertung und die Beantwortung der 75 Items und der Anmerkungen sowie Priorisierung in der Langfassung erfordern viel Zeit (Günther Gediga u. a., 1999).

Zuletzt ist noch positiv hervorzuheben, dass IsoMetrics kostenlos ist. In diesem Artikel ist der Fragebogen zu finden (G. Gediga & Hamborg, 2019). Ein ausführliches Manual befindet sich in dieser Forschungsarbeit (G. Gediga u. a., 1998).

2.1.9. SUMI

SUMI steht für Software Usability Measurement Inventory und misst die Nutzungsqualität von Software aus Sicht des Benutzers. Dieser Fragebogen basiert auf den Konstrukten Effizienz, Affekt, Hilfe und Unterstützung, Kontrollierbarkeit und Erlernbarkeit. Anhand dieser sind die 50 Items in fünf Subskalen unterteilt (Bösner, 2019; Kirakowski & Corbett, 1993).

Darüber hinaus misst der Fragebogen mithilfe einer übergreifenden Skala das Konstrukt Gebrauchstauglichkeit. Diese umfasst 25 Items. Die Nutzer beantworten

die Items auf einer 7-stufigen Ratingskala von „stimme zu“, „weiß nicht“ bis „stimme nicht zu“ (Bösser, 2019; Kirakowski & Corbett, 1993).

SUMI eignet sich für die Bewertung von Produkten während der Entwicklung, für den Vergleich mehrerer Produkte und für die Formulierung von Gestaltungszielen für die Weiterentwicklung eines Produkts. Auch hier wurde die Objektivität durch eine sorgfältige Erstellung sichergestellt. Die Validität sowie Reliabilität erfolgte durch zahlreiche Tests von unabhängigen Studien (Bösser, 2019; Kirakowski, 1994; Kirakowski & Corbett, 1993).

Zuletzt ist positiv, dass dieser auf insgesamt in 20 Sprachen verfügbar ist. Zum Beispiel auf Deutsch, Englisch und Chinesisch (Bösser, 2019; Kirakowski & Corbett, 1993). Der Fragebogen ist für akademische Zwecke kostenlos zu erhalten. Für alle anderen Ziele belaufen sich die Kosten auf 200 Euro.

2.1.10. QUIS

Der Questionnaire for User Interface Satisfaction, auch bekannt als QUIS, misst die subjektive Zufriedenheit der Nutzer mit der Schnittstelle eines Systems (Chin u. a., 1988). Dieser besteht aus 90 Items. Die allgemeine Bewertung des Systems wird anhand von 5 Items gemessen. Die restlichen 85 operationalisieren die 20 Konstrukte. Dabei existiert für jedes Konstrukt eine Hauptfrage und mehrere Unterfragen (Sarodnick & Brau, 2015, S. 190–191).

Die Items werden mit einer Zehn-Stufen-Skala zwischen Null und Neun abgefragt. Sie besteht aus gegensätzlichen Adjektiven, wie zum Beispiel schwer und einfach. Die kurze Version des Fragebogens besteht aus fünf Items für die allgemeine Bewertung des Systems und den 20 Hauptfragen der Konstrukte (Sarodnick & Brau, 2015, S. 190–191).

Der Vorteil von QUIS liegt darin, dass die Subskalen voneinander unabhängig sind. Demnach ist es möglich, den Fragebogen zu kürzen, ohne empirische Validität zu verlieren. Darüber hinaus konnten eine hohe Reliabilität und Validität bewiesen werden (Chin u. a., 1988; Harper, Slaughter, & Norman, 1997; Sarodnick & Brau, 2015, S. 190–191). Ein Nachteil ist, dass der QUIS nicht kostenfrei ist.

2.1.11. SUS

Der Fragebogen SUS, System Usability Scale, untersucht die wahrgenommene Usability der Software für die Nutzer. Wie auch der ISONORM und der IsoMetrics, basiert dieser Fragebogen auf den Dialogprinzipien (Brooke, 1996; DIN EN ISO 9241/10, 1996). Mit 10 Items ist dieser Fragebogen im Vergleich zu anderen sehr kurz. Daher wird er auch als „quick and dirty usability scale“ bezeichnet und eignet

sich, um einen Gesamtüberblick über die Usability einer Software zu erhalten (Brooke, 1996).

Der Fragebogen wurde anhand von Richtlinien erstellt und beinhaltet dadurch keine Suggestivfragen. Somit ist die Güte der Objektivität erfüllt (Brooke, 1996). Auch für den SUS wurden mehrere Studien zur Überprüfung der Validität und Reliabilität durchgeführt (Bangor, Kortum, & Miller, 2008; Lewis & Sauro, 2009). Teilnehmer der Umfrage können die einzelnen Items auf einer fünf-stufigen Likertskala von 1 „trifft nicht zu“ bis 5 „trifft voll zu“ bewerten (Brooke, 1996).

Für die Auswertung der Ergebnisse werden die Werte aller Fragen summiert und mit 2,5 multipliziert. Daraus ergibt sich der sogenannte „SUS Score“, welcher zwischen 0 und 100 liegen kann (Brooke, 1996). Bis 60 Punkte ist mit großen Usability-Problemen zu rechnen. Grenzwertige bis gute Usability hat eine Software mit einem Score von bis zu 80. Bis 100 Punkte ist die Usability gut bis exzellent (Bangor u. a., 2008). Der SUS ist kostenlos in dieser Primärliteratur zu finden (Brooke, 1996).

3. Methode

Nachdem die Grundlagen erklärt wurden, wird jetzt das Verfahren zur Beantwortung der Forschungsfrage erklärt. Dieses basiert hauptsächlich auf dem Artikel Qualitatives Software Screening (Burmester u. a., 1997). Zunächst wird das Ziel der Untersuchung konkretisiert. Für die Untersuchung wird eine passende Methode ausgewählt. Dann wird erklärt, wie die Untersuchung geplant und die Daten erhoben werden. Zuletzt wie die gewonnenen Daten ausgewertet und interpretiert werden.

3.1. Zieldefinition

Nachfolgend sind nochmal kurz alle Ziele des Artikels zusammengefasst. Der zu untersuchende Forschungsgegenstand ist die online Lernplattform Moodle. Für diese soll untersucht werden, wie gebrauchstauglich Studierende der Hochschule der Medien Moodle finden. Daraus ergab sich die Forschungsfrage:

Wie ist die Usability der online Lernplattform Moodle für die Studierenden der Hochschule der Medien?

An der Hochschule der Medien sind 4828 Studierende immatrikuliert (Hochschule der Medien, 2018). Daher beträgt die Grundgesamtheit für die Stichprobe 4828. Basierend darauf ergibt sich eine Stichprobe von 356 bei einem Konfidenzniveau von 95 % und einer Fehlermarge von 5 %.

Aufgrund der zeitlichen Einschränkung dieser Forschungsarbeit, können jedoch nur 20 Stichproben gezogen werden. Dies führt zu einem Konfidenzniveau von 95 %

und einer Fehlermarge von 22 %. Die Stichprobe wurde mit dem Rechner von SurveyMonkey kalkuliert (SurveyMonkey, 2019).

3.2. Methodenauswahl

Nachfolgend wird erläutert, wie die ISONORM 9241/110 Kurzform als Methode für die Untersuchung der Usability ausgewählt wurde. Moodle ist eine bereits fertiggestellte Software, weshalb summative Evaluationsmethoden für die Untersuchung vorteilhaft sind (Sarodnick & Brau, 2015, S. 23).

Weiter zeigt die Literaturrecherche dieses Artikels, dass die Usability von Moodle weder allgemein noch auf konkrete Verbesserungsmöglichkeiten untersucht wurde. Deshalb wird für diesen Artikel ein summativer Fragebogen verwendet (Sarodnick & Brau, 2015, S. 120, 201).

Der ISONORM 9241/110 wurde aus folgenden Gründen unter den summativen Fragebögen ausgewählt. Der SUS wurde ausgeschlossen, da dieser nur eine grobe Einschätzung über die Usability von Moodle bietet (Brooke, 1996). Das Ziel dieses Artikels ist jedoch eine umfassende Bewertung dieser relativ komplexen Software.

Darüber hinaus evaluieren viele Fragebögen nur Teile der Usability, wie der SUMI und QUIS (Sarodnick & Brau, 2015, S. 190–191). Zuletzt sollte der Fragebogen auch kurz genug sein, damit eine ordentliche Ausfüllung durch die Studierenden gewährleistet ist. Aufgrund der 75 Items des IsoMetrics und der 35 Items der Langfassung des ISONORM, wurden diese Fragebögen ausgeschlossen (Günther Gediga u. a., 1999; Prümper & Anft, 1993).

Eine Erklärung des ISONORM 9241/110 erfolgte bereits im Kapitel 4.3.1. *ISONORM*. Darüber hinaus wurde in Kapitel 4.2.8. *Empirische Evaluation* dessen hohe wissenschaftliche Güte aufgezeigt. Nun zur Operationalisierung des Fragebogens. Die Prüfaussagen befinden sich auf einer 7-stufigen Intervallskala von negativen -3 bis zur positiv +3. Dabei gibt es für jedes der sieben Dialogprinzipien 3 Aussagen. Dadurch umfasst der Fragebogen insgesamt 21 Items (Prümper, 1997, 1999).

3.3. Datenerhebung

Nun zur Planung und Durchführung der Datenerhebung. Zuerst wurde eine Umfrage mit Google Forms basierend auf der Kurzform des ISONORM 9241/110 erstellt. Die Versendung des Fragebogens fand über Social Media statt. Unabhängig von Ort und Zeit füllten die Studierenden die Umfrage selbständig zwischen dem 09.01.2019 und 16.01.2019 aus.

3.4. Datenauswertung- und -interpretation

3.4.1. Datenauswertung

Die gewonnenen Daten wurden anhand der Anweisungen des Artikels Qualitatives Software Screening (Burmester u. a., 1997) wie folgt ausgewertet. Zunächst erfolgte die Auswertung der Einstiegsfragen. Zum Beispiel über die Verteilung des Geschlechts und die Beherrschung von Moodle in der Stichprobe. Anschließend die Bewertung der Fragen über die sieben Dialogprinzipien. Übergreifend ist der Richtwert für gebrauchstaugliche Software +1. Somit besteht für alle Werte kleiner als +1 Optimierungsbedarf (Burmester u. a., 1997).

Alle Prüfaussagen wurden in Werte zwischen -3 und +3 umkodiert. Mittels Tabellen erfolgte für jeden Teilnehmer eine Zusammenfassung der 21 Einzelbewertungen in 7 Werte für die sieben Dialogprinzipien. Dies ist in *Tabelle 6: Auswertungstabelle* zu sehen. Damit wurden die Summen und Mittelwerte für die sieben Dialogprinzipien gebildet.

Tabelle 6: Auswertungstabelle

Teilnehmer	Aufgab.	Selbstb.	Erwart.	Lernf.	Steuerb.	Fehlert.	Individ.	Gesamtbeurteilung
1	3	-4	-5	0	-3	0	-1	-10
2	5	2	-1	2	4	0	-1	11
3	3	0	-3	7	0	2	-1	8
4	2	0	3	5	1	1	2	14
5	6	6	6	5	-2	4	-5	20
6	5	1	3	9	2	2	-8	14
7	2	3	2	4	-1	-1	-4	5
8	4	-1	4	6	3	2	0	18
9	5	1	1	2	-4	2	-6	1
10	4	-1	7	0	1	0	-8	3
11	0	-6	1	-1	1	-1	-3	-9
12	-2	-5	-3	-6	-2	-3	-6	-27
13	4	0	3	-4	0	-5	4	2
14	-2	-6	0	3	-1	-1	-1	-8
15	5	2	5	7	6	4	3	32
16	-1	0	-4	-2	0	0	-3	-10
17	3	-3	1	2	-3	-3	-2	-5
18	6	-1	4	2	3	-3	-6	5
19	1	-5	-4	0	-1	-2		-11
20	3	0	2	7	1	5	-3	15
Summe	56	-17	22	48	5	3	-49	68
Mittelwert	2.8	-0.85	1.1	2.4	0.25	0.15	-2.45	0.4857142857

Durch die Summierung der drei Aussagen zu einer, verdreifacht sich das Maximum auf +9 und das Minimum auf -9. Verfolgt man diese Argumentation weiter, verdreifacht sich auch die mögliche Summe auf maximal +60 beziehungsweise -60

minimal. Diese wurde in der *Tabelle 7: Bezeichnungen für Mittelwerte und Summen* nochmals übersichtlich dargestellt (Burmester u. a., 1997).

Tabelle 7: Bezeichnungen für Mittelwerte und Summen

Mittelwerte	Bezeichnung	Summen	Bezeichnung
-7 bis -9	Sehr negativ	-121 bis -180	Sehr negativ
-4 bis -6	Negativ	-61 bis -120	Negativ
-1 bis -3	Eher negativ	-1 bis -60	Eher negativ
0	Neutral	0	Neutral
+1 bis +3	Eher positiv	+1 bis +60	Eher positiv
+4 bis +6	Positiv	+61 bis +120	Positiv
+7 bis +9	Sehr positiv	+121 bis +180	Sehr positiv

Dann erfolgte die Kalkulation der Gesamtbeurteilung, siehe *Tabelle 6: Auswertungstabelle*. Dafür wurden die Werte für die sieben Dialogprinzipien summiert. Dieser Wert ermöglicht einen schnellen Überblick über die Usability zu erhalten und mit anderer Software zu vergleichen. Hier versiebenfachte sich das Maximum auf +63 und das Minimum auf -63. Weiter ist die Summe der Gesamtbeurteilung maximal +1260 und minimal -1260.

3.4.2. Dateninterpretation

Die Interpretation der Ergebnisse erfolgte wiederum anhand des Artikels *Qualitatives Software Screening* (Burmester u. a., 1997). Zum einen basiert diese auf den Umfrageergebnissen. Zum anderen auf einer Literaturrecherche und einer Diskussion mit Studierenden der Hochschule der Medien.

Zuerst wurden die allgemeinen Ergebnisse interpretiert. Zum Beispiel auf die Einwirkung der Nutzungszeit auf die Aussagekraft der Antworten. Darüber hinaus auf die Gesamtbeurteilung und mögliche Gründe dafür. Danach wurden die Resultate für die einzelnen Dialogprinzipien näher betrachtet.

4. Ergebnisse

Dieser Abschnitt ist in zwei Teile gegliedert. Zuerst zeigt der Artikel die Ergebnisse für die Stichprobe und den ISONORM 9241/110 übergreifend auf. Anschließend werden die Resultate für jedes der sieben Dialogprinzipien näher dargestellt.

4.1. Allgemein

An der Online-Umfrage zum ISONORM 9241/110 nahmen 20 Personen teil. Diese waren alle vollständig ausgefüllt, da es nur Pflichtfelder gab. Dabei waren alle Teilnehmer an der Hochschule der Medien immatrikuliert. 25 % der Studierenden weiblich und 75 % männlich. Weiter arbeiten diese im Schnitt seit 16 Monaten mit Moodle. Die tägliche Nutzungsdauer beträgt ungefähr 42 Minuten. Zuletzt liegt der Mittelwert für die Beherrschung der Lernsoftware Moodle bei 4,85.

Insgesamt beträgt der Mittelwert für die Usability von Moodle 0,46. Die genaue Berechnung ist im *Tabelle 6: Auswertungstabelle* zu finden. Dabei ist der kritische Bereich von 0 bis -9 und der positive Bereich von +1 bis +9. Der Mittelwert setzt sich aus allen Mittelwerten für die sieben Dialogprinzipien zusammen. Diese sind in *Abbildung 2: Mittelwerte der 7 Dialogprinzipien* aufgeführt.

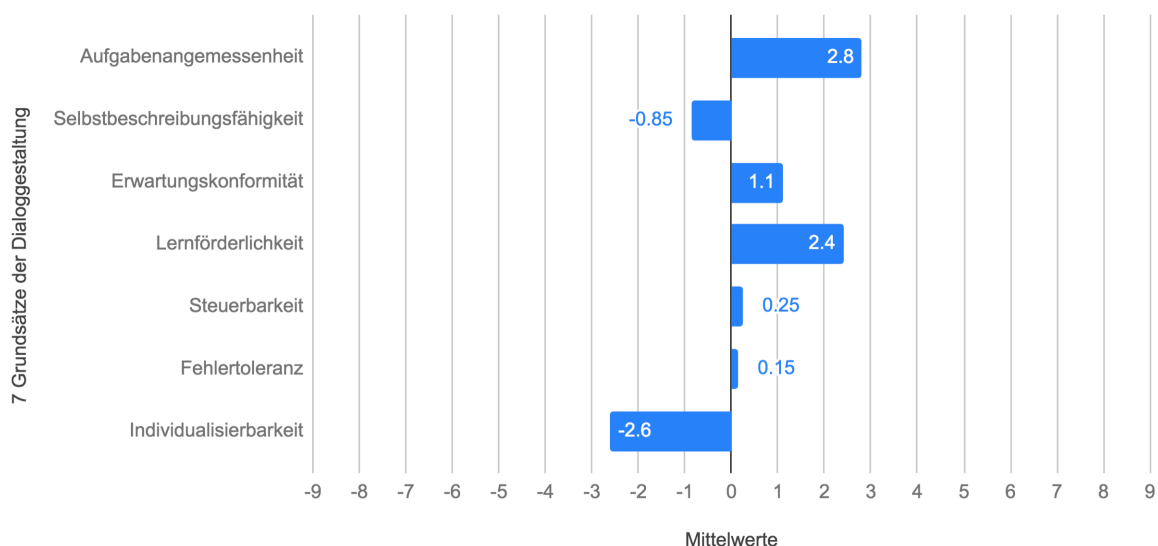


Abbildung 2: Mittelwerte der 7 Dialogprinzipien

Weiter ist die Summe für alle Prüfaussagen 68 Punkte bei einem Maximum von +1260 und Minimum von -1260. Darüber hinaus sind 72 der 140 Prüfaussagen kritisch. Die einzelnen Summen sind in der *Abbildung 3: Summen für die 7 Dialogprinzipien* dargestellt.

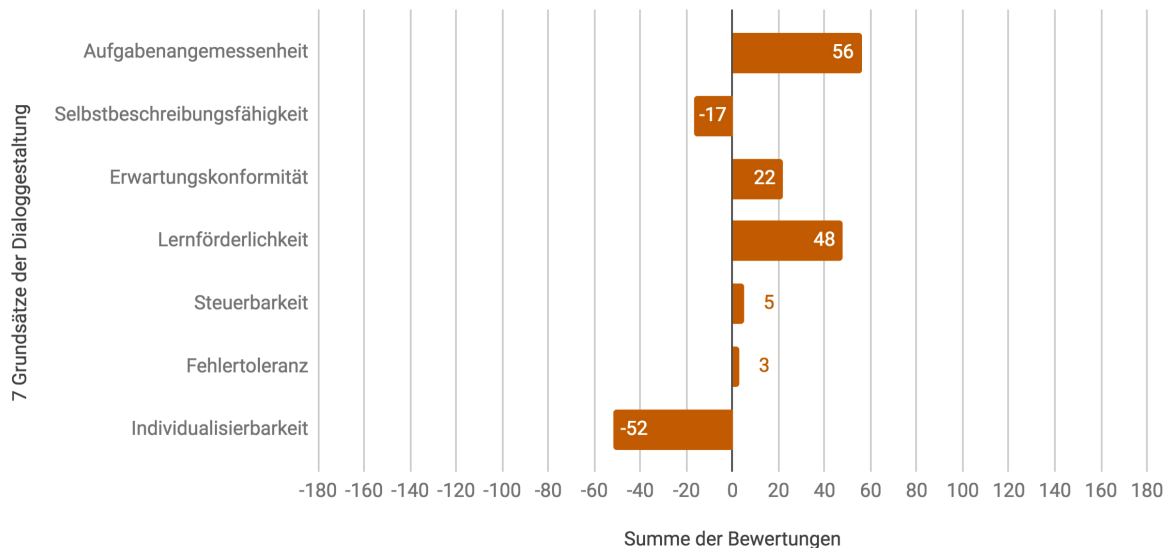


Abbildung 3: Summen für die 7 Dialogprinzipien

Zum einfacheren Verständnis der Ergebnisse wurde den möglichen Mittelwerten beziehungsweise Summen eine Bezeichnung gegeben. Eine Übersicht gibt die *Tabelle 7: Bezeichnungen für Mittelwerte und Summen*.

4.2. Grundsätze der Dialoggestaltung

Nachfolgend wird auf die Ergebnisse der Untersuchung der Usability von Moodle durch den ISONORM 9241/110 eingegangen. Besonders wird auf die kritischen Resultate für die Selbstbeschreibungsfähigkeit, die Steuerbarkeit, die Fehlertoleranz und die Individualisierbarkeit eingegangen.

4.2.1. Aufgabenangemessenheit

Für die Aufgabenangemessenheit beträgt der Mittelwert für die 20 Nutzer 2,8. Dieser kann zwischen -9 und +9 liegen. Die Summe der Bewertungen für alle Prüfaussagen über Aufgabenangemessenheit ist 56 Punkte. Wobei 180 Punkte maximal und -180 Punkte minimal möglich sind. Abschließend sind 4 Prüfaussagen kritisch, da diese kleiner als +1 sind. Einen Überblick über die Ergebnisse liefert die *Abbildung 4: Aufgabenangemessenheit*.

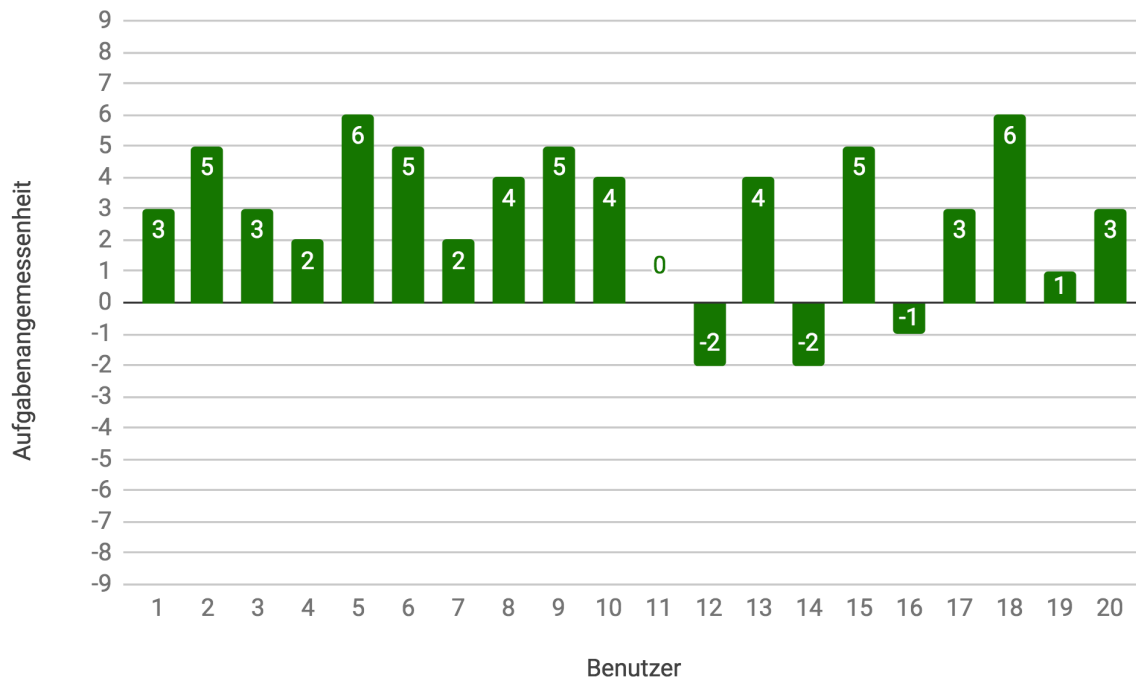


Abbildung 4: Aufgabenangemessenheit

4.2.2. Selbstbeschreibungsfähigkeit

Die Ergebnisse fielen insgesamt eher negativ aus, was in der *Abbildung 5: Selbstbeschreibungsfähigkeit* zu sehen ist. Der Mittelwert liegt bei -0,85. Die Summe der Bewertungen für alle Prüfaussagen beträgt -17 Punkte. Insgesamt sind 14 von 20 Prüfaussagen kritisch ausgefallen.

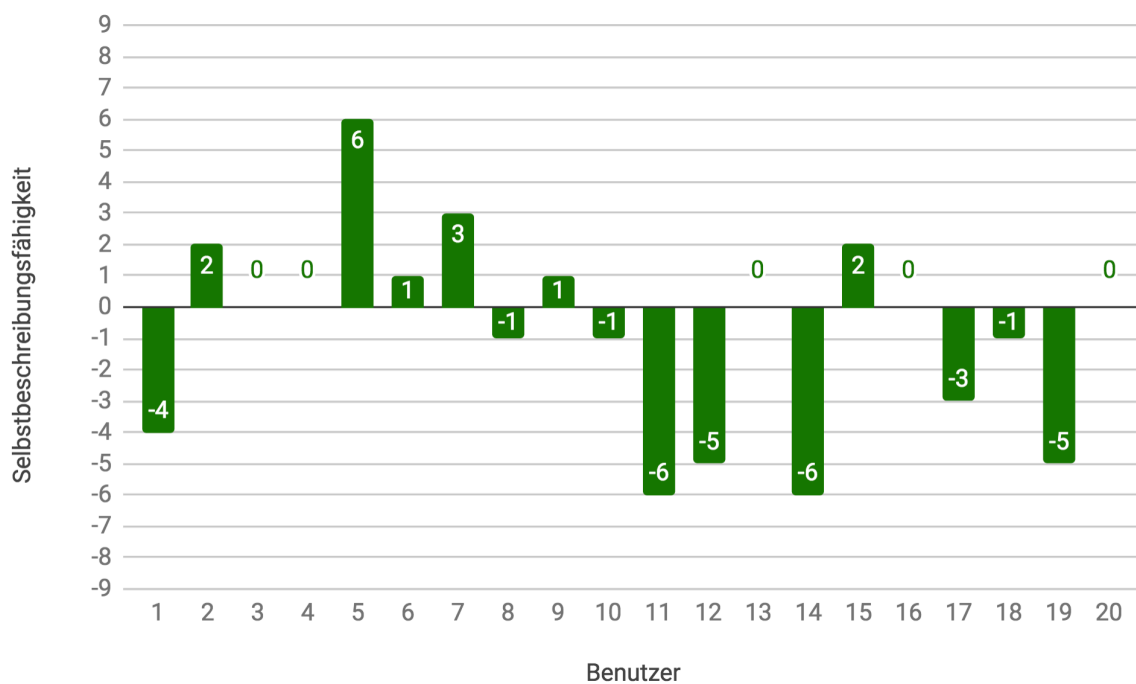


Abbildung 5: Selbstbeschreibungsfähigkeit

4.2.3. Erwartungskonformität

Mit einem Mittelwert von 1,1 sind die Ergebnisse für die Erwartungskonformität eher positiv ausgefallen. Die Summe der Bewertungen für alle Prüfaussagen liegt bei 22 Punkten. Abschließend sind 7 Prüfaussagen kritisch ausgefallen. Die Ergebnisse sind in der *Abbildung 6: Erwartungskonformität* dargestellt.

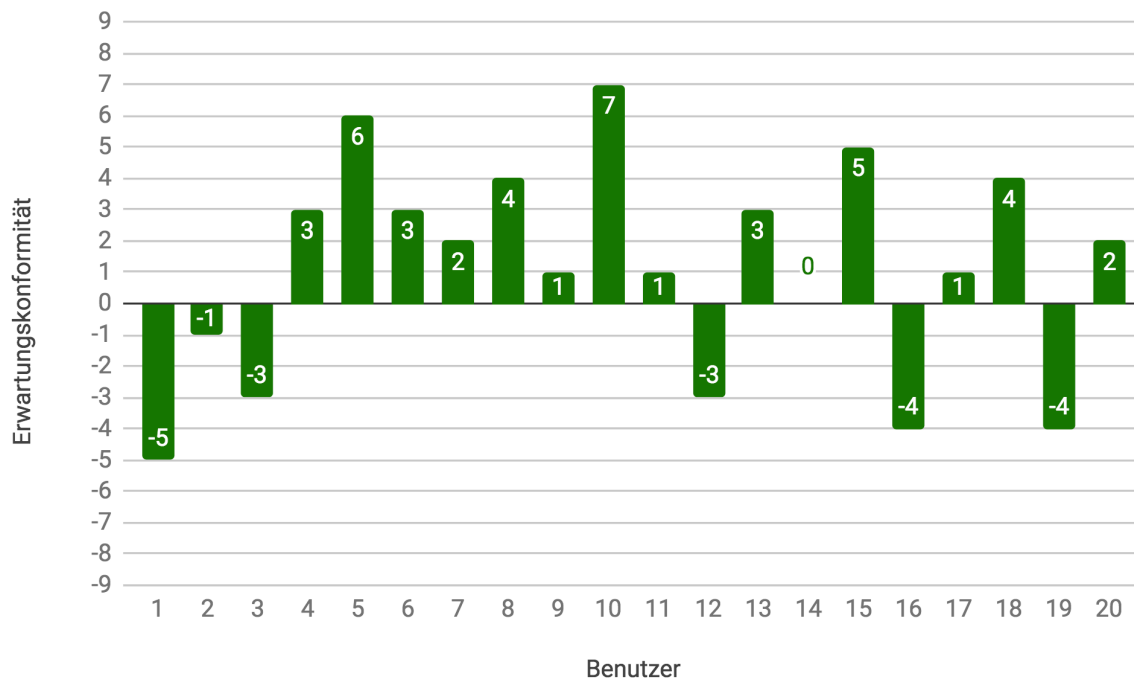


Abbildung 6: Erwartungskonformität

4.2.4. Lernförderlichkeit

Die Ergebnisse über die Lernförderlichkeit wurden auf *Abbildung 7: Lernförderlichkeit* aufgezeigt. Dabei beträgt der Mittelwert 2,4. Die Summe der Bewertungen aller Prüfaussagen beträgt 48 Punkte. 7 Prüfaussagen sind kritisch, da diese Werte kleiner oder gleich 0 sind.

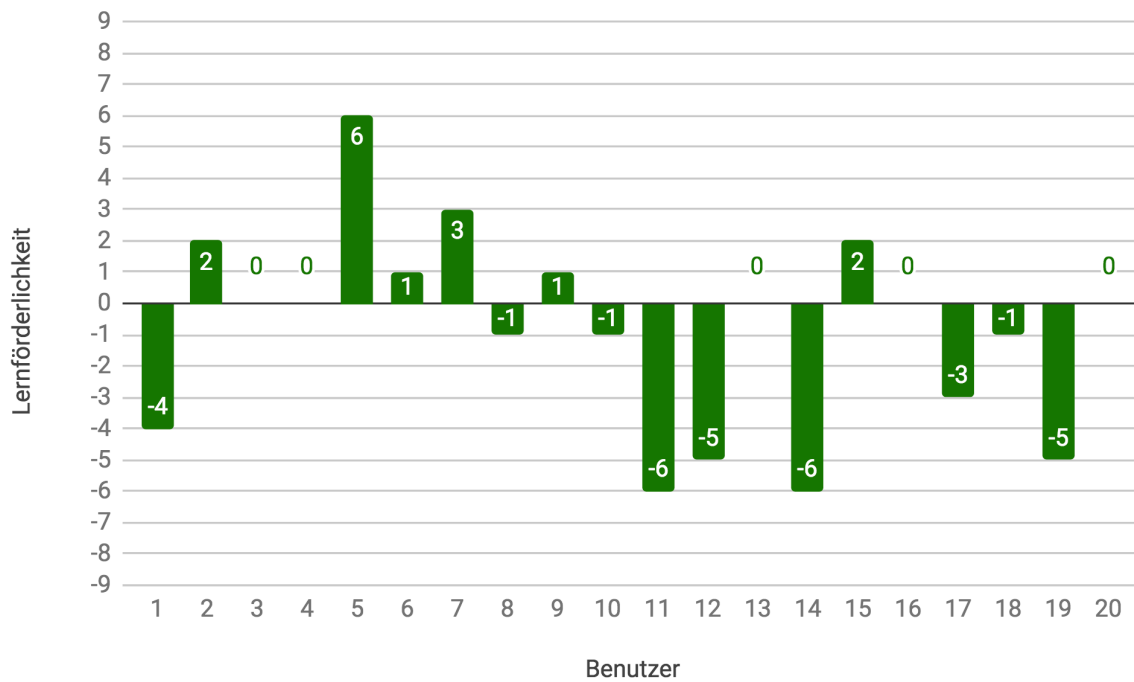


Abbildung 7: Lernförderlichkeit

4.2.5. Steuerbarkeit

Die Steuerbarkeit erreicht den Mittelwert 0,25. Die Summe der Bewertungen für alle Prüfaussagen beträgt 5 Punkte. Insgesamt fallen 11 Werte der Prüfaussagen unter 1, weshalb sie als kritisch anzusehen sind. Die Ergebnisse sind in *Abbildung 8: Steuerbarkeit* nochmals dargestellt.

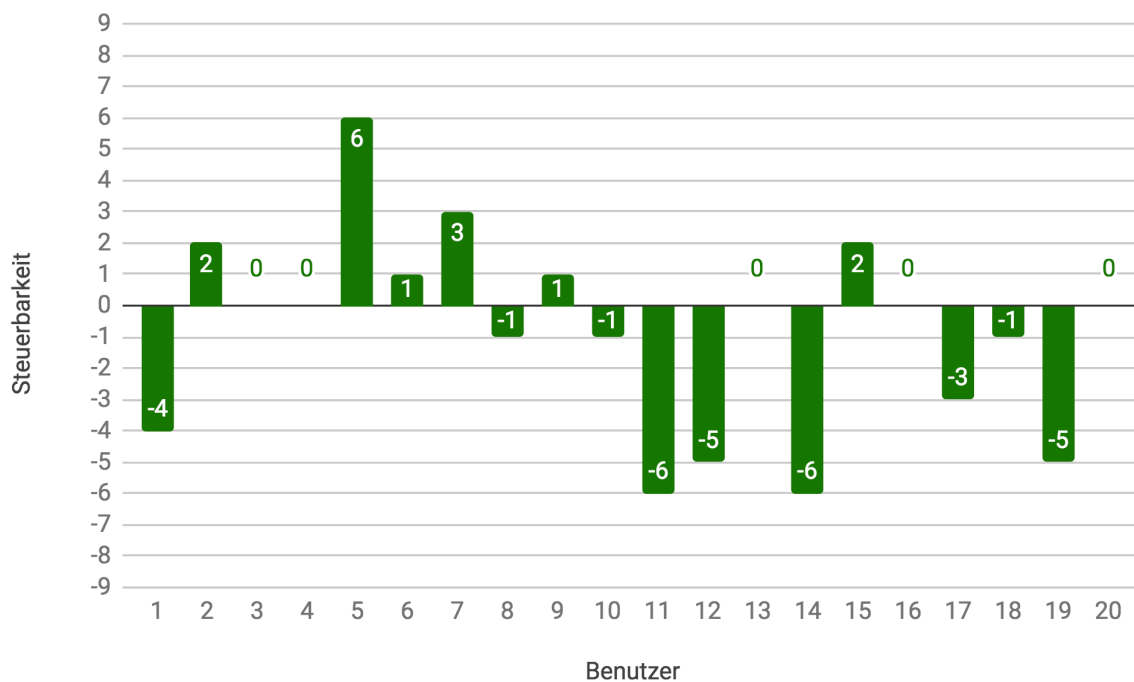


Abbildung 8: Steuerbarkeit

4.2.6. Fehlertoleranz

Die Ergebnisse für die Fehlertoleranz sind positiv ausgefallen, wie in der *Abbildung 9: Fehlertoleranz* dargestellt ist. Der Mittelwert der 20 Nutzer beträgt 0,15. Die Summe der Bewertungen für alle Prüfaussagen über Fehlertoleranz beträgt 3 Punkte. Insgesamt sind 12 Prüfaussagen kritisch.

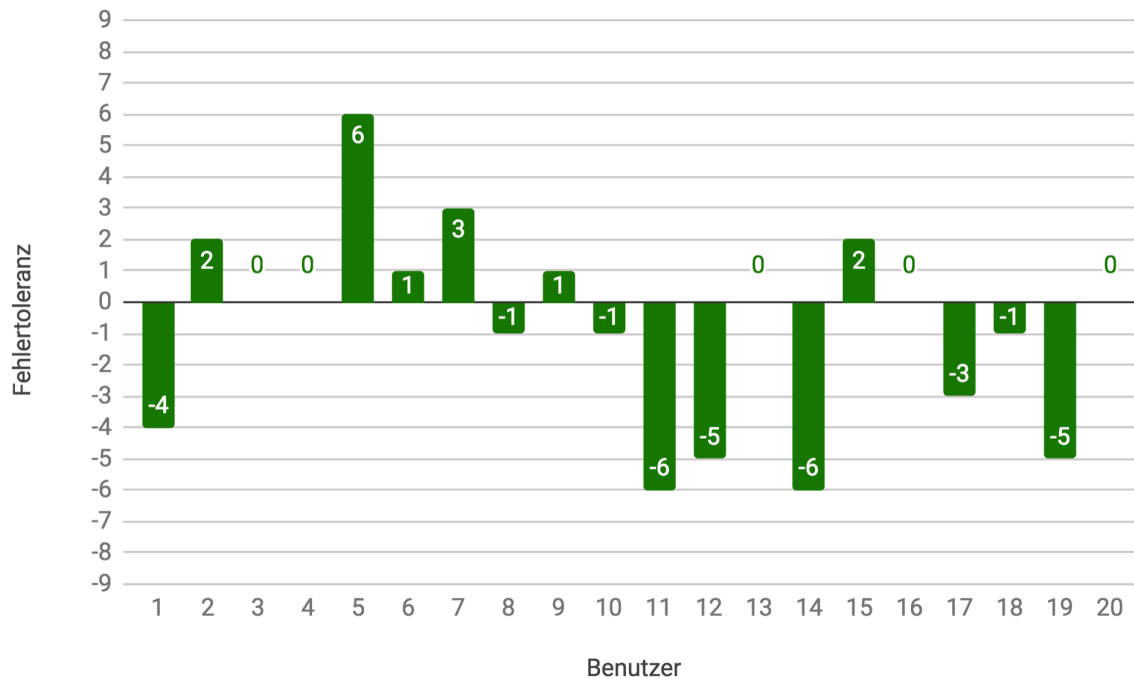


Abbildung 9: Fehlertoleranz

4.2.7. Individualisierbarkeit

Die Ergebnisse für die Individualisierbarkeit sind negativ ausgefallen, wie auf *Abbildung 10: Individualisierbarkeit* zu sehen ist. Der Mittelwert für die 20 Nutzer beträgt -2,6. Die Summe der Bewertungen für alle Prüfaussagen über Individualisierbarkeit ist -52 Punkte. Abschließend sind 17 Prüfaussagen als kritisch zu erachten, da diese Werte kleiner oder gleich 0 sind.

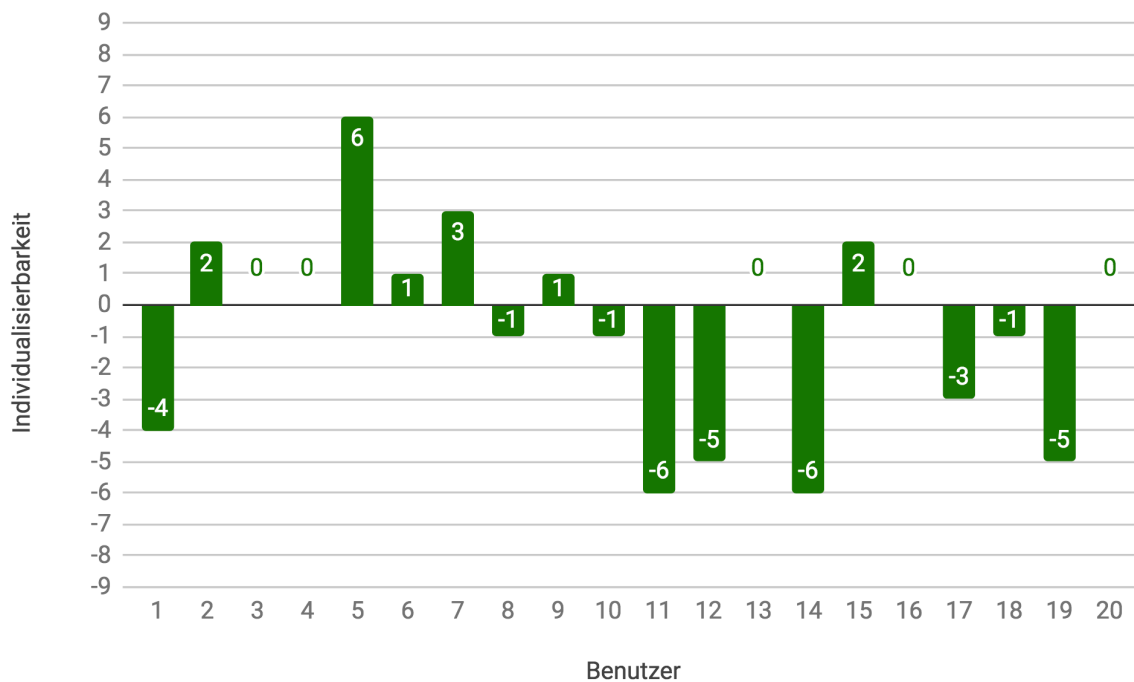


Abbildung 10: Individualisierbarkeit

5. Diskussion

Auch dieser Abschnitt besteht aus zwei Teilen. Zuerst werden die allgemeinen Resultate für die Stichprobe und die Gesamtbeurteilung des ISONORM 9241/110 interpretiert. Anschließend erfolgt eine Analyse der einzelnen Ergebnisse für jedes der sieben Dialogprinzipien.

5.1. Allgemein

Die Stichprobe umfasst 20 und nicht die notwendigen 356 Antworten. Dies ist begründet mit der zeitlichen Einschränkung für dieser Forschungsarbeit. Daher sind die Ergebnisse nicht repräsentativ für die Grundgesamtheit, sondern nur für die Stichprobe. Weiter sind mit 75 % die Männer in der Umfrage überrepräsentiert, da die Geschlechterquote an der Hochschule der Medien bei 50/50 liegt. Dies kann wiederum die gewonnen Ergebnisse verzerren (Hochschule der Medien, 2018).

Prümper und Anft (1993) geben an, dass für aussagekräftige Beurteilungen, viel Erfahrung mit der jeweiligen Software nötig ist. Kritisch ist daher zum einen die tägliche Nutzungsdauer von ungefähr 42 Minuten. Zum anderen die insgesamt Nutzungsdauer von 16 Monaten. Besonders diese ist im Vergleich zum gesamten Studium, welches 42 Monate beträgt, niedrig. Auch die Beherrschung von Moodle ist mit 0,85 eher niedrig. Dies verstärkt das zuvor angesprochene Argument, dass die Erfahrung entscheidend ist (Prümper & Anft, 1993).

Abschließend liegt die Gesamtbeurteilung der Usability mit 0,46 im kritischen Bereich. Dies zeigt, dass die Lernplattform Moodle verbessert werden muss. Jedoch wird auch deutlich, dass diese keine schlechte Software ist.

5.2. Grundsätze der Dialoggestaltung

Die nachfolgende Interpretation der Dialogprinzipien gibt Aufschluss darüber, an welchen Stellen Moodle verbessert werden muss. Dabei erfolgt eine besondere Analyse der kritischen Resultate der Selbstbeschreibungsfähigkeit, der Steuerbarkeit, der Fehlertoleranz und der Individualisierbarkeit.

5.2.1. Aufgabenangemessenheit

Die Ergebnisse zeigen, dass Moodle die Nutzer unterstützt, ihre Arbeitsaufgaben effektiv und effizient zu erledigen. Diese eher positive Tendenz zeigen der Mittelwert von 2,8, die Summe +56 und die 4 kritischen Prüfaussagen. Im Vergleich mit anderen Prinzipien sind diese Resultate am positivsten. Deshalb steht die Verbesserung der Aufgabenangemessenheit nicht im Vordergrund.

Im Gruppengespräch mit den Testteilnehmern wurde als Beispiel für die schlechte Bewertung die Verständlichkeit des Upload-Tool genannt. Sie beschwerten sich über die unnötige Komplexität im Vergleich zu Upload-Werkzeugen, wie zum Beispiel Google Drive. Also über zu hohe Anzahl an Einzelschritten, um eine Datei hochzuladen. Weiter wird die Suchfunktion bemängelt. Der Grund dafür ist, dass bei der Kurssuche viele irrelevante und ähnelnde Ergebnisse aufgelistet werden. Ansonsten wünschen sich die Studierenden die Einträge auf der Startseite sortieren zu können.

Ein Artikel weist auf die Schwierigkeiten der Kommunikations-Werkzeuge von Moodle hin (Ivanović u. a., 2013). Also dem Forum und der Chat-Funktion. Diese werden von zahlreichen Studierenden und Lehrkräften selten oder gar nicht genutzt (Ivanović u. a., 2013). Diese Forscher stellen fest, dass dieses Werkzeug oft nicht gebraucht wird, da andere Programme oder eine persönliche Kommunikation bevorzugt werden.

Darüber hinaus sind die Tools ineffizient. Demnach werden unnötig viele Schritte für das Schreiben und Erhalten von Nachrichten benötigt. Besonders durch die Abundanz zahlreicher Kommunikationstools, wie WhatsApp und Google Mail, haben viele Nutzer heutzutage hohe Anforderungen (Ivanović u. a., 2013).

Trotz der Kritik gab es allgemeine Zustimmung darüber, dass die Aufgabenangemessenheit von Moodle zufriedenstellend ist. Die Verfasser dieses Artikels mutmaßen, dass die meisten Aufgaben grundsätzlich erledigt werden können und kein großer Aufwand dafür notwendig ist.

Somit sind die Studierenden mit dem Funktionsumfang allgemein zufrieden. Also mit dem Bereitstellen und Abrufen von Vorlesungsmaterial, der Kalenderfunktion, dem Lösen von Aufgaben und dem Upload von Abgaben. Auch ein Artikel stützt die These (Kakasevski u. a., 2008).

5.2.2. Selbstbeschreibungsfähigkeit

Dieses Prinzip wurde insgesamt eher negativ bewertet. Diese Interpretation basiert auf dem Mittelwert von -0,85, der Summe von -17 und der 14 kritischen Prüfaussagen. Somit sollten die einzelnen Dialogschritte durch Rückmeldungen besser verständlich gemacht oder auf Anfrage erklärt werden.

Während der Diskussion, gaben die Studierenden an, dass das Lesen von Studienmaterial innerhalb der Kursangebote unnötig komplex ist. Vor allem die Navigation zwischen den einzelnen Kapiteln aufgrund der fehlenden Beschreibungen der Pfeile. Diese führten vermehrt zu Fehleingaben, wodurch die Nutzer auf unerwünschte Seiten geleitet wurden.

Darüber hinaus wurde die Übersichtlichkeit der Startseite als mangelhaft bezeichnet. Die Informationen werden dort vertikal aufgelistet. Dies ist vorteilhaft für eine mobile Ansicht. Jedoch wird bei bereiten Bildschirmen in der Desktop-Version die Anzeige nicht angepasst. Durch eine sowohl vertikale als auch horizontale Verteilung der Information auf dem Bildschirm, könnten Informationen schneller erfasst werden.

Ein Artikel zeigt (Rakoczi, 2010), dass die Kursangebote visuell nicht klar getrennt sind. Um die deutlichere Unterscheidbarkeit der Kurse zu erzielen und damit die Startseite übersichtlicher zu machen, empfiehlt Rakoczi die Verwendungen von klaren Konturen oder starken Kontrasten, für die einzelnen Kursangebote.

Die bereits angesprochene geringe Nutzung der Kommunikations-Tools von Moodle, könnten auch auf eine nicht ausreichende Selbstbeschreibungsfähigkeit zurückzuführen sein. Höbl und Welzer (2010) vermuten, dass die Funktionen, aufgrund schlechter Verständlichkeit, nicht entdeckt beziehungsweise nach der Nutzung vergessen wurden. Weiter werden das Mitteilungs-Symbol, welche sich oben rechts im Interface befindet, und das Forums-Symbol nicht genügend hervorgehoben. Dadurch fallen diese Funktionen den Studierenden oft nicht auf.

5.2.3. Erwartungskonformität

Der Mittelwert ist 1,1, die Summe der Aussagen +22 und es gibt 7 kritischen Prüfaussagen. Die meisten Teilnehmer denken, dass Moodle konsistent ist und ihren Merkmalen entspricht. Also mit den Erfahrungen und allgemein anerkannten Konventionen übereinstimmt. Somit ist eine Optimierung von Moodle für dieses Prinzips eher zweitrangig.

In der Diskussion nannten die Studierenden, dass sie den Zugang zum Forum und der Chat-Funktion an einer anderen Stelle erwartet hätten. Daher haben diese die Funktionen oft übersehen. Weiter haben sie erwartet, dass geänderte und neue Inhalte offensichtlicher dargestellt werden.

Von der Startseite hätten sich diese erwartet, dass dort alle ihre Kurse und die Neuigkeiten auf einen Blick aufzufinden sind. Jedoch wird dies nicht von der Startseite geboten. Durch die vertikale Anzeige, die Auflistung aller besuchten Kurse ohne Semester Zugehörigkeit und die Aufzählung von mehr als 30 Dozenten ist diese unübersichtlich.

5.2.4. Lernförderlichkeit

Nach Einschätzung der Teilnehmer können die Studierenden ohne großen Aufwand die Nutzung von Moodle erlernen. Somit wird dieses Prinzip eher positiv wahrgenommen. Dies zeigt auch der Mittelwert von 2,4 und die Summe von +48 Punkten. Von den Prüfaussagen sind 7 kleiner +1, weshalb dieses Prinzip trotz der anderen hohen Werte, näher untersucht werden sollte.

Ein Grund für die positive Einschätzung von Moodle liegt in der durchgängigen Verwendung von Symbolen und Icons. Diese ermöglichen eine intuitive Bedienung der Benutzeroberfläche. Auch sind die meisten Seiten einfach gestaltet, wodurch nur eine geringe kognitive Leistung erforderlich ist. Dadurch fällt es den Studierenden einfach, die Website zu verstehen und zu bedienen (Debevc & Bele, 2008).

Ausbaufähig ist die Lernförderlichkeit in Bezug auf die bereits genannten Kommunikationstools. Vor allem für die Chat-Funktion und das Forum. Deren Funktion ist für Studierende meist nicht offensichtlich, wodurch hier Verbesserungspotential besteht. Eine auffälligere Gestaltung oder ein erklärender Text könnte dieses Problem lösen (Hölbl & Welzer, 2010).

5.2.5. Steuerbarkeit

Die Abbildung 8: Steuerbarkeit zeigt, dass die Wahrnehmung der Steuerung von Moodle von Nutzer zu Nutzer stark schwankt. Dies ist auch an den 11 kritischen Aussagen zu erkennen. Im Durchschnitt sind diese jedoch mit einem Mittelwert von 0,25 und einer Summe der Prüfaussagen von +5 neutral. Die Studierenden sind sich also nicht einig, wie gut sie sich in der Lage sehen, den Dialogablauf zu starten und seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist. Somit sollte dieses Prinzip unbedingt näher betrachtet werden, obwohl das Ergebnis neutral ist.

Leider führten die Diskussion und Literaturrecherche zu keinen brauchbaren Hinweisen auf die Bewertung der Steuerbarkeit. Deshalb wird nachfolgend die Meinung der Verfasser dieser Studie dargelegt. Zum einen ist es jeder Zeit möglich zwischen zwei Seiten mithilfe der Navigation zu wechseln. Darüber hinaus sind die Arbeitsschritte einheitlich gestaltet und einfach steuerbar.

Jedoch steckt Verbesserungspotential für die Navigation des Studienmaterials. Dabei sind die Navigationselemente innerhalb der Kapitel oft missverständlich platziert. Dadurch kann es sein, dass Studierende auf ungewünschte Seiten geleitet werden.

5.2.6. Fehlertoleranz

Dieses Prinzip liefert ein ähnliches, neutrales Resultat wie die Steuerbarkeit. Hier sollten Nutzer trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder minimalem Korrekturaufwand ihr Arbeitsziel erreichen können. Der Mittelwert beträgt 0,15 und die Summe der Prüfaussagen ist +3. Die 12 kritischen Prüfaussagen sowie die *Abbildung 9: Fehlertoleranz* veranschaulichen, dass sich die Beurteilungen auch hier stark unterscheiden und Moodle für dieses Prinzip optimiert werden muss.

Wieder lieferte die Diskussion und die Recherche Hinweise für die Bewertung. Zum einen sind die Fehlermeldungen von Moodle sehr auffällig. Weiter werden Falscheingaben entweder durch ein Pop-up-Fenster oder durch rote Anmerkungen neben dem fehlerhaften Dialogfeld angezeigt. Auch wird eine Fehlermeldung ausgegeben, wenn Nutzer versuchen, zu große Dateien hochzuladen.

Jeder Schritt kann wieder rückgängig gemacht werden. So können sich die Studierenden aus einem Kurs abmelden, sollten sich diese in den falschen eingeschrieben haben. Bei der Abgabe einer Prüfungsleistung ist es möglich die Datei nach dem Upload wieder zu ersetzen oder zu entfernen. Kalendereinträge sind ebenfalls löschar. Allgemein gibt es keine Arbeitsschritte, deren Korrektur einen hohen Aufwand erfordert. Da den Verfassern dieser Studie keine negativen Aspekte für die Fehlertoleranz aufgefallen sind, sollte dieses Prinzip unbedingt näher betrachtet werden.

5.2.7. Individualisierbarkeit

Der Mittelwert dieses Prinzips ist -2,6, die Summe liegt bei -52 Punkten und 17 von 20 Aussagen sind kritisch. Somit wird die Individualisierbarkeit eher negativ wahrgenommen. Darauf basierend kann angenommen werden, dass eine Anpassung an die Anforderungen für ihre Arbeitsaufgabe sowie an ihre Fähigkeiten und Präferenzen für die Nutzer nur schwer möglich ist. Im Vergleich zu den anderen

Prinzipien sind diese Resultate am schlechtesten. Daher sollte der Verbesserung dieses Prinzips eine hohe Priorität in der Weiterentwicklung der Software erhalten.

In einer offenen Diskussionsrunde mit den Probanden kam der konkrete Wunsch auf, die Bezeichnung der Kurse zu individualisieren. Die kryptischen Namen der Veranstaltungen verwirren die Studierenden und verlangsamen, aufgrund ihres schlechten Wiedererkennungswertes, die Arbeit mit Moodle.

Auch wünschen sich die Studierenden, die Auflistung der Veranstaltungen auf der Startseite nach unterschiedlichen Kriterien sortieren zu können. Zum Beispiel nach dem Alphabet oder Datum. Auch sollte es möglich sein, die Schriftgröße und den Kontrast einzustellen, um die Barrierefreiheit von Moodle zu gewährleisten. Jedoch sollten diese individuellen Anpassungen nicht die Usability beeinträchtigen. Zum Beispiel, dass die Farbe der Schrift so angepasst werden kann, dass diese nicht mehr lesbar ist und der Nutzer somit feststeckt (DIN EN ISO 9241/11, 2015).

Im Hinblick auf Individualisierung wäre es denkbar das Interface auf die Vertrautheit des Nutzers mit Moodle anzupassen. So könnte Experten ein komplexeres Layout angezeigt werden, welches ihnen erlaubt schneller zu arbeiten. Anfänger würden ein reduziertes Interface vorfinden, durch welches sie Zugriff auf die standardmäßigen Funktionen von Moodle haben (Tee, Wook, & Zainudin, 2013).

6. Fazit

Das Fazit ist in zwei Abschnitte aufgeteilt. Zuerst die theoretischen und dann die praktischen Schlussfolgerungen.

6.1. Theoretisches Fazit

Ein hoher Zeitfaktor bei der Durchführung der Studie lag in der Auswahl des korrekten Fragebogens. Die Frage, ob Moodle als klassische Software anzusehen ist, welche sich mit dem ISONORM testen lässt, war nicht sofort zu beantworten. Auch die hohe Anzahl an verschiedenen Usability-Fragebögen war anfangs ein Hindernis. Der Anspruch war, eine Ausgewogenheit zwischen einer effizienten Befragung und einer großen Menge gesammelter Informationen zu erhalten.

Innerhalb der Recherche über den Fragebogen-Erstellungsprozess ergab sich das Bild einer großen Diskrepanz zwischen vielen Experten. Die Aufteilung der Schritte, die es benötigt einen Fragebogen zu erstellen und die dabei gelegten Schwerpunkte unterschieden sich deutlich.

Abschließend lässt sich feststellen, dass Fragebögen eher für eine allgemeine Bewertung der Usability geeignet sind. Konkrete Verbesserungsmöglichkeiten

können zum Beispiel mit Einzelinterviews und Fokusgruppen gefunden werden (Burmester u. a., 1997).

6.2. Praktisches Fazit

Fazit für ISONORM

Nachfolgend wird aufgezeigt, weshalb wir davon überzeugt sind, dass der ISONORM 110 eine gute Möglichkeit darstellt, um die Usability einer Software zu evaluieren. Der Fragebogen bildet sehr gut die 7 Dialogprinzipien ab. Dadurch konnte herausgefunden werden, in welchen Bereichen die Usability optimiert werden sollte.

Da der ISONORM auf einer wissenschaftlichen Norm basiert und die Gütekriterien, wie zuvor gezeigt, erfüllt, ist mit einer Aussagekraft der Ergebnisse zu rechnen (DIN EN ISO 9241/110, 2006; Prümper, 1997, 1999).

Der ISONORM 110 ist besonders gut für Teilnehmer geeignet, die wenig Zeit haben beziehungsweise sich wenig Zeit nehmen wollen. In unserem Fall waren das die Studierenden der Hochschule der Medien in Stuttgart. Die Kürze und die Prägnanz des Fragebogens haben dies gewährleistet, wie wir aus der Diskussion erfahren haben.

Fazit für Moodle

Die neutrale Gesamtbeurteilung mit 0,46 zeigt, dass die Usability von Moodle ist zwar verbesserungswürdig, jedoch ausreichend ist (Burmester u. a., 1997). Somit ist die Benutzerfreundlichkeit von Moodle schon auf einem hohen Niveau. Weiter wäre es sinnvoll dieses Resultat mit der Bewertung anderer online Lernplattformen zu vergleichen, um das Ausmaß der Probleme besser einschätzen zu können.

Das beste Ergebnis von 2,8 erfuhr die Aufgabenangemessenheit. Dies lässt vermuten, dass die Studierenden grundsätzlich zufrieden sind, da sie zielführend ihre Aufgaben erledigen können.

Die Selbstbeschreibungsfähigkeit, Steuerbarkeit und Fehlertoleranz erhielten neutrale Ergebnisse. Diese sind ausreichend, aber dennoch stark ausbaufähig. Weiterführend würde es sich am ehesten lohnen, die Selbstbeschreibungsfähigkeit zu steigern.

Die Verständlichkeit von Moodle ist bereits mit einfachen Mitteln gut zu optimieren. Die Umgestaltung der Oberfläche nach Grundsätzen wie die der guten Form oder ähnlichen Gestaltungs-Heuristiken können hier mit wenig Aufwand bereits einen großen Unterschied erzielen.

Den letzten Platz in der Bewertung erhält die Individualisierbarkeit. Jedoch ist an diesem Punkt anzumerken, dass die Dialogprinzipien von den Testteilnehmern im Fall von Moodle nicht als gleichwertig empfunden werden. Individualisierbarkeit mag auf der Lernplattform geringfügig verfügbar sein. Dennoch ist es gut möglich, dass sich die Studierenden keineswegs daran stören.

Die Verbesserung der Individualisierbarkeit könnte in verschiedenen Layout Varianten der Startseite verwirklicht werden. Eine Moduldarstellung wäre möglich. Auch ein Fähigkeiten-orientiertes Interface wäre denkbar, welches sich auf die Kenntnisse des Nutzers anpasst (Tee u. a., 2013).

Der weitere Prozess der Optimierung der Usability im Fall von Moodle erfordert definitiv zusätzliche Methoden. Wie bereits erwähnt, eignet sich der ISONORM 110 für ein allgemeines Bild der Usability und gibt erste Hinweise. Um die kritischen Punkte näher zu betrachten, kann eine Spezifizierung anhand der Fokusgruppen Methode stattfinden. Denkbar sind auch Einzelinterviews (Burmester u. a., 1997).

Alternativ käme eine Schulung der Studierenden infrage. In Form von Workshops könnten regelmäßig Neulinge in die Software eingeführt werden. Dies würde gewährleisten, dass die Studierenden alle Funktionen kennen und die Verwendung dieser beherrschen (Kiget, Wanyembi, & Peters, 2014).

An letzter Stelle der Optionen steht das Umsteigen auf eine neue Software. Der Austausch von Moodle würde jedoch einen unnötig hohen Aufwand bedeuten (Burmester u. a., 1997). Die Ergebnisse sind mit Sicherheit nicht so gravierend, dass diese Möglichkeit in Betracht gezogen werden sollte.

Um die Verbesserung von Moodle anzustreben, wäre es denkbar, die Studie mit den Entwicklern von Moodle zu teilen.

Literatur

- Babbie, E. R., & Mouton, J. (2001). *The practice of social research*. Cape Town: Oxford University Press Southern Africa.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6), 574–594.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer Medizin Verlag.
- Bösser, T. (2019, Januar 30). SUMI Background Reading. Abgerufen 30. Januar 2019, von <http://sumi.uxp.ie/>
- Breakwell, G. M., Smith, J. A., & Wright, D. B. (2012). *Research Methods in Psychology*. SAGE.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4–7.
- Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. Pearson Deutschland GmbH.
- Burmester, M., Görner, C., Vossen, P. H., Zolleis, T. M., & Zouboulides, V. (1997). Qualitatives software screening. *Das SANUS-Handbuch. Bildschirmarbeit EU-konform*, 2–33.
- Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an Instrument Measuring User Satisfaction of the Human-computer Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 213–218). New York, NY, USA: ACM. <https://doi.org/10.1145/57167.57203>
- Debevc, M., & Bele, J. L. (2008). Usability testing of e-learning content as used in two learning management systems. *European Journal of Open, Distance and E-learning*, 11(1).
- DIN EN ISO 9241/10. (1996). Teil 10: Grundsätze der Dialoggestaltung. In *Ergonomie der Mensch-System-Interaktion*. Berlin: Beuth Verlag.
- DIN EN ISO 9241/11. (2015). Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte. In *Ergonomie der Mensch-System-Interaktion*. Berlin: Beuth Verlag.
- DIN EN ISO 9241/110. (2006). Teil 110: Grundsätze der Dialoggestaltung. In *Ergonomie der Mensch-System-Interaktion*. Berlin: Beuth Verlag.
- DIN EN ISO 9241/210. (2011). Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme. In *Ergonomie der Mensch-System-Interaktion*. Berlin: Beuth Verlag.

- Endmann, A., Ackermann, L., Thielsch, M. T., & Hess, S. (2014). Arbeitskreis User Research-Was ist User Research? In *UPI4-Workshops*. German UPA.
- Figl, K. (2010). Deutschsprachige Fragebögen zur Usability-Evaluation im Vergleich. *Zeitschrift für Arbeitswissenschaft*, 4, 321–337.
- Figl, K., & Mensch, U.-F. im V. T. (2009). ISONORM 9241/10 und Isometrics: Usability-Fragebögen im Vergleich. In *Mensch & Computer* (Bd. 9, S. 143–152).
- Gediga, G., Hamborg, K. C., & Willumeit, H. (1998). The IsoMetrics Manual (Osnabrücker Schriftenreihe Software-Ergonomie Nr. 7). *Osnabrück: Universität Osnabrück, Fachbereich Humanwissenschaften*, 73–84.
- Gediga, G., & Hamborg, K.-C. (2019, Januar 30). IsoMetrics Questionnaires. Abgerufen 30. Januar 2019, von <http://www.isometrics.uni-osnabrueck.de/qn.htm>
- Gediga, Günther, & Hamborg, K.-C. (1999). IsoMetrics: An Usability Inventory Supporting Summative and Formative Evaluation of Software Systems. In *Proceedings of HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I - Volume I* (S. 1018–1022). Hillsdale, NJ, USA: L. Erlbaum Associates Inc.
- Gediga, Günther, Hamborg, K.-C., & Düntsch, I. (1999). The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behav. Inf. Technol.*, 18(3), 151–164.
- Gediga, Günther, Hamborg, K.-C., & Düntsch, I. (2002). Evaluation of software systems. *Encyclopedia of computer science and technology*, 45(supplement 30), 127–53.
- Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-computer interaction*, 13(3), 203–261.
- Harper, B. D., Slaughter, L. A., & Norman, K. L. (1997). Questionnaire Administration Via the WWW: A Validation & Reliability Study for a User Satisfaction Questionnaire. In *WebNet* (Bd. 97, S. 1–4).
- Heinsen, S. (2003). *Usability praktisch umsetzen: Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte*. Hanser.
- Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: ensuring usability through product & process*. John Wiley & Sons, Inc.
- Hochschule der Medien. (2018). *Jahresbericht für das Akademische Jahr 2016/2017*. Abgerufen von https://www.hdm-stuttgart.de/hochschule/profil/hdmberichte/HdM_Jahresbericht_2016_2017.pdf
- Ivanović, M., Putnik, Z., Komlenov, Ž., Welzer, T., Hölbl, M., & Schweighofer, T. (2013).

Usability and privacy aspects of Moodle: students' and teachers' perspective. *Informatica*, 37(3).

Jonkisz, E., Moosbrugger, H., & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In *Testtheorie und Fragebogenkonstruktion* (S. 27–74). Springer.

Kabungaidze, T., Mahlatshana, N., & Ngirande, H. (2013). The impact of job satisfaction and some demographic variables on employee turnover intentions. *International Journal of Business Administration*, 4(1), 53.

Kakasevski, G., Mihajlov, M., Arsenovski, S., & Chungurski, S. (2008). Evaluating usability in learning management system moodle. In *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on* (S. 613–618). IEEE.

Kallus, K. W. (2016). *Erstellung von Fragebogen* (2. Aufl., Bd. 4465). Wien: Facultas.

Kiget, N. K., Wanyembi, G., & Peters, A. I. (2014). Evaluating usability of e-learning systems in universities. *International Journal of Advanced Computer Science and Applications*, 5(8), 97–102.

Kirakowski, J. (1994). The use of questionnaire methods for usability assessment. *Unpublished manuscript. Recuperado el*, 12.

Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. *British journal of educational technology*, 24(3), 210–212.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78.

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In *International conference on human centered design* (S. 94–103). Springer.

Martin, L., Martínez, D. R., Revilla, O., Aguilar, M. J., Santos, O. C., & Boticario, J. G. (2008). Usability in e-Learning Platforms: heuristics comparison between Moodle, Sakai and dotLRN. In *Sixth International Conference on Community based environments. Guatemala* (S. 12–16).

Melton, J. (2006). The LMS Moodle: A usability evaluation. *Prefectural University of Kumamoto. Retrieved February, 21, 2008*.

Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative usability evaluation. *Behaviour & Information Technology*, 23(1), 65–74.

Molich, R., Thomsen, A. D., Karyukina, B., Schmidt, L., Ede, M., Van Oel, W., & Arcuri, M. (1999). Comparative evaluation of usability tests. In *CHI'99 extended abstracts on Human factors in computing systems* (S. 83–84). ACM.

Nielsen, J. (1989). Usability engineering at a discount. In *Proceedings of the third international conference on human-computer interaction on Designing and using human-computer interfaces and knowledge based systems (2nd ed.)* (S. 394–401). Elsevier Science Inc.

Nokelainen, P. (2006). An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. *Journal of Educational Technology & Society*, 9(2).

Prümper, J. (1997). Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In R. Liskowsky, B. M. Velichkovsky, & W. Wünschmann (Hrsg.), *Software-Ergonomie '97: Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung* (S. 253–262). Wiesbaden: Vieweg+Teubner Verlag.

Prümper, J. (1999). Test It: ISONORM 9241/10. In *HCI (1)* (S. 1028–1032). f3.htw-berlin.de.

Prümper, J. (2008). ISONORM 9241/110 Langfassung: Beurteilung von Software auf Grundlage der Internationalen Ergonomie-Norm DIN EN ISO 9241-110. Abgerufen von http://projekt.kke.tu-berlin.de/wp-content/uploads/2015/09/Methode_Isonorm-Fragebogen.pdf

Prümper, J., & Anft, M. (1993). Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel. In *Software-Ergonomie'93* (S. 145–156). Springer.

Rakoczi, G. (2010). Cast your eyes on moodle: An eye tracking study investigating learning with Moodle. In *Proceedings of the 4th International Conference Moodle. si.*

Sarodnick, F., & Brau, H. (2015). *Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung*. Hans Huber.

SurveyMonkey. (2019). Stichprobenrechner: Stichprobengrößen verstehen. Abgerufen 21. Januar 2019, von <https://www.surveymonkey.de/mp/sample-size-calculator/>

Tee, S. S., Wook, T., & Zainudin, S. (2013). User testing for moodle application. *International Journal of Software Engineering and Its Applications*, 7(5), 243–252.

Van Nuland, S. E., & Rogers, K. A. (2016). The anatomy of E-Learning tools: Does software usability influence learning outcomes? *Anatomical sciences education*, 9(4), 378–390.

Vredenburg, K., Mao, J.-Y., Smith, P. W., & Carey, T. (2002). A survey of user-centered design practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (S. 471–478). ACM.

Younes, A. M., Stewart, J., & Kyriakidou, N. (2013). MTDPNA in non-oil international organisations in Libyan post crisis. *Business and Management Research*, 2(1), 18.

Zemsky, R., & Massy, W. F. (2004). Thwarted innovation. *What happened to e-learning and why, A final report for the Weather station Project of the Learning Alliance at the University of Pennsylvania in cooperation with the Thomson Corporation, Pennsylvania.*