

Zbl Arbeitsmed 2020 · 70:256–263
<https://doi.org/10.1007/s40664-020-00399-2>
Online publiziert: 28. April 2020
© Springer-Verlag GmbH 2020



Stefanie Schmidt · Annemarie Minow · Irina Böckelmann

Bereich Arbeitsmedizin, Medizinische Fakultät, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Deutschland

Einsatz und Aussagekraft etablierter quantitativer Usability-Fragebögen in einem User-Test

In Zeiten des digitalen Fortschritts sind Mensch-Technik-Interaktionen der betrieblichen Praxis Bestandteil vieler wissenschaftlicher Diskurse, insbesondere im Kontext der Arbeitssicherheit, ergonomischen Gestaltung und Gesundheit. Für das Gelingen der Interaktion gilt es, neben den technischen Herausforderungen vor allem die nutzerzentrierten Aspekte bei dem Einsatz neuer digitaler Assistenzsysteme stärker zu fokussieren. Eine frühzeitige Analyse der Gebrauchstauglichkeit digitaler Technologien unter Einbezug von Mitarbeitern bildet eine wesentliche Grundlage für den erfolgreichen und gesundheitsunbedenklichen Einsatz im Betrieb. Aus diesem Grund widmete sich die vorliegende Studie der Auseinandersetzung mit den Möglichkeiten, die Gebrauchstauglichkeit verschiedener digitaler Assistenzfunktionen zu untersuchen. Dabei rückten die in wissenschaftlichen Studien [1–3] verbreiteten quantitativen Fragebögen der System Usability Scale (SUS), der DIN EN ISO 9241/110 und des Technologieakzeptanzmodells (TAM) in den Fokus. Das wissenschaftliche Interesse gilt insbesondere der SUS und der damit einhergehenden augenscheinlich eindimensionalen Operationalisierung des Konstruktes Gebrauchstauglichkeit.

Hintergrund

Im Zentrum der heutigen Arbeitswelt stehen sich wandelnde Anforderungen an die Arbeitsorganisation, Mitarbeiterkompetenzen und Wertschöpfungspro-

zesse, denen Unternehmen begegnen müssen. Chancen hierfür bieten sich in der Nutzung digitaler Assistenzsysteme, wie beispielsweise Tablets, Smartwatches oder Datenbrillen, die bereits vermehrt im Produktionskontext, in Bereichen der Montage, der Instandhaltung und der Kommissionierung zum Einsatz kommen [4]. Entscheidend für den Erfolg oder Misserfolg dieser Systeme im betrieblichen Kontext ist auch die Akzeptanz durch den Benutzer, d. h. die Bereitschaft Beschäftigter, die innovativen Technologien für ihre Arbeitsaufgaben anzuwenden [5]. Damit die Benutzerakzeptanz gegeben ist und somit die Unterstützungspotenziale digitaler Assistenzsysteme ausgeschöpft werden können, gilt es, die entstehende Mensch-Technik-Interaktion von Beginn an nachhaltig und gesundheitsgerecht zu fördern. Dabei ist davon abzusehen, innovative Systeme per *Helikopter* über Arbeitsprozesse und Mitarbeiter zu streuen. Für das Gelingen der Mensch-Technik-Interaktion sollten Beschäftigte als Nutzer bereits in den Forschungs- und Entwicklungsprozess der Systeme aktiv eingebunden werden, da nur sie über das nötige Expertenwissen zu einzelnen Arbeitsbedingungen und -abläufen verfügen und so bereits optimal Einfluss auf Gestaltung und Design digitaler Assistenzsysteme nehmen können [6]. Aber auch während und nach der Einführung moderner Assistenzsysteme in die betriebliche Praxis gilt es, alle Beteiligten, die einen täglichen Umgang mit den neuen Technologien pflegen, ausrei-

chend für den technologischen Wandel zu sensibilisieren und nutzerbezogene Aspekte angemessen zu berücksichtigen. Zusammenfassend kann also gesagt werden, dass „ein praktischer Mehrwert in der Mensch-Maschine-Interaktion [...] dann entsteht, wenn der Mitarbeiter im Arbeitsprozess *mitdenkt*; wenn er mit seiner technischen Umgebung *vertraut* ist, diese *akzeptiert* und *nutzt*“ [7].

Ein Faktor, der für die Forschungs- und Entwicklungsprozesse innovativer Systeme eine wesentliche Rolle spielt, ist die Gebrauchstauglichkeit („Usability“). Laut dem internationalen Standard der DIN EN ISO 9241, die Richtlinien zur Gestaltung der *Ergonomie der Mensch-System-Interaktion* formuliert, wird in Teil 11 die Gebrauchstauglichkeit als das „Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden können, um festgelegte Ziele effektiv, effizient und zufriedenstellend zu erreichen“ beschrieben [8]. Dabei werden die Begriffe „effektiv“, „effizient“ und „zufriedenstellend“ folgendermaßen definiert [8]:

- Effektivität: „Genauigkeit und Vollständigkeit, mit denen Benutzer bestimmte Ziele erreichen“;
- Effizienz: „im Verhältnis zu den erreichten Ergebnissen eingesetzte Ressourcen“;
- Zufriedenstellung: „Wahrnehmungen und Reaktionen einer Person, die aus der Benutzung eines Systems, eines

Produkts oder einer Dienstleistung resultieren“.

Zur exakten Beurteilung der Gebrauchstauglichkeit sind eine Reihe von quantitativen und qualitativen Forschungsmethoden vorhanden.

Um hinsichtlich der Möglichkeiten der Bewertung der Gebrauchstauglichkeit digitaler Assistenztechnologien zu einer konkreten Einschätzung zu gelangen, wurde eine systematische Recherche (Stichtag 26.09.2019) durchgeführt. Diese Literaturrecherche erfolgte u. a. in den Datenbanken „Web of Science“, „Scopus“, „IEEE“ und „dblp“. Für die gezielte Suche wurden relevante Schlüsselbegriffe, zum Beispiel „Usability“, „Augmented Reality“ und „Industry“ verwendet und durch das Hinzufügen von Booleschen Operatoren zu Suchstrings miteinander verknüpft. Die systematische Recherche zeigte eine erhöhte Anwendung quantitativer Verfahren in Studien. Auffallend dabei war die häufige Verwendung der SUS zur Evaluierung der Gebrauchstauglichkeit. Weiterhin wurde deutlich, dass in Bezug auf die Bewertung der Gebrauchstauglichkeit in der Literatur immer wieder Stellung zu den in Teil 110 der DIN EN ISO 9241 aufgeführten Grundsätzen der Dialoggestaltung sowie zu dem TAM genommen wird. Zur besseren Einordnung des wissenschaftlichen Interesses dieser Studie werden die benannten Verfahren kurz erläutert.

System Usability Scale

Die SUS, die es durch ihre quantitative Methodik ermöglicht, technische Geräte und Systeme jeder Art bezüglich ihrer Gebrauchstauglichkeit zu bewerten, wurde 1986 von Brooke begründet [9]. Vor dem Hintergrund der praktischen Durchführbarkeit wurde ein Fragebogen entwickelt, der schnell und einfach anzuwenden ist. Dieser enthält 10 Aussagen, welche zur Hälfte positiv und zur Hälfte negativ formuliert sind. Die Antwortoptionen reichen auf einer fünfstufigen Likert-Skala von „vollständige Ablehnung“ bis „vollständige Zustimmung“ [9].

Nach der Durchführung der Bewertung der Gebrauchstauglichkeit mittels SUS-Fragebogens werden die Antwort-

möglichkeiten mit einem Wert von 0 bis 4 kodiert. So wird bei voller Zustimmung einer positiv formulierten Antwortmöglichkeit der Wert 4 und bei voller Zustimmung und einer negativ formulierten Antwortmöglichkeit der Wert 0 vergeben. Wird einer Aussage nicht zugestimmt, ist eine umgekehrte Bewertung vorzunehmen. Anschließend werden die Einzelwerte summiert und mit dem Wert 2,5 multipliziert. Die sich daraus ergebende Endsumme wird als SUS-Score bezeichnet und ergibt einen Globalwert zwischen 0 und 100 [9]. Dieser Wert ist als Prozentsatz zu verstehen, welcher dementsprechend die Gebrauchstauglichkeit von technischen Systemen definiert. Hierfür wurde von Bangor und den Koautoren folgende Kategorisierung vorgenommen [10]:

- 100 % entsprechen einem „perfekten“ System ohne Usability-Probleme,
- >80 % entsprechen einer „guten“ bis „exzellenten“ Usability,
- 60–80 % entsprechen einer „grenzwertig“ bis „guten“ Usability und
- <60 % entsprechen einem System mit „erheblichen Usability-Problemen“.

ISONORM 9241/110-Fragebogen

Ein sich an den im Teil 110 der DIN EN ISO 9241 verankerten Grundsätzen der Dialoggestaltung orientierender Fragebogen wurde bereits Anfang der 1990er von Prümper und Anft entwickelt [11]. Mithilfe des Fragebogens kann die Gebrauchstauglichkeit innovativer Systeme, beginnend vom Prototypen bis hin zur finalen Version, evaluiert werden [12]. Dabei wird in der Methodik Bezug zu den sieben verankerten „Grundsätzen der Dialoggestaltung“ der DIN EN ISO 9241-110 genommen, die wie folgt lauten [13]:

- Aufgabenangemessenheit,
- Selbstbeschreibungsfähigkeit,
- Steuerbarkeit,
- Erwartungskonformität,
- Fehlertoleranz,
- Individualisierbarkeit und
- Lernförderlichkeit.

Entsprechend der Dialogprinzipien sind in der Langfassung des Fragebogens 35 und in der Kurzform, die auch in der vorliegenden Studie genutzt wurde, 21 Aus-

sagen enthalten. Zur Beantwortung dieser Aussagen stehen Optionen in Form einer siebenstufigen Likert-Skala von „gut“ bis „schlecht“ zur Verfügung, welche mit den Werten von –3 bis +3 kodiert werden. Für die anschließende Auswertung bieten sich 3 Möglichkeiten. So können Mittelwerte bzw. Summen-Scores für einzelne Fragen oder Dialogprinzipien oder für den gesamten Fragebogen gebildet werden. Letzteres empfiehlt sich für einen schnellen Vergleich zwischen mehreren Assistenzsystemen. Für alle 3 Möglichkeiten gilt aufgrund der Wertekodierung der Wert 1 als Mindestmaß für eine gute Gebrauchstauglichkeit der Systeme [12].

Technologieakzeptanzmodell

Auch das 1989 von Davis begründete Akzeptanzmodell befasst sich inhaltlich mit der Gebrauchstauglichkeit von Technologieinnovationen, in dem vor allem die Variablen „perceived usefulness“ (wahrgenommene Nützlichkeit) und „perceived ease of use“ (wahrgenommene Benutzerfreundlichkeit) die Nutzung einer Technologie bestimmen [14]. Auf Grundlage dieser inhaltlichen Ausführungen des TAM wurden 2 Teilskalen entwickelt. Jede dieser Teilskalen enthält vier entsprechende Aussagen, die in einer Likert-Skala mit 7 Antwortoptionen von „stimme gar nicht zu“ bis hin zu „stimme voll und ganz zu“ bewertet werden können. Alle Aussagen sind positiv formuliert und erhalten bei voller Zustimmung den Wert 7. Ähnlich dem ISONORM-Fragebogen ist es möglich, mithilfe von Mittelwerten bzw. einem Summen-Score innerhalb der einzelnen Teilskalen oder für beide Skalen insgesamt eine Auswertung vorzunehmen.

Wissenschaftliches Interesse

Ausgehend von den theoretischen Grundlagen ist eindeutig, dass auf der Suche nach einem praktikablen Instrument zur Bewertung der Gebrauchstauglichkeit von digitalen Systemen bzw. Funktionen im industriellen Kontext mit der SUS ein Verfahren entwickelt wurde, das einfach und in kurzer Zeit durchgeführt werden kann. Jedoch wird dieses Verfahren selbst nicht nur als „quick“

(schnell), sondern als „dirty“ bezeichnet [9], was in diesem Zusammenhang mit den Worten „unscharf, unpräzise“ assoziiert werden kann. Diese Aussage spiegelt die Grenze des Evaluierungsinstrumentes und somit die Problemlage für die Praxis wider. Denn während sich sowohl mit dem ISONORM-Fragebogen als auch mittels der Teilskalen des TAM und der damit verbundenen mehrdimensionalen Operationalisierung der Gebrauchstauglichkeit konkrete Hinweise zur Gestaltung sowie zur Nützlichkeit und Benutzerfreundlichkeit von digitalen Assistenzsystemen bzw. -funktionen abbilden lassen, scheint dies durch die SUS nicht gegeben. Es ist zwar davon auszugehen, dass ein niedriger SUS-Score ein Hinweis für das Vorhandensein von Problemen ist, diese jedoch aufgrund der Erhebung eines Globalwertes nicht konkretisiert werden können. Dies würde bedeuten, dass eine differenzierte Betrachtung der Gebrauchstauglichkeit der zu untersuchenden digitalen Technologie nicht möglich ist.

Vor diesem Hintergrund wurde die vorliegende Studie initiiert, die Aussagekraft des SUS-Scores hinsichtlich der Gebrauchstauglichkeit von digitalen Assistenzen zu beurteilen. Aufgrund der Tatsache, dass alle drei oben genannten Verfahren das Konstrukt Gebrauchstauglichkeit abbilden, wird der hypothetischen Annahme nachgegangen, dass Zusammenhänge zwischen dem SUS-Score und den Dimensionen der anderen beiden Evaluierungsmethoden bestehen. Dazu wurden die Beziehungen zwischen dem Globalwert der SUS und den in der DIN EN ISO 9241 und dem TAM verankerten Aspekten der Dialoggestaltung sowie der wahrgenommenen Nützlichkeit und der wahrgenommenen Bedienbarkeit analysiert. Forschungsleitend ergaben sie folgende primäre Fragestellungen:

- Bestehen Zusammenhänge zwischen dem Globalwert der SUS und den in der DIN EN ISO 9241-110 enthaltenen Dimensionen „Aufgabenangemessenheit“, „Selbstbeschreibungsfähigkeit“, „Steuerbarkeit“, „Erwartungskonformität“, „Fehlertoleranz“, „Individualisierbarkeit“ und „Lernförderlichkeit“?

Zbl Arbeitsmed 2020 · 70:256–263 <https://doi.org/10.1007/s40664-020-00399-2>
© Springer-Verlag GmbH 2020

S. Schmidt · A. Minow · I. Böckelmann

Einsatz und Aussagekraft etablierter quantitativer Usability-Fragebögen in einem User-Test

Zusammenfassung

Hintergrund. Analysen der Gebrauchstauglichkeit (Usability) digitaler Assistenzsysteme sind zur Förderung der Nutzerakzeptanz neuer Technologien unabdingbar. Für diese Analysen sind verschiedene Methoden vorhanden, von denen in diesem Beitrag drei etablierte quantitative Verfahren an einem praktischen Beispiel vorgestellt und deren Aussagekraft näher beleuchtet werden.

Ziel der Arbeit. Ziel war es, die in Studien häufig verwendete System Usability Scale (SUS) als hinreichendes Maß zur Bewertung der Gebrauchstauglichkeit zu untersuchen und die Aussagekraft des SUS-Scores zu beurteilen. Hierfür wurden Zusammenhänge zu den Grundsätzen der Dialoggestaltung nach DIN EN ISO 9241-110 und dem Technologieakzeptanzmodell (TAM) untersucht.

Material und Methoden. Es wurde ein User-Test durchgeführt, in dem zwei digitale Assistenzfunktionen, basierend auf Augmented Reality (AR) und Pick-to-light (PtI) von 25 Probanden getestet und hinsichtlich ihrer

Gebrauchstauglichkeit durch die 3 Verfahren bewertet wurden. Die gewonnenen Daten wurden durch Mittelwertvergleiche und Korrelationsanalysen ausgewertet und interpretiert.

Ergebnisse. Insgesamt zeigen sich durch die Mittelwertvergleiche in allen 3 Verfahren ähnliche Aussagen. Mithilfe der Korrelationsanalysen konnten mittlere bis hohe positive Zusammenhänge zwischen den Usability-Evaluierungsmethoden aufgezeigt werden.

Diskussion. Die Ergebnisse sprechen für die SUS als geeignetes Maß zur umfangreichen Evaluierung der Gebrauchstauglichkeit digitaler Assistenzsysteme. So sind Bezüge zu den Grundsätzen der Dialoggestaltung sowie zur Nützlichkeit und Benutzerfreundlichkeit einer Technologie erkennbar.

Schlüsselwörter

Digitale Assistenzsysteme · Gebrauchstauglichkeit · System Usability Scale · DIN EN ISO 9241 · Technologieakzeptanzmodell

Use and significance of established quantitative usability questionnaires in a user test

Abstract

Background. Analyses of the usability of digital assistance systems are indispensable for promoting user acceptance of new technologies. Various methods are available for these analyses. In this article three established quantitative procedures are presented with a practical example and their significance is examined in more detail.

Objective. The aim was to examine the system usability scale (SUS), which is often used in studies, as an adequate measure for evaluating usability and to assess the validity of the SUS score. For this purpose, associations with the principles of dialogue design according to DIN EN ISO 9241-110 and the technology acceptance model (TAM) were investigated.

Material and methods. A user test was carried out in which two digital assistance functions based on augmented reality (AR) and pick-to-light (PtI) were tested by 25 test persons and their usability was evaluated

by the 3 methods. The data obtained were evaluated and interpreted by means of mean value comparisons and correlation analyses.

Results. Overall, the mean value comparisons in all three procedures showed similar statements. With the help of correlation analyses medium to high positive correlations between the usability evaluation methods could be shown.

Conclusion. The results suggest that SUS is an appropriate measure for the comprehensive evaluation of the usability of digital assistance systems. References to the principles of dialogue design as well as to the usefulness and user-friendliness of a technology can be identified.

Keywords

Digital assistance systems · Usability · System usability scale · ISO 9241 · Technology acceptance model

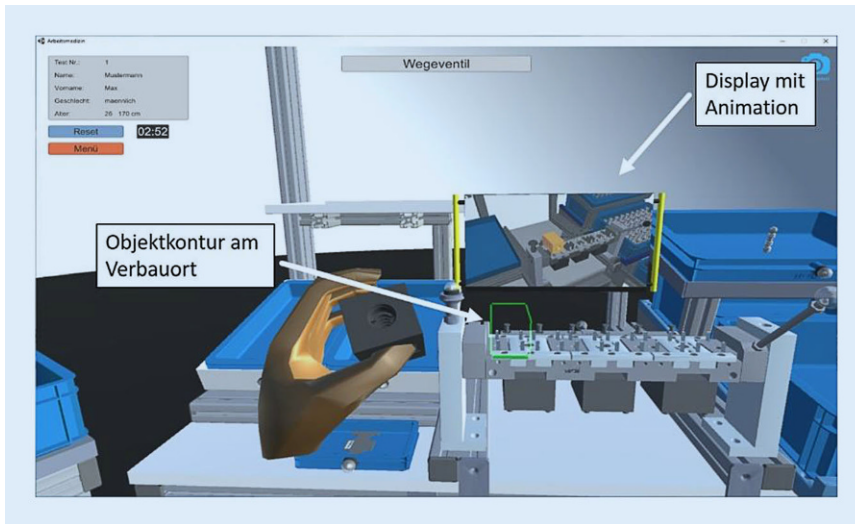


Abb. 1 ▲ Montageszenario unter Verwendung von Augmented-Reality(AR)-Konturen (Foto: LIVING-SOLIDS GmbH)

- Bestehen Zusammenhänge zwischen dem Globalwert der SUS und den in Teilskalen des TAM verankerten Dimensionen „wahrgenommene Nützlichkeit“ und „wahrgenommene Benutzerfreundlichkeit“?

Eine sekundäre Fragestellung lautet:

- Unterscheiden sich die in dieser Studie betrachteten Assistenzfunktionen in der Bewertung der Gebrauchstauglichkeit?

Die Beantwortung der sekundären Fragestellung erfolgt in den Ergebnissen zuerst, um eine Grundlage für die Beantwortung der primären Fragestellungen zu liefern.

Studiendesign und Untersuchungsmethoden

Zur Beantwortung der forschungsleitenden Fragen wurde vom Bereich Arbeitsmedizin der Medizinischen Fakultät der Otto-von-Guericke-Universität Magdeburg eine Studie auf Basis eines Crossover-Designs konzipiert. Hierfür wurde ein experimenteller User-Test im Rahmen des vom BMBF-geförderten Verbundvorhabens (FKZ: 03ZZ0441E) „3D-Montageassistent“ mit der LIVING-SOLIDS GmbH durchgeführt. Die Aufgabe für die Studienteilnehmer bestand darin, zwei virtuell simulierte digitale Assistenzfunktionen, basierend auf Augmented Reality(AR)- und Pick-to-

light(Ptl)-Funktionen, hinsichtlich ihrer Gebrauchstauglichkeit zu testen. Zur Simulation wurde eine HTC Vive Virtual Reality(VR)-Brille (HTC Corporation, Taoyuan, Taiwan) (Head-Mounted-Display; HMD) mit 2 Basisstationen zur Erfassung von Position und Orientierung der Brille im Raum genutzt. Zur Registrierung der für die Montageaufgaben relevanten Nutzereingaben (z. B. Greifen von Bauteilen) wurde die VR-Brille mit einem Leap-Motion-Sensor ausgestattet, der die Hände erkennt und eine intuitive Interaktion mit virtuellen Objekten ermöglicht. Für die Simulation der Arbeitsumgebung wurde ein Raum gestaltet und darin ein realer Arbeitsplatz mit Hilfe von 3-D-Modellen der Einzelobjekte nachgebildet. Die für die Arbeitsaufgabe nötigen Behälter, Vorrichtungen und Bauteile wurden als 3-D-Objekte in die Szene integriert und soweit erforderlich bedienbar gestaltet.

Ausgehend von einer ausführlichen Aufklärung der Teilnehmer zum Studienablauf sowie einer positiven Einwilligung zur Teilnahme ihrerseits, wurden die Probanden vor Beginn des Untersuchungsablaufes nach ihren soziodemografischen Daten und bisherigen Erfahrungen im Umgang mit digitalen Assistenzsystemen und VR-Brillen befragt. Da keiner der Studienteilnehmer vorher in der Montage tätig war, erhielten alle Probanden zunächst eine Trainingseinheit,

um den Montageablauf des Wegeventils zu üben.

Die Wegeventilmontage wurden anschließend von jedem Probanden in zwei randomisierten Durchgängen unter Nutzung der jeweiligen Assistenzfunktion (Augmented Reality und Pick-to-light) durchgeführt. Dazu wurden die Teilnehmer aufgefordert, in jedem Durchgang ein Wegeventil, drei geometrische Figuren und nochmals das gleiche Wegeventil zu montieren. Um mögliche Lerneffekte zu vermeiden, fand auch die Randomisierung von Varianten der geometrischen Figuren statt.

Im Montageszenario wurden bei der AR-basierten Assistenz simulierte Konturen eingesetzt (▣ Abb. 1), um den Probanden den Verbauort der Einzelteile am Werkstück anzuzeigen. Bei dem Pick-to-light-System (▣ Abb. 2) wiederum wurde den Probanden durch ein grün aufleuchtendes Lämpchen angezeigt, aus welcher Kiste das nächste zu verbauende Teil gegriffen werden muss.

Am Ende der Montageabläufe wurden die jeweiligen Assistenzfunktionen hinsichtlich ihrer Gebrauchstauglichkeit mittels der SUS, dem DIN EN ISO 9241-Fragebogen und den beiden Teilskalen des TAM bewertet. Die so gewonnenen intervallskalierten Daten wurden zunächst mit dem Shapiro-Wilk-Test auf Normalverteilung überprüft. Der Großteil der Variablen war nicht normalverteilt. Aus diesem Grund wurden in der statistischen Auswertung nicht-parametrische Tests verwendet. Die im Anschluss erfolgten Mittelwertvergleiche unter Verwendung des Wilcoxon-Vorzeichen-Rang-Tests lieferten konkrete Hinweise zur bevorzugten Assistenzfunktion. Um orientiert an der Zielstellung dieser Studie zu arbeiten, wurden zudem Korrelationsanalysen nach Spearman innerhalb der jeweiligen Assistenzfunktion zwischen der SUS und dem ISONORM-Fragebogen sowie dem TAM vorgenommen. Bei allen Berechnungen wurde ein Signifikanzniveau von 5% zugrunde gelegt.

In den nun folgenden Ausführungen werden Mittelwerte mit MW, Standardabweichungen mit SD, Median mit MD und Minimum und Maximum mit Min und Max abgekürzt. Zur statistischen

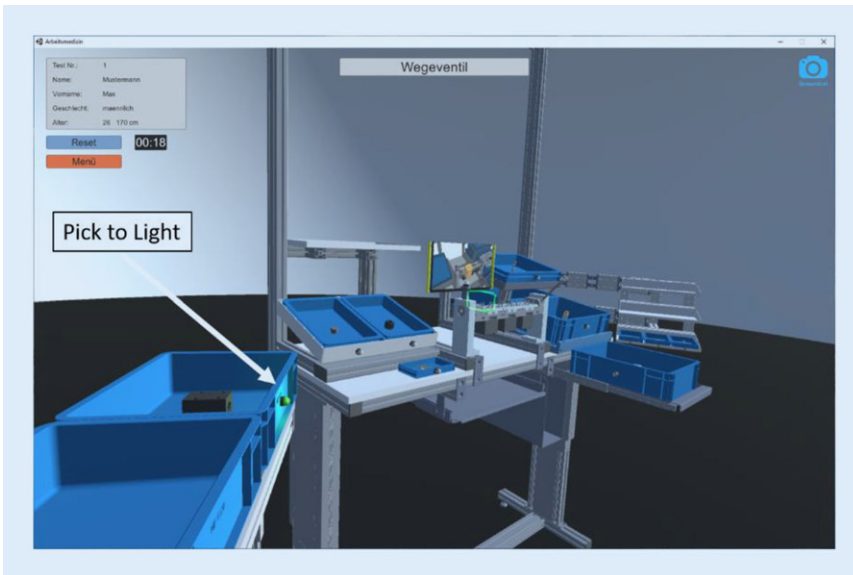


Abb. 2 ▲ Montageszenario unter Verwendung von Pick-to-light (Foto: LIVINGSOLIDS GmbH)

Auswertung wurde SPSS Statistics 26 (IBM Corporation, Armonk, NY, USA) genutzt.

Ergebnisse

Für die Studie konnten 25 gesunde Teilnehmer gewonnen werden, davon waren 12 (48 %) weiblichen und 13 (52 %) männlichen Geschlechts. Bei den Probanden handelte es sich vorwiegend um Studierende und Hochschulabsolventen aus nichttechnischen Fachrichtungen im Alter von 21–59 Jahren (Durchschnittsalter: $26,8 \pm 8,93$ Jahre). Außerdem berichteten 52 % der Teilnehmer ($n = 13$) von keinerlei Erfahrungen mit VR-Brillen, während 48 % ($n = 12$) diese bereits für andere Forschungszwecke oder im Freizeitbereich ausprobiert haben.

Unterschiede in der Beurteilung zweier Assistenzsysteme

Anhand der durchgeführten Mittelwertvergleiche können ähnliche Ergebnisse in der Bewertung der Gebrauchstauglichkeit der digitalen Assistenzsysteme in allen Evaluierungsmethoden festgestellt werden. Beginnend mit den Ergebnissen der SUS zeigt sich im Vergleich eine signifikant bessere Bewertung der Pick-to-light-Funktion als für die Augmented-Reality-basierte (Pick-to-light $78,7 \pm 11,99$ Punkte und Aug-

mented Reality $74,6 \pm 13,84$ Punkte; $p_{\text{Wilcoxon}} = 0,031$). Dennoch kann die Gebrauchstauglichkeit beider Assistenzsysteme nach der Einordnung von [10] als „gut“ eingestuft werden.

Anders als bei der SUS weisen die Mittelwertvergleiche beider Assistenzfunktionen in den 7 Dialogprinzipien des ISONORM-Fragebogens keine statistisch signifikanten Unterschiede ($p_{\text{Wilcoxon}} > 0,05$; geringe Effektstärken) auf. Eine eindeutige Aussage darüber, welche Assistenzfunktion von den Probanden bevorzugt wird, kann anhand dieser Ergebnisse nicht abgeleitet werden (Tab. 1). Jedoch ist bei beiden Funktionen aufgrund der positiven Mittelwerte in allen Dimensionen des Fragebogens von einem vorhandenen Mindestmaß an guter Gebrauchstauglichkeit auszugehen. In den Bewertungen der 2 Teilskalen „wahrgenommene Nützlichkeit“ und „wahrgenommene Benutzerfreundlichkeit“, die durch das TAM definiert werden, zeichnen sich keine statistisch bedeutsamen Unterschiede bei geringen Effektstärken zwischen den Assistenzfunktionen ab.

Zusammenfassend ist zu sagen, dass die Bewertungen hinsichtlich der Gebrauchstauglichkeit der Assistenzfunktionen für alle drei quantitativen Evaluierungsmethoden ähnlich gut sind und dass, obwohl eine tendenzielle Präferenz der Pick-to-light-Funktionen vorliegt,

die Ergebnisse insgesamt auf wenig Optimierungsbedarf der jeweiligen Assistenz verweisen.

Zusammenhänge zwischen den Beurteilungsverfahren

In der Tab. 2 sind die Ergebnisse der Korrelationsanalysen zwischen dem SUS-Score und den Mittelwerten der Grundsätze der Dialoggestaltung sowie den zwei Teilskalen des TAM für die jeweilige digitale Assistenzfunktion (AR und Ptl) dargestellt. Beginnend mit den AR-Konturen können zwischen dem Wert der SUS und 5 der 7 Dialogprinzipien des DIN EN ISO 9241 mittlere bis hohe positive Korrelationen mit einer geringen Irrtumswahrscheinlichkeit ($p \leq 0,05$ bzw. $p \leq 0,01$) nachgewiesen werden. Lediglich die ISONORM-Dimensionen der „Steuerbarkeit“ und der „Fehlertoleranz“ weisen keine signifikante Korrelation zur SUS auf. Sehr bedeutsam hingegen, mit einem Korrelationskoeffizienten von 0,542, ist der positive Zusammenhang zwischen dem Globalwert der SUS und dem Prinzip der „Erwartungskonformität“. Die SUS korreliert hoch positiv mit den Teilskalen des Technologieakzeptanzmodells. Das heißt, steigt bzw. sinkt der ermittelte Wert der SUS, kann dies in gleicher Weise für die Werte der „wahrgenommenen Nützlichkeit“ sowie der „wahrgenommenen Benutzerfreundlichkeit“ festgestellt werden. Diese Zusammenhänge sind statistisch höchst bedeutsam ($p \leq 0,001$).

In der Anwendung der Pick-to-light-Funktion zeichnen sich signifikante mittlere bis hohe positive Korrelationen zwischen dem ermittelten Wert der SUS und den ISONORM-Dimensionen „Aufgabenangemessenheit“, „Steuerbarkeit“, „Erwartungskonformität“ und „Lernförderlichkeit“ ab. Mit einem Korrelationskoeffizienten von 0,635 fällt dabei der Zusammenhang zwischen dem SUS-Wert und der „Aufgabenangemessenheit“ am stärksten aus. Für die Dimensionen „Selbstbeschreibungsfähigkeit“, „Fehlertoleranz“ und „Individualisierbarkeit“ konnte kein statistisch signifikanter Zusammenhang zur SUS hergestellt werden.

Tab. 1 Vergleich beider Assistenzfunktionen mittels ISONORM 9241-Grundsätzen und Technologieakzeptanzmodell

Variable		Augmented Reality	Pick-to-light	pWilcoxon	Effektstärke
		MW ± SD MD (Min–Max)			
ISONORM 9241 – Grundsätze	Aufgabenangemessenheit	1,4 ± 1,19 1,7 (–1–3)	1,6 ± 0,95 1,7 (–0,3–3)	0,332	0,14
	Selbstbeschreibungsfähigkeit	1,1 ± 1,28 1 (–1,7–3)	1,0 ± 1,30 1 (–2,3–3)	0,541	0,09
	Steuerbarkeit	0,8 ± 1,44 0,5 (–2–3)	0,7 ± 1,26 0,5 (–1,5–3)	0,529	0,09
	Erwartungskonformität	2,1 ± 0,65 2 (0,7–3)	2,0 ± 0,79 2 (0,3–3)	0,753	0,05
	Fehlertoleranz	0,6 ± 1,24 0,3 (–1,3–3)	0,8 ± 0,94 0,7 (–0,7–2,7)	0,671	0,06
	Individualisierbarkeit	1,1 ± 1,00 1 (–1,5–2,5)	0,6 ± 0,99 0,5 (–1–3)	0,123	0,22
	Lernförderlichkeit	2,0 ± 0,87 2 (–0,7–3)	2,2 ± 0,76 2,3 (0–3)	0,105	0,23
Technologieakzeptanzmodell	Wahrgenommene Nützlichkeit	4,5 ± 1,74 5 (1–7)	4,9 ± 1,72 5 (1–7)	0,111	0,23
	Wahrgenommene Benutzerfreundlichkeit	5,4 ± 1,25 5,5 (2,8–7)	5,7 ± 1,08 5,8 (3–7)	0,230	0,17

Die mögliche Spannweite des ISONORM-Fragebogens liegt bei –3 bis 3, Die mögliche Spannweite in den Items des Technologieakzeptanzmodells liegt bei 1–7

MW Mittelwert, SD Standardabweichungen, MD Median, Min Minimum, Max Maximum

Tab. 2 Korrelation zwischen „System Usability Scale“ und ISONORM 9241-Grundsätzen bzw. Technologieakzeptanzmodell für die Augmented-Reality- und Pick-to-light-Anwendung

		Augmented-Reality-Anwendung	Pick-to-light-Anwendung
		Roh („System Usability Scale“)	
ISONORM 9241 – Grundsätze	Aufgabenangemessenheit	0,487*	0,635**
	Selbstbeschreibungsfähigkeit	0,428*	n. s.
	Steuerbarkeit	n. s.	0,566**
	Erwartungskonformität	0,542**	0,471*
	Fehlertoleranz	n. s.	n. s.
	Individualisierbarkeit	0,405*	n. s.
	Lernförderlichkeit	0,559**	0,510**
Technologieakzeptanzmodell	Wahrgenommene Nützlichkeit	0,651***	0,643**
	Wahrgenommene Benutzerfreundlichkeit	0,798***	0,471*

n. s. nicht signifikant

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

Im Gegensatz dazu korrelieren die Mittelwerte der 2 Teilskalen des TAM mittel bis hoch positiv mit dem ermittelten Wert der SUS. Das bedeutet, je besser die Bewertungen von digitalen Assistenzen durch die SUS ausfallen, desto höher auch die „wahrgenommene Nützlichkeit“ und die „wahrgenommene Benutzerfreundlichkeit“ der Systeme oder Funktionen. Noch stärker ist dabei der Zusammenhang zur „wahrgenommenen Nützlichkeit“ ($r = 0,643$).

In beiden Fällen liegt eine statistische Signifikanz vor.

Diskussion

Ausgehend von der hypothetischen Annahme bestehender Zusammenhänge zwischen dem von der SUS ermittelten Score und den Bewertungen der einzelnen Dimensionen des DIN EN ISO 9241-Fragebogens sowie der Teilskalen des TAM wurden vergleichende Analysen durchgeführt. Bereits der Vergleich der

Mittelwerte und die resultierenden konsistenten Ergebnisse festigten die Vermutung tatsächlicher Beziehungen der Usability-Fragebögen untereinander. So wiesen alle 3 Evaluierungsmethoden auf eine gute Gebrauchstauglichkeit bei beiden hier vorgestellten Assistenzfunktionen hin.

Zur detaillierten Betrachtung der Zusammenhänge und der damit einhergehenden Beantwortung der forschungsleitenden Fragen wurden Korrelationsanalysen durchgeführt. Die Ergebnisse

dieser Analysen weisen bei beiden Assistenzfunktionen positive Abhängigkeiten zwischen dem SUS-Score und den Dialogprinzipien des ISONORM-Fragebogens sowie zwischen dem SUS-Score und den TAM-Dimensionen auf.

Die Resultate der Studie bestätigen die hypothetische Annahme bestehender Zusammenhänge und deuten auf die Konkretisierung des Globalwertes der SUS durch einzelne Prinzipien, der in der DIN EN ISO 9241-110 und des TAM verankerten Aspekten einer digitalen Assistenzfunktion. Die Ergebnisse zeugen aber auch von kontextueller Gebundenheit; so scheinen vor allem Zusammenhänge zu den Dialogprinzipien „Steuerbarkeit“, „Selbstbeschreibungsfähigkeit“, „Fehlertoleranz“ und „Individualisierbarkeit“ aufgrund unterschiedlicher Signifikanzen technologieabhängig zu sein. Diese Annahme führt gleichzeitig zu den Limitationen, welche die vorliegende Studie trotz guter methodischer Herangehensweise aufweist.

So ist in Folgestudien zu prüfen, ob die Ergebnisse auf andere Anwendungsfälle und Stichproben übertragbar sind. Hierfür sollte u. a. die Anwendung weiterer Assistenzfunktionen erfolgen. Aufgrund der zeitlichen Begrenzung des User-Tests wurde sich lediglich an einem Anwendungsszenario eines Projektpartners des Verbundvorhabens orientiert. Hierfür waren die unter AR-Brillen simulierte Assistenzfunktionen in Form von AR-Konturen und Pick-to-light die vielversprechendsten Lösungen. Grundsätzlich können unter VR-Brillen aber auch andere Assistenzfunktionalitäten simuliert und getestet werden. Dabei gilt es, die Probanden darauf hinzuweisen, dass sie ausschließlich die simulierte Assistenzfunktion und nicht die Hardware (AR-Brille) bewerten, da dies zu Verzerrungen in den Ergebnissen führen würde. Außerdem sollte für die Prüfung der Übertragbarkeit der Ergebnisse in Folgestudien eine andere Zusammensetzung der Stichprobe gewählt werden. Da es sich bei dem User-Test ausschließlich um junge Probanden ohne Montageerfahrung handelte, könnte in weiteren Experimenten der Fokus auf älteren Arbeitnehmern oder gelernten Montagearbeitern liegen. Fol-

gestudien im Rahmen des Projektes bei einem Anwendungspartner unter realen Arbeitsbedingungen im Unternehmen sind geplant.

Zusammenfassend kann jedoch gesagt werden, dass laut der vorliegenden Studienergebnisse mithilfe des SUS-Scores die Möglichkeit besteht, weit mehr zu interpretieren als die bloße Unzufriedenheit bzw. Zufriedenheit mit der Gebrauchstauglichkeit einer Technologie durch den Nutzer. So können eindeutige Bezüge zur Gestaltung, der Bedienbarkeit sowie dem Nutzen hergestellt und eine differenzierte Betrachtung der Gebrauchstauglichkeit digitaler Assistenz vorgenommen werden. Bei Bedarf und zur Ergänzung der durch die SUS erzeugten Ergebnisse eignen sich zudem vertiefende qualitative Analysen (Einzelinterviews, Fokusgruppen, Thinking-Aloud-Methode¹), mit denen eine detaillierte Erkundung gegebener Schwachstellen der Technologien möglich ist und so entsprechende Verbesserungsvorschläge generiert werden können.

Fazit für die Praxis

- Die System Usability Scale (SUS) ist ein geeignetes Instrument für eine schnelle, kostengünstige und einfache Analyse der Gebrauchstauglichkeit digitaler Technologien.
- Mithilfe des SUS-Scores kann eine Beurteilung der Gebrauchstauglichkeit vorgenommen werden, die Aufschluss darüber gibt, ob Probleme vorliegen oder nicht.
- Verweist der SUS-Score auf Schwächen der Technologie, so können diese mithilfe der nachgewiesenen Zusammenhänge zu den Gestaltungsprinzipien (DIN EN ISO 9241) sowie der Nützlichkeit und Benutzerfreundlichkeit (Technologieakzeptanzmodell) konkretisiert werden.
- Zur Ergänzung der Ergebnisse durch die quantitative Methodik der SUS wäre es sinnvoll, noch weitere Analy-

¹ Thinking-Aloud-Methode – Methode, bei der alle Gedanken, die dem Probanden bei der Lösung der Aufgabe durch den Kopf gehen, vollständig und ungefiltert verbalisiert werden.

sen in qualitativer Form einzubeziehen.

Korrespondenzadresse



Stefanie Schmidt, B.A.
Bereich Arbeitsmedizin,
Medizinische Fakultät, Otto-
von-Guericke-Universität
Magdeburg
Leipziger Str. 44,
39120 Magdeburg,
Deutschland
stefanie1.schmidt@
med.ovgu.de

Danksagung. Wir bedanken uns beim Bundesministerium für Bildung und Forschung (BMBF) für die Förderung des FuE-Projektes (FKZ: 03Z0441E) der Allianz „3Dsensation“ im Rahmen des Programms „Zwanzig20“. Ein weiterer Dank gilt den Mitarbeiterinnen und Mitarbeitern der LIVING SOLIDS GmbH, die uns technisch unterstützt haben.

Einhaltung ethischer Richtlinien

Interessenkonflikt. S. Schmidt, A. Minow und I. Böckelmann geben an, dass kein Interessenkonflikt besteht.

Alle beschriebenen Untersuchungen am Menschen wurden mit Zustimmung der zuständigen Ethik-Kommission, im Einklang mit nationalem Recht sowie gemäß der Deklaration von Helsinki von 1975 (in der aktuellen, überarbeiteten Fassung) durchgeführt. Von allen beteiligten Probanden liegt eine Einverständniserklärung vor.

Literatur

1. Kretschmer V, Schier A, Mättig B (2018) Digitale Assistenten für die Verpackungslogistik. Studie zur Kognitiven Ergonomie unterschiedlicher digitaler Packanleitungen. Werkstatttech Online 108(7/8):549–554
2. Hartmann M (2008) Usability Untersuchung eines Internetauftrittes nach DIN EN ISO 9241. Am Praxisbeispiel der Firma MAFI Transport-Systeme GmbH. Diplomica, Hamburg
3. Daling L, Abdelrazeq A, Sauerborn C, Hees F (2020) A comparative study of augmented reality assistant tools in assembly. In: Ahram T, Falcão C (Hrsg) Advances in usability and user experience. Proceedings of the AHFE 2019. Springer Nature, Cham, S 755–767
4. Niehaus J (2017) Mobile Assistenzsysteme für Industrie 4.0. Gestaltungsoptionen zwischen Autonomie und Kontrolle. FGW, Düsseldorf
5. Dillon A (2001) User acceptance of information technology. In: Karwowski W (Hrsg) Encyclopedia of human factors and ergonomics. Taylor and Francis, London
6. Minow A, Böckelmann I (2019) Empfehlungen für Forschung und Praxis zu Nutzerakzeptanz, Usability und User Experience bei der Entwicklung und beim Einsatz digitaler Assistenzsysteme. 59.

-
- Jahrestagung der Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e. V.
7. Schenk M, Haase T, Keller A, Berndt D (2016) Herausforderungen der Mensch-Technik-Interaktion für die Gestaltung zukünftiger Arbeitssysteme. In: Schlick C, Wissenschaftliche Gesellschaft für Arbeits- und Betriebsorganisation, Hochschulgruppe Arbeit- und Betriebsorganisation e.V. München (Hrsg) Megatrend Digitalisierung – Potenziale der Arbeits- und Betriebsorganisation. Tagungsband anlässlich des Jubiläumsseminars zum 30-jährigen Bestehen der WGAB e. V., ehemals Hochschulgruppe Arbeits- und Betriebsorganisation HAB e. V. GITO, Berlin, S 131–139
 8. DIN Deutsches Institut für Normung e. V. (2018) DIN EN ISO 9241-11. Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (DIN EN ISO 9241-11:2018)
 9. Brooke J (1996) SUS—A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL (Hrsg) Usability evaluation in industry. Taylor and Francis, London
 10. Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 4(3):114–123
 11. Use Tree – Berliner Kompetenzzentrum für Usability Maßnahmen (2020) Methode: ISONORM 9241/110-Fragebogen. Gestaltungsanforderungen prüfen per Fragebogen. <http://projekt.kke.tu-berlin.de/methode-isonorm-9241110-fragebogen/>. Zugegriffen: 7. Jan. 2020
 12. Use Tree – Berliner Kompetenzzentrum für Usability Maßnahmen (2015) ISONORM 9241/110-Fragebogen. http://projekt.kke.tu-berlin.de/wp-content/uploads/2015/09/Methode_Isonorm-Fragebogen.pdf. Zugegriffen: 10. Dez. 2019
 13. 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: Rödinger K-H (Hrsg) Software-Ergonomie'93. Von der Benutzungsoberfläche zur Arbeitsgestaltung. Vieweg und Teubner, Wiesbaden, S 145–156
 14. Davis FD (1985) A technology acceptance model for empirically testing new end-user information systems: theory and results. Institute of Technology, Massachusetts