

Usability und Technologieakzeptanz – ein Prozessmodell

Der Fakultät Bildungs-, Kultur- und Sozialwissenschaften der
Leuphana Universität Lüneburg zur Erlangung des Grades

**Doktor der Philosophie
– Dr. phil. –**

vorgelegte Dissertation von

Stefan Rothermel

geboren am 09. Juni 1971 in Plochingen

eingereicht am: 19.06.2009

Betreuer und Gutachter: Prof. Dr. Lutz Schumacher

Gutachter: Prof. Dr. Rainer Höger

Gutachter: Prof. Dr. Rainer Guski

Tag der Disputation: 26.10.2010

Inhaltsverzeichnis

Einleitung.....	1
Nutzung und Bedeutung der Informations- und Kommunikationstechnologien.....	1
Usability und Akzeptanzforschung.....	4
Aufbau und Vorgehen dieser Arbeit.....	8
Theoretischer Teil.....	11
Technology Acceptance Model (TAM).....	11
Theorie des überlegten Handelns (Theory of reasoned action).....	11
Struktur des TAM.....	12
Perceived Usefulness.....	14
Perceived Ease of Use.....	14
Entwicklung der PEOU-Skala.....	15
Stand der Forschung zum TAM.....	25
Experimentelle Überprüfungen zum TAM.....	25
Erweiterungen des TAM.....	31
Zusammenfassung.....	44
Kognitionspsychologische Grundlagen.....	45
Mentale Modelle.....	47
Der Modell-Begriff.....	47
Analogiehaftigkeit mentaler Modelle.....	48
Schema und mentale Modelle.....	48
Inhaltsabhängigkeit von schlussfolgerndem Denken.....	50
Mentale Modelle als kognitive Simulationen.....	52
Mentale Modelle und komplexe Systeme.....	53
Lernen und mentale Modelle.....	54
Zusammenfassung.....	55
Usability.....	56
Ursprung des Begriffs.....	56
Usability in der EN ISO 9241-10.....	58
Usability nach Ulich.....	60
Usability nach Nielsen.....	62
Usability in der EN ISO 9241-11.....	64
Zusammenfassung.....	66
Usability vor dem Hintergrund mentaler Modelle.....	69
Mentale Modelle und das Erlernen der Benutzung.....	69
Vermittlung konzeptioneller Modelle.....	71
Mentale Modelle bei der Nutzung moderner Informationsverarbeitung.....	73
Vergleich einer Theorie mentaler Modelle mit den Usability-Konzepten.....	74
Präzisierung der Fragestellung.....	80
Definition der neuen Konstrukte.....	84
Empirischer Teil.....	87

Skalenentwicklung.....	87
Befragung zu den Konstruktdefinitionen.....	89
Analyse ähnlicher Skalen.....	90
Itempool.....	94
Erste Itemselektion (Sortieraufgabe).....	96
Methode.....	96
Erhebung.....	98
Auswertung der Daten.....	98
Ergebnisse.....	100
Zusammenfassung der ersten Itemauswahl.....	121
Zweite Itemselektion – Beurteilung von Routenplanern.....	122
Untersuchung.....	123
Aufgabenstellung.....	123
Auswertungen.....	124
Verteilungsform der Itemantworten.....	124
Korrelationsmatrix der Items.....	127
Faktorenanalyse der Items.....	133
Zusammenfassung der zweiten Itemauswahl.....	136
Empirische Überprüfung des Prozessmodells.....	137
Design der Studie.....	138
Der Online-Routenplaner mit Alertness-Management.....	139
Untersuchung.....	149
Stichprobe.....	151
Auswertungen.....	151
Verteilungsform der Itemantworten.....	152
Analyse der Interitemkorrelationen.....	157
Bildung der Skalenwerte.....	163
Verteilungsform der Skalenwerte.....	164
Modellierung der Nutzungsintention.....	168
Modellvergleich.....	172
Analyse des Prozessmodells.....	177
Zusammenfassung.....	184
Diskussion.....	185
Die Ergebnisse im Hinblick auf die Hypothesen.....	185
Validität der Arbeit.....	187
Validität der Skalen.....	187
Validität der neuen Skalen für PL und PEUEU.....	187
Validität der Skala für PEOU.....	188
Validität der Skalen für PU und der Nutzungsintention.....	189
Repräsentativität der Stichprobe.....	189
Repräsentativität der eingesetzten Software.....	190
Operationalisierung der Erfahrung der Benutzer.....	191

Ausblick: Einfluss der Erfahrung auf den Bewertungsprozess.....	192
Praktische Bedeutung der Ergebnisse.....	194
Literaturverzeichnis.....	197
Anhang A.....	206
Anhang B.....	211
Anhang C.....	215
Anhang D.....	219
Anhang E.....	221
Anhang F.....	223
Anhang G.....	225
Anhang H.....	228
Anhang I.....	250
Anhang J.....	280
Anhang K.....	286
Anhang L.....	294
Anhang M.....	342

Einleitung

Die letzten 30 Jahre waren geprägt von einer rasanten Entwicklung der digitalen Informations- und Kommunikationstechnik. Nach dem Einzug des Computers in die Arbeitswelt in den 80er Jahren des vorigen Jahrhunderts ist die digitale Technik inzwischen in nahezu allen Lebensbereichen anzutreffen. Den großen Potentialen dieser Entwicklung stehen verschiedene Probleme und Gefahren gegenüber. So behindern mangelnde Akzeptanz und Schwierigkeiten mit der Bedienung die Verbreitung und effiziente Nutzung moderner Informations- und Kommunikationstechnologien. Hierdurch wird der erwünschte Nutzen durch den Einsatz von Computern verfehlt. Im Extremfall führt dies zu einer Ausgrenzung von im Umgang mit Computern ungeübten Personengruppen aus gesellschaftlichen Bereichen, die durch Informations- und Kommunikationstechnologie geprägt sind. Häufig sind hierbei gerade ältere Menschen betroffen.

Das Ziel dieser Arbeit ist es, das Verständnis der Prozesse, die die Akzeptanz neuer Informations- und Kommunikationstechnologien bestimmen, weiterzuentwickeln. Dabei liegt der Schwerpunkt auf dem Zusammenhang zwischen Usability – also der Bedienbarkeit oder „Benutzerfreundlichkeit“ – und der Akzeptanz einer neuen Technologie. Hierbei ergibt sich scheinbar eine Diskrepanz zwischen dem augenscheinlichen gewichtigen Einfluss der Bedienbarkeit auf die Akzeptanz und Forschungsergebnissen, die dieser nur eine untergeordnete Bedeutung zusprechen. Diese Arbeit versucht, diesen Widerspruch zu erklären. Hierzu wird der Einfluss der Erfahrung eines Benutzers auf dessen Wahrnehmung der Bedeutung von Bedienbarkeit untersucht.

Nutzung und Bedeutung der Informations- und Kommunikationstechnologien

Der Betrieb und die Nutzung von Computern war lange Zeit die Domäne speziell ausgebildeter Fachkräfte. Diese verfügten zumeist über Programmierkenntnisse und die Kommunikation mit dem Computer erfolgte durch spezielle Befehle, die in eine Textkonsole eingegeben wurden. Mit sinkenden Hardwarepreisen bei gleichzeitig steigender Rechenleistung entdeckte die Arbeitswelt vor ungefähr 30 Jahren zunehmend die Möglichkeiten, die die digitale Verarbeitung von Informationen bot. Von zentraler Bedeutung war hierbei die Verknüpfung der Informationstechnik mit der Telekommunikation.

Die Entwicklung des so genannten Personal Computers (PC), wie etwa dem Xerox Alto oder dem Apple II, führte zu einem zunehmenden Einsatz von Computern. Damit begann sich die Gruppe der Computerbenutzer systematisch zu verändern: die computerspezifischen Kenntnisse von Computerbenutzern nahmen mit zunehmender Verbreitung des Computers immer weiter ab. Da für immer mehr Aufgaben Software zur Verfügung stand, wurde es möglich, Computer auch ohne

Programmierkenntnisse sinnvoll zu nutzen.

Die Entwicklung von Standardsoftware für allgemeine Arbeitsaufgaben wie beispielsweise Textverarbeitung oder Tabellenkalkulation verbreiterte die Gruppe potentieller und tatsächlicher Computernutzer. Inzwischen hat der Computer in der Arbeitswelt eine Bedeutung erreicht, die Krings (2004) beschreibt „als ein zentrales Arbeitsinstrument in nahezu allen Sektoren“. Die überwiegende Mehrheit der Beschäftigten wurde hierdurch zu Computerbenutzern, an die „neue qualitative Anforderungen herangetragen“ (Krings, 2004) wurden. Anwender ohne den entsprechenden mathematisch-technischen Hintergrund hatten regelmäßig gravierende Schwierigkeiten bei der Nutzung des Computers. Dies rührte daher, dass die Interaktion zwischen Benutzer und Computer nicht mit Fokus auf den Benutzer, sondern im Hinblick auf die Möglichkeiten des Computers gestaltet wurde.

Als die Mehrheit der Benutzer begann, den Computer primär als Arbeitswerkzeug und Hilfsmittel und nicht mehr als Entwicklungs- und Forschungsgegenstand zu betrachten, wurden Schwierigkeiten im Umgang mit dem Computer nicht mehr länger als Herausforderung, sondern zunehmend schlicht als störend und lästig empfunden. Dies führte dazu, dass sich die durch den Einsatz der Informationstechnik im professionellen Umfeld erhofften Verbesserungen häufig nicht realisierten, da den Systemen die Akzeptanz der Nutzer fehlte.

In dieser Zeit entdeckte die Wissenschaft das Spannungsfeld zwischen Mensch und Computer als Forschungsgegenstand. Ben Shneiderman publizierte 1980 die Monographie *Software Psychology* und gilt damit als Begründer dieser Forschungsrichtung. Hierbei interessierte, wie Benutzerschnittstellen ergonomischer gestaltet werden könnten ebenso wie die Frage der Akzeptanz der neuen Technologien. Einen bedeutsamen Beitrag zur Erforschung der Akzeptanz lieferte 1989 Davis mit der Publikation seines *Technology Acceptance Models* (TAM), welches auch in dieser Arbeit als theoretisches Grundgerüst genutzt wird. Der Stellenwert dieses Themenbereichs wird ersichtlich an einer Literaturstudie von Han (2003). Für den Zeitraum zwischen 1989 und 2002 konnte der Autor bereits 92 Untersuchungen zur Einführung und Akzeptanz neuer Informationstechnologien in betriebliche Organisationen auswerten.

In den letzten 30 Jahren hat sich die generelle Bedeutung der Informations- und Kommunikationstechnologien deutlich gesteigert. Der Einsatz digitaler Datenverarbeitung ist nicht länger auf die Arbeitswelt beschränkt, sondern hat alle gesellschaftlichen Bereiche durchdrungen. Die Veränderungen sind so gravierend, dass von einem neuen Zeitalter – dem Informationszeitalter – gesprochen wird. In einem Bericht des Bundesministeriums für Wirtschaft war 1996 zu lesen:

Die führenden Industrieländer und damit auch die Bundesrepublik Deutschland stehen an der Schwelle des 21. Jahrhunderts vor einem Sprung in ihrer wirtschaftlich-technologischen Entwicklung, der hin zur Informationsgesellschaft führt. Dieser Wandel ist keine Vision, sondern bereits in vollem Gange.

Die von den Autoren des Bericht vorhergesehenen Möglichkeiten für den Einsatz der neuen Informations- und Kommunikationstechnologien sind heute größtenteils realisiert. Dabei handelt es sich außerhalb des geschäftlichen Bereichs um Anwendungsfelder im Verkehr, in Bildung und Forschung, im Gesundheitswesen, in der Unterhaltungsbranche sowie in der öffentlichen Verwaltung.

Da diese Arbeit im Rahmen von Forschung im Verkehrsbereich entstanden ist, beschränkt sich die Darstellung der Möglichkeiten moderner Informations- und Kommunikationstechnologie im Folgenden auf diesen Bereich. Die nachfolgend geschilderten Beispiele sollen verdeutlichen, wie umfassend Informationstechnik in diesem Bereich zum Einsatz kommt und wie weitreichend mangelnde Akzeptanz den erhofften Nutzen gefährden kann.

Der Einsatz von Computer hat die Reiseplanung und den Fahrkartenerwerb im öffentlichen Personenverkehr grundlegend verändert. Hier fand eine Verlagerung von Tätigkeiten auf den Kunden statt, der heute entweder mithilfe von Fahrkartenautomaten oder über das Internet selbstständig die Planung einer Reise (beispielsweise über www.bahn.de) und den Kauf des Tickets vornimmt. Bei diesen Tickets sucht der Kunde online nach einer geeigneten Verbindung und bezahlt das zugehörige Ticket per Lastschrift oder Kreditkarte. Das eigentliche Ticket stellt ein Ausdruck einer pdf-Datei dar, welche der Kunde nach der Buchung per E-Mail zugeschickt bekommt. Der gesamte Vorgang kann vom Kunden selbst bearbeitet werden, wobei er hierbei mit Systemen in Kontakt kommt, die früher durch geschultes Personal an den Verkaufsschaltern bedient wurden.

Eine ähnliche Möglichkeit zum Ticketkauf bietet im Luftverkehr die Lufthansa unter der Produktbezeichnung etix[®] an. Ab April 2008 können Tickets für Inlandsflüge bei der Lufthansa nur noch online als etix[®] gebucht werden. Das etix[®] ermöglicht eine Buchung bis spätestens zwei Stunden vor Abflug. Passagieren, die ein etix[®] erworben haben, können selbstständig an einem Check-In-Automaten für den Flug einchecken. Auch hier findet eine Verlagerung von Tätigkeiten weg vom Schalterpersonal hin zum Kunden statt, der hierfür selbst direkt mit der erforderlichen Technik interagieren muss.

Im Bereich des Personennahverkehrs haben beispielsweise die Berliner Verkehrsbetriebe (BVG) unter der Produktbezeichnung tick.et[®] mit der Einführung eines elektronischen Tickets begonnen. Eine Chipkarte ersetzt dabei den herkömmlichen Fahrschein. Diverse Kundendaten werden auf der Karte abgespeichert. Der Fahrgast muss sich vor Fahrtbeginn mit dieser Karte an einem Automaten anmelden und bei Fahrtende wieder abmelden. Die Chipkarte stellt weitere Funktionalität bereit, die über den klassischen Papierfahrschein hinausgeht. Während bei klassischen Papierfahrschein am Entwerteraufdruck problemlos zu erkennen war, ob man sich im Besitz eines gültigen Fahrscheins befindet, muss bei der Chipkarte über ein Dialogsystems beim Einsteigen erkannt werden, ob korrekt eingeloggt wurde. Der Fahrgast kann den Status seines „Fahrscheins“ während der

Fahrt nicht mehr mit einfachen Mitteln wie einer visuellen Kontrolle selbst überprüfen, sondern gerät in eine Abhängigkeit von den zugehörigen Automaten.

Auch im Bereich des Individualverkehrs steigt die Verbreitung moderner Informations- und Kommunikationstechnologie an. Die Fahrzeuge werden – teils von den Herstellern, teils später durch Nachrüstlösungen – mit einer Vielzahl von Assistenzsystemen ausgestattet. Eine hohe Nachfrage erleben aktuell Navigationsassistenten, die den Fahrer bei der Routenplanung und Wegfindung unterstützen.

Die Fähigkeiten dieser Systeme wird ständig weiterentwickelt. Eine Koppelung der Navigationsgeräte an die Verkehrsnachrichten (via TMC übertragen) ermöglicht eine dynamische Anpassung der Route an die aktuelle Verkehrssituation. Da es bei den meisten Erweiterungen der Funktionalität geboten ist, Vorlieben des Benutzers zu berücksichtigen, steigt mit dem Funktionsumfang auch die Komplexität der Interaktion zwischen Benutzer und System an.

Dabei existiert ein generelles Problemfeld hinsichtlich der Wirkung von Assistenzsystemen: der Unterstützung bei der Fahraufgabe steht eine ablenkende Wirkung durch die Bedienung der Geräte gegenüber. Im Hinblick auf Mobiltelefone trägt der §23 der StVO diesem Konfliktfeld Rechnung und untersagt die Benutzung von Handys, falls diese hierzu in der Hand gehalten werden müssen. Andere Informations- und Kommunikationssysteme, wie PDAs oder Navigationsassistenten werden durch die StVO bislang nicht berücksichtigt.

Vielfältige Unterstützungsmöglichkeiten bieten Plattformen zur Online-Routenplanung, wie beispielsweise map24.de, mappy.de oder Viamichelin.de. Neben der reinen Routenplanung unterstützen diese Dienste bei der Suche nach Übernachtungsmöglichkeiten entlang der Strecke, berechnen die voraussichtliche Fahrtdauer und erlauben die Eingabe diverser Optimierungsparameter für die Streckenauswahl. Dies umfasst meistens die Wahl zwischen kürzester und schnellster Strecke, aber auch komplexere Optimierungen wie beispielsweise „von Michelin empfohlen“ bei ViaMichelin.de.

Die Schilderung von Einsatzszenarien der modernen Informations- und Kommunikationstechnologien dient lediglich einem groben Überblick und beschränkt sich auf den Verkehrsbereich. Sie soll verdeutlichen, wie gravierend sich die durch die technische Entwicklung ausgelösten Veränderungen auswirken und wie hierdurch neue Problemfelder entstehen. Um das volle Potential der technischen Möglichkeiten im praktischen Einsatz zu realisieren, müssen die konkreten Lösungen so gestaltet werden, dass die eine hohe Nutzungsquote dank guter Akzeptanz erzielen.

Usability und Akzeptanzforschung

Mit der wachsenden Verbreitung des Computers und den damit einhergehenden Problemen nahm das Interesse der Wissenschaft am Themenbereich

der Interaktion zwischen Mensch und Computer zu. Die Ergebnisse dieser Forschung stellen die theoretischen Grundlagen dieser Arbeit dar. Im Folgenden wird ein kurzer Überblick über diesen Themenbereich gegeben.

Mit zunehmender Verbreitung von Informationstechnik und einer rasant wachsenden Benutzergruppe veränderte sich der durchschnittliche Kenntnisstand der Computerbenutzer. Das Hintergrundwissen über Computer nahm ab, der Anteil von Benutzern mit Programmierkenntnissen wurde geringer und vor allem verringerte sich die Bereitschaft, die Auseinandersetzung mit dem Computer als lohnende Herausforderung zu betrachten. Computerbenutzer wollten zunehmend zielstrebig Aufgaben mit Hilfe des Computers lösen und sich hierfür nicht langwierig in die Bedienung des Computers selbst einarbeiten. Am klarsten kann der Wandel dadurch beschrieben werden, dass sich nicht länger der Benutzer die Sprache und das Denken des Computers zu eigen machen musste, sondern dass die Softwareprogrammierer begannen, den Computer an die Sprache und das Denken des Benutzers anzupassen.

Das Interesse der Forschung an der Interaktion zwischen Mensch und Computer nahm parallel zum Interesse der Arbeitswelt am Computer als Werkzeug, zu. Für Software wie Textverarbeitung und Tabellenkalkulation galt es Bedienkonzepte zu entwickeln, die diese Anwendungen einer breiten Nutzergruppe zugänglich machen sollten. Für die Forschung stellte sich die Frage, wie sich die Benutzbarkeit unterschiedlicher Bedienkonzepte messen lässt, wie sich die Bedienkonzepte hinsichtlich der Benutzbarkeit unterscheiden und wie gute Bedienkonzepte systematisch entwickelt werden konnten.

Mit Ben Shneidermans (1980) Monographie „Software psychology“ entstand eine neue Forschungsdomäne der Psychologie, welche auf die Erforschung und das Verständnis der Interaktion zwischen Benutzern und moderner Informations- und Kommunikationstechnologie ausgerichtet ist. Der Fokus der Forschungsbemühungen lag häufig weniger in einem verbesserten theoretischen Verständnis der Interaktion zwischen Mensch und Computer, sondern vielmehr in der Beantwortung konkreter Fragen zur Gestaltung von Bedienoberflächen.

Ein zentraler Begriff im Bereich der Software-Psychologie ist *Usability* (Gebrauchstauglichkeit, Übers. nach EN ISO 9241). Der Begriff geht vermutlich auf einen Ansatz von Boehm, Brown und Lipow (1976) zurück, die grundlegende Anforderungen an Software definieren. Sie benutzen den Begriff *Usability*, um die Einschätzung der Tauglichkeit einer Software durch den Endbenutzer zu beschreiben.

Heute regelt die EN ISO 9241 als europäische Norm die *Ergonomischen Anforderungen für Bürotätigkeiten mit Bildschirmgeräten*. Wichtige Gestaltungsregeln für die Benutzerschnittstelle finden sich in Teil 10¹: Grundlagen der Dialoggestaltung. Die Autoren der Norm vertreten eine Sichtweise auf Usability, in der diese

1seit 01.04.2006 gilt der überarbeitete Teil 110, der inhaltlich weitgehend dem dieser Arbeit zugrundeliegenden, älteren Teil 10 entspricht

weitgehend als Eigenschaft der Software und somit losgelöst vom Benutzer und dessen Verwendungszweck betrachtet wird. Dies scheint insofern problematisch, als Usability nur im Rahmen der Interaktion eines Benutzers mit einer Software für die Lösung eines gegebenen Problems beobachtbar wird (wie der Begriff „Dialoggestaltung“ im Titel der EN ISO 9241 schon vermuten lässt). In der Usability-Forschung herrscht bislang keine einheitliche Auffassung, inwieweit Usability sinnvoll unabhängig von einer Benutzergruppe und einem Einsatzzweck definiert werden kann.

Ungeachtet dieser Uneinigkeit bezüglich der Definition von Usability herrscht Einigkeit in der Software-Ergonomie und Usability-Forschung darüber, dass Usability ein wichtiges Erfolgskriterium für ein System darstellt. Jakob Nielsen (2003) fasst dies für die Gestaltung von Webseiten folgendermaßen zusammen:

On the Web, usability is a necessary condition for survival. If a website is difficult to use, people leave. If the homepage fails to clearly state what a company offers and what users can do on the site, people leave. If users get lost on a website, they leave. If a website's information is hard to read or doesn't answer users' key questions, they leave. Note a pattern here?

Im Internet ist Usability eine lebenswichtige Eigenschaft. Falls eine Webseite schwierig zu benutzen ist, werden die Besucher sie verlassen. Falls die Startseite einer Firma keine deutliche Aussage zum Angebot der Firma und den Möglichkeiten auf der Webseite macht, werden die Besucher sie verlassen. Falls sich Benutzer auf einer Webseite verirren, werden die Besucher sie verlassen. Falls die Informationen einer Webseite schlecht zu lesen sind oder diese die zentralen Fragen der Benutzer nicht beantworten, werden die Besucher sie verlassen. Wird das Muster klar? (Übers. d. Verf.)

In einer Studie *Kundenkompass Onlineshopping*, die das F.A.Z.-Institut zusammen mit der novomind AG im Jahr 2004 durchgeführt hat, wurde festgestellt, dass 40 % der potentiellen Kunden ihren Einkauf aufgrund technischer Probleme abgebrochen haben. Als weitere Gründe für abgebrochene Einkäufe nennt die Studie unzureichende Informationen, zu hohe Versandkosten und zu aufwendige Bestellvorgänge. Abgesehen von den Versandkosten fallen somit alle in der Studie genannten Gründe für den Abbruch eines Einkaufs in den Bereich der Usability.

Diese Aussagen aus dem Bereich der Usability-Forschung lassen erwarten, dass die Akzeptanz eines Informationsverarbeitungssystems eng an dessen Gebrauchstauglichkeit gekoppelt ist. Die Akzeptanz moderner Technologien ist Gegenstand einer – der Usability-Forschung verwandten – eigenen Forschungsrichtung. Interessanterweise sind die Befunde der Akzeptanzforschung bezüglich des Stellenwerts, den Usability für die Akzeptanz einnimmt, nicht so eindeutig, wie auf Grundlage der Ergebnisse der Usability-Forschung zu erwarten wäre.

Das bereits weiter oben erwähnte TAM nennt – vereinfacht dargestellt – zwei Determinanten für die Akzeptanz einer neuen Informations- und

Kommunikationstechnologie: die *wahrgenommene Nützlichkeit* („perceived utility“) und die *wahrgenommene Bedienbarkeit* (perceived ease of use). Während sich die *wahrgenommene Nützlichkeit* in einer Vielzahl von Studien konstant als wichtiger Prädiktor der Technologieakzeptanz bewährt hat, schwankt der Einfluss der *wahrgenommenen Bedienbarkeit* deutlich.

Hierbei ist zu beachten, dass das TAM keinerlei Aussagen über einen Zusammenhang zwischen einer – wie auch immer zu definierenden – objektiven Nützlichkeit bzw. objektiven Bedienbarkeit und der Akzeptanz macht. Statt dessen wird die Wahrnehmung eben dieser Systemeigenschaften als handlungsrelevant betrachtet. Die Bemühungen der Usability-Fachleute zielen ebenso nicht primär auf eine objektiv gebrauchstaugliche Gestaltung von Technologie. Es ist sogar fragwürdig, ob eine solche objektive Gebrauchstauglichkeit – also Gebrauchstauglichkeit unabhängig von einem benutzenden Subjekt – überhaupt sinnvoll definiert werden kann.

Die widersprüchlichen Befunde bezüglich der Rolle der Gebrauchstauglichkeit können entsprechend nicht einfach dadurch erklärt werden, dass von den beiden Forschungsrichtungen Aussagen über zwar ähnliche, aber doch substantiell verschiedene Konstrukte, getroffen werden. Vielmehr ist es so, dass die *Gebrauchstauglichkeit* der Usability-Forschung und die *wahrgenommene Bedienbarkeit* der Akzeptanz-Forschung als hinreichend ähnlich betrachtet werden müssen.

Die in dieser Arbeit untersuchte Erklärung für den schwankenden Einfluss der Bedienbarkeit liegt – stark verkürzt – darin, dass sich die Anforderungen eines Benutzers mit zunehmender Erfahrung desselben verändern. Je erfahrener ein Benutzer mit der Bedienung einer Software ist, desto unwichtiger wird für diesen deren Erlernbarkeit (Erlernbarkeit sowohl im Usability-Konzept als auch im TAM ein gewichtiger Einflussfaktor). Gleichzeitig wird für den Benutzer immer wichtiger, dass sich die Software effizient bedienen lässt. Die Akzeptanz einer Informations- und Kommunikationstechnologie würde also neben ihrer Nützlichkeit zunächst von der Erlernbarkeit, im Verlauf der Benutzung aber statt dessen zunehmend von der Möglichkeit zur effizienten Benutzung, abhängen.

Die Modellierung dieser Veränderung in den Bewertungsprozessen erfordert eine differenziertere Betrachtung der Wahrnehmung der Usability einer Technologie, die im TAM durch das Konstrukt *Perceived Ease of Use* (PEOU, wahrgenommene Leichtigkeit der Benutzung, Übers. d. Verf.) repräsentiert ist. Hierzu werden in dieser Arbeit zwei neue Konstrukte definiert, die gemeinsam den Bereich des TAM-Konstrukts *Perceived Ease Of Use* abdecken. Das erste der beiden neuen Konstrukte fokussiert sich inhaltlich auf die wahrgenommene Erlernbarkeit eines Bedienkonzepts, während das zweite neue Konstrukt sich auf die wahrgenommene Effizienz und Effektivität der Bedienung durch erfahrene Benutzer beschränkt. Diese Aufteilung ermöglicht eine Betrachtung des Einflusses unterschiedlicher Aspekte des

TAM-Konstrukts *Perceived Ease Of Use*. Dabei wird ein Einfluss der Erfahrung, die ein Benutzer mit einem System hat, erwartet.

Das nächste Kapitel stellt das für diese Arbeit gewählte Vorgehen zur Untersuchung der Hypothese dar. Dabei wird ein Überblick über die inhaltliche Gliederung der Arbeit gegeben.

Aufbau und Vorgehen dieser Arbeit

Der theoretische Teil dient einer begründeten Herleitung der Hypothesen der Arbeit. Im Rahmen der theoretischen Weiterentwicklung des TAM wurden zwei neue Konstrukte definiert, die das TAM zu einem Prozessmodell verändern. Im empirischen Teil der Arbeit diente die Mehrzahl der Studien der zunächst erforderlichen Entwicklung der Skalen für die neuen Konstrukte. In der letzten Studie wurden die neuen Skalen zur empirischen Überprüfung der Hypothesen eingesetzt.

Der theoretische Teil der Arbeit beginnt mit einer detaillierten Betrachtung des TAM. Hierfür wird zunächst der theoretische Vorläufer des TAM, die sozialpsychologische Theorie des überlegten Handelns (Fishbein & Ajzen, 1975; Ajzen & Fishbein, 1980) knapp dargestellt. Es folgt eine Beschreibung der Struktur des TAM mit den beiden Konstrukten „Perceived Usefulness“ (PU, wahrgenommene Nützlichkeit, Übers. d. Verf.) und „Perceived Ease of Use“ (PEOU, wahrgenommene Leichtigkeit der Bedienung, Übers. d. Verf.), die im TAM als Determinanten der Nutzungsabsicht fungieren.

Einen Schwerpunkt bei der Betrachtung des TAM bildet der Prozess der ursprünglichen Skalenentwicklung zum Konstrukt PEOU. Im Prozess dieser Skalenentwicklung veränderte sich der Fokus der Messskala zunehmend von der Konstruktdefinition weg. Während die Konstruktdefinition noch die Benutzbarkeit eines Systems durch alle Benutzer umfasste, wurde bei der Entwicklung der originalen Skala PEOU zunehmend mit Erlernbarkeit gleichgesetzt.

Die Darstellung der Forschungsergebnisse zum TAM ist in zwei Teile gegliedert. Im ersten Teil sind Studien zusammengefasst, deren Absicht in der empirischen Überprüfung des TAM lag. Der zweite Teil besteht aus Studien, in denen das TAM erweitert oder mit anderen Theorien verschmolzen wurde. Als wesentliche Ergebnisse aller Studien zeigt sich zum einen, dass einem konstant hohen Einfluss des Konstrukts PU auf die Nutzungsintention ein deutlich schwankender Einfluss des zweiten Konstrukts PEOU gegenüber steht. Die Studien des zweiten Teils zeigen darüber hinaus, dass keine der vorgenommenen Veränderungen am TAM eine substantielle Verbesserung der Modellierung erzielen konnte.

Den Hintergrund für die Weiterentwicklung des TAM zum Prozessmodell bilden kognitionspsychologische Theorien, insbesondere das Konzept der mentalen Modelle. Darunter versteht man – vereinfacht dargestellt – die Vorstellungen, die Personen von ihrer Umwelt haben und die ihnen als Handlungsgrundlage dienen. Nach der

Diskussion des TAM mit den zugehörigen Forschungsergebnissen erfolgt entsprechend eine Darstellung des Konzepts mentaler Modelle und die Anwendung dieses Konzeptes auf die Interaktion zwischen Menschen und modernen Informations- und Kommunikationstechnologien.

Ein mentales Modell spielt eine entscheidende Rolle bei der Interaktion mit komplexer Technik. Es erlauben als mentale Simulation, gefahrlos verschiedene Handlungsoptionen durchzuspielen und die Ergebnisse der jeweiligen Möglichkeiten gegeneinander abzuwägen. Mentale Modelle bilden hierzu komplexe Sachverhalte in vereinfachter, für die jeweilige Aufgabenstellung funktionaler Darstellung ab. Zur Hypothesenbildung dieser Arbeit trägt das Konzept mentaler Modelle somit bei, weil es eine relevante Verständnisgrundlage für die Interaktion zwischen Mensch und Computer darstellt.

Eng mit dem Konzept eines mentalen Modells ist der Schema-Begriff verknüpft. Vereinfacht dargestellt repräsentieren Schemata erlerntes, im Langzeitgedächtnis gespeichertes Wissen. Das mentale Modell einer Situation wird bei Bedarf anhand des gelernten schematischen Wissens gebildet. Der Lernvorgang wird für diese Arbeit, in der die Erfahrung eines Benutzers einen entscheidenden Einflussfaktor im Prozess der Technologiebewertung darstellt, als Wechselspiel zwischen mentalem Modell und schematischem Wissen betrachtet.

Die Kernhypothese dieser Arbeit entstand im Spannungsfeld zwischen den Forschungsergebnissen der Akzeptanz- und der Usability-Forschung. Entsprechend folgt eine Darstellung des Usability-Konzepts. Dabei werden zunächst der Ursprung dieses jungen Begriffs und seine Bedeutung beleuchtet. Im Anschluss werden vier wichtige Usability-Konzeptionen näher betrachtet.

Den Abschluss des theoretischen Teils dieser Arbeit bildet die zusammenfassende Betrachtung des theoretischen Wissens im Hinblick auf die widersprüchlichen Befunde zum Einfluss der Bedienbarkeit in der Akzeptanz- und der Usability-Forschung. Hieraus ergibt sich die Formulierung der Untersuchungshypothesen mit der Definition zweier neuer Konstrukte, die das TAM zu einem Prozessmodell erweitern. Die Überprüfung der Hypothesen erfolgt anschließend im empirischen Teil der Arbeit.

Die empirische Untersuchung der formulierten Hypothesen erfolgt in drei grundlegenden Schritten. Zunächst mussten zwei neue Messskalen zur Erfassung der neu formulierten Konstrukte des Prozessmodells entwickelt werden. In einem zweiten Schritt wurden zur Validierung mit den Skalen Daten erhoben und die Itemauswahl anhand der Ergebnisse verfeinert. Mit den neu entwickelten Skalen erfolgte abschließend die Erhebung der Daten für den Modellvergleich zwischen TAM und Prozessmodell. Die theoretischen Erläuterungen im empirischen Teil dienen der Begründung der Wahl der Untersuchungs- und Auswertungsmethoden und sind insofern für die Kernfrage dieser Arbeit irrelevant. Entsprechend sind sie – zur Verbesserung des Verständnisses – in den empirischen Teil der Arbeit eingeflochten.

Der erste Schritt der Skalenkonstruktion bestand in der Bestimmung des Bereichs, der durch eine valide Skala inhaltlich abgedeckt werden muss. Hierzu wurden verschiedene erfahrene Computerbenutzer befragt, welche Eigenschaften einer Software relevant sind für PL bzw. PEUEU. Die gesammelten Aussagen konnten zu neun Aspekten gegliedert werden. Ergänzend wurden Skalen zur Messung ähnlicher Konstrukte analysiert, um sicherzustellen, dass alle relevanten Aspekte der Konstrukte in den neuen Skalen berücksichtigt wurden. Auf der Grundlage der gesammelten Aspekte wurde ein Itempool mit jeweils 17 Items pro Konstrukt formuliert.

Zur Auswahl geeigneter Items musste festgestellt werden, wie sehr die Items inhaltlich mit dem Konstrukt verknüpft sind und wie ähnlich sich die Items untereinander sind. Items, die inhaltlich wenig Bezug zum zu messenden Konstrukt aufweisen, sind nicht für eine Skala geeignet. Ebenso sind mehrere ähnliche Items nicht wünschenswert, da der Informationsgewinn, den jedes Item zum Skalenwert beiträgt, umso geringer wird, je ähnlicher sich die Items sind. Die Ähnlichkeit der Items untereinander und ihr Bezug zum Zielkonstrukt der Skalen wurden durch verschiedene Sortieraufgaben erhoben. Eine Auswertung der Itemsortierungen erfolgte non-parametrisch durch eine multi-dimensionale Skalierung (MDS).

Für die Validierung der neuen Skalen wurden mit der getroffenen Itemauswahl Daten erhoben. Die statistische Auswertung der Daten erfolgte auf der Grundlage der Interitemkorrelationen. Eine detaillierte Betrachtung der Korrelationsmatrix ermöglichte die Auswahl von Items, deren Korrelationen sich im Wesentlichen innerhalb des von Clark und Watson (1995) empfohlenen Bereichs bewegten. Eine Faktorenanalyse bestätigte die inhaltliche Trennung der beiden Konstrukte sowie die Skalenhomogenität innerhalb der beiden Skalen.

Den Abschluss des empirischen Teils dieser Arbeit bildet der statistische Vergleich der Modellierung durch das TAM und das Prozessmodell. Die Daten für die Parameterschätzung der beiden Modelle wurden an einem Prototyp für einen Routenplaner mit integrierter Pausenplanung erhoben.

Diese Software war eine Neuentwicklung, was die experimentelle Kontrolle der Erfahrung der Probanden ermöglichte. Zur Erhebung unterschiedlich erfahrener Benutzer wurden zwei Messungen durchgeführt, zwischen denen die Probanden mit der Software arbeiten mussten, um Erfahrung aufzubauen. Es ergab sich ein Versuchsdesign mit Messwiederholung.

Vor der Auswertung der Daten im Hinblick auf die Hypothesen erfolgte zunächst eine Kontrolle der Skaleneigenschaften der beiden neu entwickelten Skalen. Die Schätzung der Modellparameter erfolgte – entsprechend des Designs mit Messwiederholung – im Rahmen einer sogenannten gemischten Modellierung. Dabei erzielte das Prozessmodell eine signifikant bessere Modellierung der Versuchsdaten gegenüber dem TAM, was die Haupthypothese dieser Arbeit bestätigte.

Theoretischer Teil

Technology Acceptance Model (TAM)

Der realisierte Nutzen technischer Lösungen hängt maßgeblich von deren Akzeptanz durch potenzielle Benutzer ab. Während der Etablierungsphase von Computern in der Arbeitswelt in den 1980er Jahren konnten erhoffte Produktivitätssteigerungen häufig mangels Akzeptanz durch die Mitarbeiter nicht realisiert werden (Bowen, 1986). Diese Tatsache nahm Davis (1986) zum Anlass, in seiner Dissertation ein Modell der Akzeptanz von Informationstechnologie zu entwickeln – das „Technology Acceptance Model.“ Das TAM modelliert die allgemeinen, grundlegenden Prozesse, die zu Technologieakzeptanz führen. Das TAM basiert in seinen Grundzügen auf der bewährten sozialpsychologischen Theorie des überlegten Handelns (Fishbein & Ajzen, 1975; Ajzen & Fishbein, 1980). Wie in dieser für menschliches Handeln allgemein formulierten Theorie führt auch im TAM die persönliche Einstellung zur Ausbildung einer Verhaltensintention. Die relevanten Einstellungen des TAM sind – entsprechend dessen Ausrichtung auf Informationstechnologie – „Perceived Usefulness“ (PU, wahrgenommene Nützlichkeit, Übers. d. Verf.) und „Perceived Ease of Use“ (PEOU, wahrgenommene Leichtigkeit der Bedienung, Übers. d. Verf.). Im TAM gibt es keinen direkten Einfluss von Eigenschaften der Informationstechnologie (z. B. Bedienoberfläche) auf die Verhaltensintention und das Verhalten. Nach der Publikation (Davis, 1989) bildete das TAM die Grundlage für zahlreiche wissenschaftliche Untersuchungen zur Computerakzeptanz und wurde im Zuge dessen erweitert und verändert. Von den Veränderungen ist im Kontext dieser Arbeit insbesondere die durch Davis, Bagozzi und Warshaw (1989) erwähnenswert. Sie vereinfachten das theoretische Modell durch den Verzicht auf das Einstellungskonstrukt, ohne dass sich die Güte der Modellierung nennenswert verschlechterte. Die meisten späteren wissenschaftlichen Arbeiten nutzten entsprechend das vereinfachte Modell ohne Einstellungskonstrukt als Grundlage. Im Folgenden werden zunächst die sozialpsychologischen Ursprünge des Modells und seine Komponenten betrachtet.

Theorie des überlegten Handelns (Theory of reasoned action)

Die Theorie des überlegten Handelns (Fishbein & Ajzen, 1975; Ajzen & Fishbein, 1980) sieht den Menschen als rationales Wesen, dessen Verhalten auf Einstellungen zum Verhalten und auf wahrgenommenen sozialen Erwartungen beruht. Abb. 1 veranschaulicht die Theorie des geplanten Handelns.

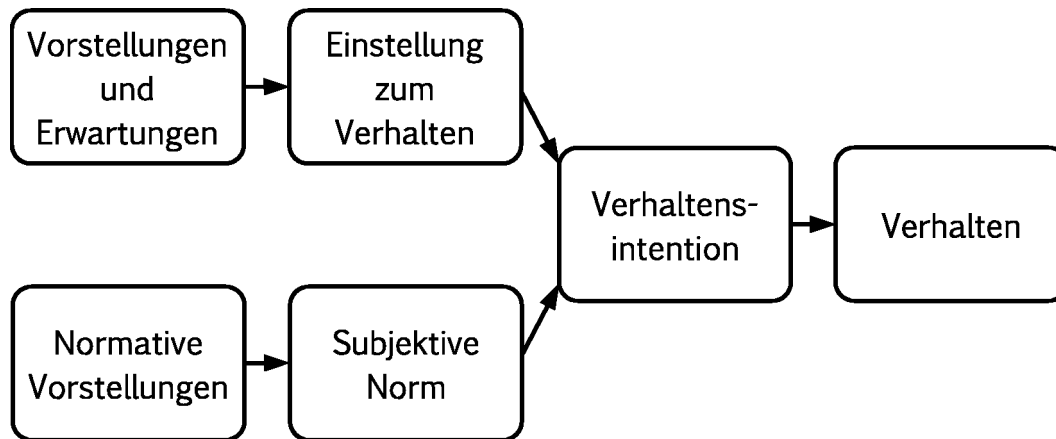


Abb. 1: Theorie des überlegten Handelns

Die Einstellung bezüglich eines konkreten Verhaltens ergibt sich aus den erwarteten Konsequenzen des Verhaltens. Werden vorteilhafte Konsequenzen erwartet, so führt dies zu einer positiven Einstellung zum Verhalten. *Normative Vorstellungen* sind die wahrgenommenen Erwartungen des sozialen Umfelds einer Person. Die *subjektive Norm* ist die gewichtete Summe dieser – teilweise unterschiedlichen – sozialen Erwartungen. Die unterschiedlichen Erwartungen werden nach der Bedeutung des Erwartungsträgers gewichtet. Gemeinsam führen die Einstellung zum Verhalten und die *subjektive Norm* zur Ausbildung einer Verhaltensintention. Äußere Einflüsse wirken nicht direkt auf die Verhaltensintention, sondern beeinflussen diese über die Einstellung oder die *subjektive Norm*. Nach dem Modell ist die Verhaltensintention der Prädiktor für tatsächliches Verhalten. Die Theorie des überlegten Handelns definiert Verhalten als nicht direkt von äußeren Einflüssen abhängig. Diese – von Fishbein und Ajzen (1975) als „äußere Variablen“ (Übers. d. Verf.) bezeichneten – Größen wirken indirekt durch ihren Einfluss auf die *Einstellung* oder die *subjektive Norm*.

Ajzen erweiterte die Theorie 1985 durch die Einführung des Konstrukts der „wahrgenommene Verhaltenskontrolle“ zur Theorie des geplanten Verhaltens.

Struktur des TAM

Für das TAM hat Davis die einstellungsabhängigen Komponenten der Theorie des überlegten Handelns übernommen. Die Aspekte des sozialen Einflusses auf die *Verhaltensintention* (*normative Vorstellungen*, *subjektive Norm*) sind nicht im TAM enthalten. Während für die Theorie des überlegten Handelns der Anwendungsbereich offen bleibt, beschränkt sich das TAM auf die Akzeptanz von Informationstechnologie. Infolgedessen konnte Davis spezifizieren, welche Vorstellungen und Erwartungen für die Einstellung relevant sind. Auf der Grundlage einer Literaturstudie hat er „Perceived Usefulness“ (PU, wahrgenommene Nützlichkeit, Übers. d. Verf.) und „Perceived Ease of Use“ (PEOU, wahrgenommene Leichtigkeit der Benutzung, Übers. d. Verf.) als relevante Vorläufer der Einstellung

definiert (Davis, 1989). Zusätzlich postulierte Davis einen direkten Einfluss der PU auf die Verhaltensintention, den die Theorie des überlegten Handelns nicht vorsieht. Es ergibt sich eine Grundstruktur für das TAM wie in Abb. 2 dargestellt (hellgrau sind die zusätzlichen Konstrukte der Theorie des überlegten Handelns).

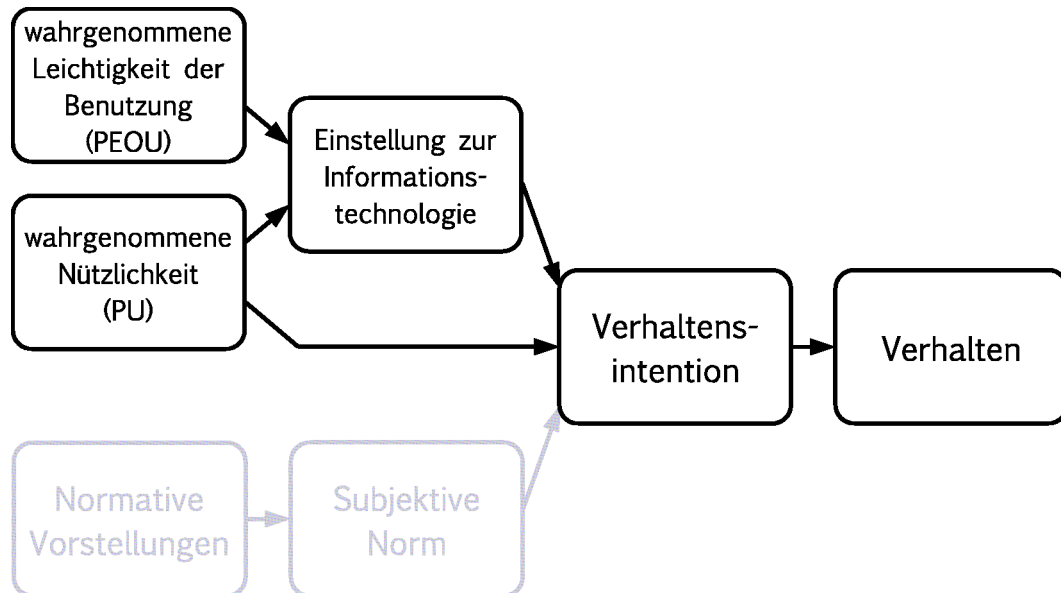


Abb. 2: Struktur des Technology Acceptance Models

In einer Studie von Davis, Bagozzi und Warshaw (1989) zur Akzeptanz einer Textverarbeitungssoftware wurde das TAM direkt mit der Theorie des überlegten Handelns verglichen. Das in beiden Theorien enthaltene Einstellungskonstrukt trug zur Erklärung der empirischen Daten nur wenig bei. Infolgedessen wurde die Einstellung als Mittler zwischen den Vorstellungen und Erwartungen und der Verhaltensintention aus dem TAM entfernt. Darüber hinaus zeigten die empirischen Daten einen indirekten Einfluss der PEOU auf die Verhaltensintention über PU. Das entsprechende Modell ist in Abb. 3 zu sehen und wurde in dieser Form von Davis in späteren Untersuchungen eingesetzt (Venkatesh & Davis, 1996, 2000).

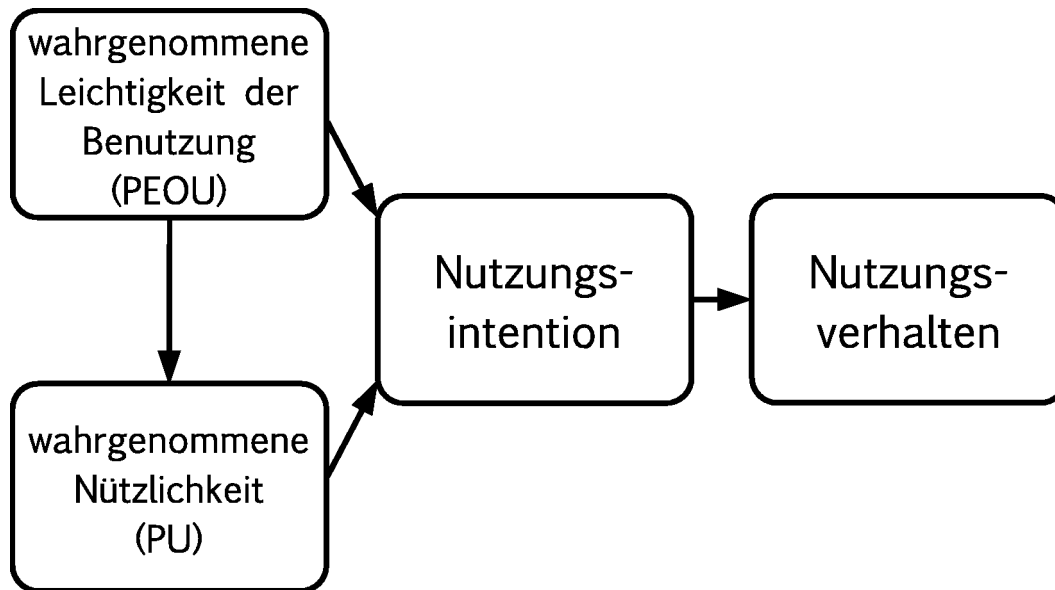


Abb. 3: Technology Acceptance Model nach Davis, Bagozzi & Warshaw (1989)

Perceived Usefulness

Die erste der beiden relevanten Vorstellungen eines Benutzers im TAM (Abb. 3) ist die wahrgenommene Nützlichkeit einer Informationstechnologie (PU). Die *Nutzungsintention* hängt entsprechend davon ab, ob der Benutzer sich von der Informationstechnologie eine Unterstützung bei der Lösung seiner Probleme verspricht. Davis Forschungsinteresse galt der beruflichen Computernutzung, weshalb sich die Definition auf den beruflichen Einsatz von Informationstechnologie bezieht. Davis definiert PU als „the degree to which a person believes that using a particular system would enhance his or her job performance“ (der Grad, zu dem eine Person glaubt, dass die Benutzung eines bestimmten Systems seine/ihre berufliche Leistungsfähigkeit steigern würde, Über. d. Verf.).

Davis Konstrukt PU geht auf Robey (1979) zurück. Ein wesentlicher Aspekt des Konstrukts ist sein indirekter Bezug zu Eigenschaften des technischen Systems. Es geht ausdrücklich nicht um die objektive Nützlichkeit desselben, sondern die subjektive Einschätzung derselben durch den Benutzer.

Perceived Ease of Use

Die zweite relevante Vorstellung eines Benutzers im TAM (Abb. 3) ist die wahrgenommene Leichtigkeit der Benutzung (PEOU). Der Einfluss der Leichtigkeit der Benutzung scheint hoch plausibel, wenn man sich zwei in Bezug auf deren Nützlichkeit identische Systeme vorstellt, die einen unterschiedlichen Bedienungsaufwand erfordern. Das TAM postuliert eine stärkere *Nutzungsintention* für eine Informationstechnologie, wenn diese leicht zu bedienen ist.

Leichtigkeit bezeichnet in Davis Definition ein möglichst geringes Maß an Anstrengung, das die Benutzung erfordert. Dabei bezieht Davis die Anstrengung sowohl auf das Erlernen der Benutzung wie auch auf die Benutzung selbst. Er zitiert

in diesem Zusammenhang eine Studie von Roberts und Moran (1983) in der „ease of use“ (Leichtigkeit der Benutzung, Übers. d. Verf.) und „ease of learning“ (Leichtigkeit des Erlernens, Übers. d. Verf.) eine signifikante Korrelation von 0,79 aufweisen.

Davis postuliert zwei Wirkmechanismen von PEOU auf die *Nutzungsintention*. Der eine Einfluss führt über eine gesteigerte Wirksamkeitseinschätzung des Benutzers. Eine als leichter benutzbar wahrgenommene Informationstechnologie führt zu einer höheren Erfolgserwartung für die Benutzung. Erfolgserwartung ist nach Bandura (1982) einer der Hauptfaktoren für intrinsische Motivation. Zum anderen hat die Bedienbarkeit Einfluss auf die Nützlichkeit einer Informationstechnologie. Ein leichter bedienbares System bindet weniger Ressourcen des Benutzers. Die freien Ressourcen können die inhaltliche Arbeit mit dem System verbessern und/oder beschleunigen und so die Produktivität der Informationstechnologie steigern. Somit hat PEOU auch einen indirekten Einfluss auf die *Nutzungsintention* über PU.

Entwicklung der PEOU-Skala

Für diese Arbeit ist ein genaues Verständnis des Konstrukts PEOU und des Prozesses der zugehörigen Skalenentwicklung von zentraler Bedeutung. Dies liegt daran, dass sich in der Forschung zum TAM (die im Kapitel „Stand der Forschung zum TAM“ detailliert diskutiert wird) ein stark schwankender Einfluss des Konstrukts zeigt, dessen Ursprung in dieser Arbeit wesentlich im Prozess der Skalenentwicklung vermutet wird. Entsprechend folgt an dieser Stelle eine ausführliche Analyse von Davis (1989) Skalenentwicklung. Diese zeigt, wie sich die Skalen inhaltlich zunehmend von der ursprünglichen Konstruktdefinition entfernt haben.

Was die methodischen Grundlagen zur Skalenentwicklung angeht, wird im Folgenden auf Kapitel Skalenentwicklung im empirischen Teil dieser Arbeit verwiesen. Die entsprechenden Methoden und Begriffe sind dort genauer beschrieben, da sie für den empirischen Teil der Arbeit entscheidend sind.

Das Konstrukt PEOU erfasst die wahrgenommene Leichtigkeit der Benutzung. Davis versteht Leichtigkeit als Gegenteil von Anstrengung, die sich sowohl auf das Erlernen wie auch auf das Benutzen beziehen kann. Nach Davis ist ein Informationsverarbeitungssystem also leicht zu benutzen, wenn sowohl die Benutzung leicht zu erlernen ist, als auch die Benutzung durch einen erfahrenen Benutzer ohne Anstrengung zu bewerkstelligen ist. Davis hat die erlebte Anstrengung unerfahrener und erfahrener Benutzer in einem gemeinsamen Konstrukt zusammengefasst auf der Grundlage der Ergebnisse von Roberts und Moran (1983). In deren Studie berechneten die Autoren eine hohe Korrelation von 0.79 zwischen Messwerten für „ease of learning“ (Leichtigkeit des Erlernens, Übers. d. Verf.) und „ease of use“ (Leichtigkeit der Benutzung, Übers. d. Verf.). Davis zitiert in diesem Zusammenhang außerdem Whiteside et al. (1985), die in ihrer Untersuchung keine Unterschiede in den Präferenzen von erfahrenen Computernutzern im Vergleich zu

Anfängern bezüglich der Bedienoberfläche feststellen konnten. Diesen Befund interpretiert Davis als Kongruenz zwischen der Leichtigkeit des Erlernens und der Leichtigkeit der Benutzung.

Um das Konstrukt PEOU besser zu verstehen, lohnt es sich, den Prozess der Skalenentwicklung näher zu betrachten, den Davis (1989) sehr detailliert beschreibt. Für die Konstruktion der Skala definierte Davis das Konstrukt PEOU als „the degree to which a person believes that using a particular system would be free from effort“ (der Grad zu dem eine Person glaubt, dass die Nutzung einer bestimmten Software nicht aufwendig ist, Übers. d. Verf.). Ausgehend von der Definition formulierte Davis zunächst 14 Items für das Konstrukt. In der folgenden Auflistung der Items wurde die Formulierung geringfügig abgeändert: Davis hatte die Items im Hinblick auf ein E-Mail-System formuliert („electronic mail system“), das Gegenstand seiner Untersuchung war; diese Formulierung ist im Folgenden durch „SOFTWARE“ als Platzhalter ersetzt, um die Kernaussage der Items unabhängig von ihrem Bezug auf ein konkretes System zu formulieren:

- I often become confused when I use the SOFTWARE.
(Ich werde oft verwirrt, wenn ich mit SOFTWARE arbeite, Übers. d. Verf.)
- I make errors frequently when using the SOFTWARE.
(Ich mache häufig Fehler, wenn ich mit SOFTWARE arbeite, Übers. d. Verf.)
- Interacting with the SOFTWARE is often frustrating.
(Die Interaktion mit SOFTWARE ist oftmals frustrierend, Übers. d. Verf.)
- I need to consult the user manual often when using SOFTWARE.
(Ich muss häufig das Benutzerhandbuch heranziehen, wenn ich SOFTWARE benutze, Übers. d. Verf.)
- Interacting with SOFTWARE requires a lot of mental effort.
(Die Interaktion mit SOFTWARE erfordert eine Menge geistiger Anstrengung, Übers. d. Verf.)
- I find it easy to recover from errors encountered while using SOFTWARE.
(Ich finde es leicht, während der Benutzung von SOFTWARE Fehler zu korrigieren, Übers. d. Verf.)
- The SOFTWARE is rigid and inflexible to interact with.
(SOFTWARE ist in der Interaktion starr und unflexibel, Übers. d. Verf.)
- I find it easy to get the SOFTWARE to do what I want it to do.
(SOFTWARE lässt sich leicht für meine Zwecke einsetzen, Übers. d. Verf.)
- The SOFTWARE often behaves in unexpected ways.
(SOFTWARE verhält sich häufig unerwartet, Übers. d. Verf.)
- I find it cumbersome to use SOFTWARE.
(Ich finde die Benutzung von SOFTWARE mühsam, Übers. d. Verf.)
- My interaction with SOFTWARE is easy for me to understand.
(Die Bedienung von SOFTWARE ist für mich leicht verständlich, Übers. d. Verf.)
- It is easy for me to remember how to perform tasks using the SOFTWARE.

(Ich kann mir leicht merken, wie ich Aufgaben mit SOFTWARE erledigen kann, Übers. d. Verf.)

- The SOFTWARE provides helpful guidance in performing tasks.
(SOFTWARE unterstützt mich bei meinen Aufgaben mit hilfreichen Anleitungen, Übers. d. Verf.)
- Overall, I find the SOFTWARE easy to use.
(Insgesamt finde ich SOFTWARE leicht zu benutzen, Übers. d. Verf.)

Um zu einer Auswahl aus diesen Items zu gelangen, ließ Davis die Items gemäß zweier verschiedener Vorgaben von 15 erfahrenen Computernutzern sortieren. Dieses Vorgehen ist eine modifizierte Version von Sherif und Sherifs (1967) „own category procedure“ (Methode [zum Auffinden] eigener Kategorien [einer Versuchsperson], Übers. d. Verf.). Die eine Sortierung sollte aufdecken, inwieweit das Item den Inhalt der Konstruktdefinition trifft. Hierfür wurden die Items in eine Rangreihenfolge sortiert, bei der die Items mit der größten inhaltlichen Übereinstimmung zur Konstruktdefinition ganz an das eine Ende, die Items mit der kleinsten Übereinstimmung ganz an das andere Ende sortiert wurden. Durch die zweite Sortierung sollte sichergestellt werden, dass relevante inhaltliche Teilaspekte des Konstrukts durch die Items der Skala abgedeckt wurden. Hierfür sollten die Items nach Ähnlichkeit zu drei bis fünf Stapeln gruppiert werden. Für die Auswahl der Items wurden die Ergebnisse beider Sortieraufgaben simultan betrachtet. Die Clusterung der Items auf drei bis fünf Stapel sollte aufzeigen, welche Teilaspekte des Konstrukts berücksichtigt werden mussten und dadurch verhindern, dass alle Items eines solchen Teilbereichs zugunsten anderer Items aus der Skala entfernt wurden. Die Rangreihenfolge bildete die Entscheidungsgrundlage bezüglich der Items innerhalb eines Stapels. Items wurden einem gemeinsamen Stapel zugeordnet, wenn sie von mindestens sieben Teilnehmern auf einen gemeinsamen Stapel sortiert worden waren. Im Ergebnis ergab sich für die Items eine Clusterung wie in Tab 1 dargestellt:

C l u s t e r A	<p>I find it easy to get the SOFTWARE to do what I want it to do. (SOFTWARE lässt sich leicht für meine Zwecke einsetzen.)</p> <p>I find it cumbersome to use SOFTWARE. (Ich finde die Benutzung von SOFTWARE mühsam.)</p> <p>The SOFTWARE is rigid and inflexible to interact with. (SOFTWARE ist in der Interaktion starr und unflexibel.)</p>
C l u s t e r B	<p>Interacting with the SOFTWARE is often frustrating. (Die Interaktion mit SOFTWARE ist oftmals frustrierend.)</p> <p>My interaction with SOFTWARE is easy for me to understand. (Die Bedienung von SOFTWARE ist für mich leicht verständlich.)</p> <p>Interacting with SOFTWARE requires a lot of mental effort. (Die Interaktion mit SOFTWARE erfordert eine Menge geistiger Anstrengung.)</p> <p>I often become confused when I use the SOFTWARE. (Ich werde oft verwirrt, wenn ich mit SOFTWARE arbeite.)</p>
C l u s t e r C	<p>It is easy for me to remember how to perform tasks using the SOFTWARE. (Ich kann mir leicht merken, wie ich Aufgaben in SOFTWARE erledigen kann.)</p> <p>I need to consult the user manual often when using SOFTWARE. (Ich muss häufig das Benutzerhandbuch heranziehen, wenn ich SOFTWARE benutze.)</p> <p>The SOFTWARE provides helpful guidance in performing tasks. (SOFTWARE unterstützt mich bei meinen Aufgaben mit hilfreichen Anleitungen.)</p>
ü b r i g e	<p>I make errors frequently when using the SOFTWARE. (Ich mache häufig Fehler, wenn ich mit SOFTWARE arbeite.)</p> <p>I find it easy to recover from errors encountered while using SOFTWARE. (Ich finde es leicht, während der Benutzung von SOFTWARE Fehler zu korrigieren.)</p> <p>The SOFTWARE often behaves in unexpected ways. (SOFTWARE verhält sich häufig unerwartet.)</p>
u n b e w e r t e t	<p>Overall, I find the SOFTWARE easy to use. (Insgesamt finde ich SOFTWARE leicht zu benutzen.)</p>

Tab 1: Clusterung der PEOU-Items bei Davis (1989)

Das erste Cluster interpretierte Davis als auf „physical effort“ (physische Anstrengung, Übers. d. Verf.) bezogen, während die Items des zweiten Clusters seiner Ansicht nach „mental effort“ (geistige Anstrengung, Übers. d. Verf.) erfassen. Zur Interpretation des dritten Clusters sagte Davis:

The third cluster is somewhat more difficult to interpret but appears to be tapping perceptions of how easy a system is to learn.

(Das dritte Cluster ist etwas schwieriger zu interpretieren, steht aber anscheinend in Bezug zur Wahrnehmung, wie leicht ein System zu erlernen ist, Übers. d. Verf.)

Die Zuordnung der übrigen vier Items durch Davis Versuchspersonen war so uneinheitlich, dass diese Items keinem weiteren Cluster zugeordnet werden konnten. Davis entfernte diese Items daraufhin aus dem Itempool für die PEOU-Skala. Die Formulierung des letzten Items orientiert sich eng an der Konstrukt-Definition, was für Davis eine ausreichende Begründung darstellt, dieses Item nicht zu testen.

Die Reihenfolge der Cluster ergab sich aus der Rangreihenfolge, in die die Items von den Teilnehmern sortiert wurden. Cluster A enthielt die Items, die die höchsten Rangplätze (Rang 1, 2 und 6) bezüglich ihrer Übereinstimmung mit der Konstrukt-Definition hatten. Die Items des Clusters B wurden auf Rang 3, 4, 5 und 7 sortiert und die Items des Clusters C auf Rang 8, 9 und 12. Die nicht geclusterten Items belegten in der Rangreihenfolge Rang 10, 11 und 13, was ebenfalls dafür sprach, sie aus der Skala zu entfernen.

In einem ersten Schritt zur Verkürzung und Verbesserung der Skala reduzierte Davis die Itemanzahl für jedes Cluster auf drei Items. Aus den vier Items des Clusters B wurde das Item

I often become confused when I use the SOFTWARE.

(Ich werde oft verwirrt, wenn ich mit SOFTWARE arbeite.)

aus der Skala entfernt, da es mit Rang 7 bezüglich seiner Übereinstimmung mit der Skalen-Definition innerhalb des Clusters am schlechtesten beurteilt worden war. Um den inhaltlichen Aspekt dieses Items nicht aus der Skala zu verlieren, formulierte Davis ein verbliebenes Item leicht um. Er änderte hierzu das Item

My interaction with SOFTWARE is easy for me to understand.

(Die Bedienung von SOFTWARE ist für mich leicht verständlich.)

in

My interaction with SOFTWARE is clear and understandable.

(Die Bedienung von SOFTWARE ist klar und verständlich.)

ab.

Zwei der Items aus Cluster C wurden von Davis umformuliert, da sie seiner Einschätzung nach den durch das Cluster abgedeckten Aspekt der Erlernbarkeit nur indirekt erfassten. Die Items

I need to consult the user manual often when using SOFTWARE.

(Ich muss häufig das Benutzerhandbuch heranziehen, wenn ich SOFTWARE benutze.)

und

The SOFTWARE provides helpful guidance in performing tasks.

(SOFTWARE unterstützt mich bei meinen Aufgaben mit hilfreichen Anleitungen.)

wurden durch

Learning to operate the SOFTWARE is easy for me

(Es ist leicht für mich, die Bedienung von Software zu erlernen, Übers. d. Verf.)

und

I find it takes a lot of effort to become skillful at using SOFTWARE.

Ich finde es anstrengend, ein fachkundiger Benutzer von SOFTWARE zu werden,
Übers. d. Verf.)

ersetzt. Davis nahm an, dass die neuen Items ebenfalls zu Cluster C gehören würden, obwohl dies von ihm nicht empirisch überprüft wurde. Mit den vorgenommenen Veränderungen ergab sich eine Skala mit zehn Items, die Davis in einer ersten Studie einsetzte. Die Items der Skala und ihre inhaltliche Zuordnung zu den Clustern ist in Tab 2 abgebildet.

C l u s t e r A	<p>I find it easy to get the SOFTWARE to do what I want it to do. (SOFTWARE lässt sich leicht für meine Zwecke einsetzen.)</p> <p>I find it cumbersome to use SOFTWARE. (Ich finde die Benutzung von SOFTWARE mühsam.)</p> <p>The SOFTWARE is rigid and inflexible to interact with. (SOFTWARE ist in der Interaktion starr und unflexibel.)</p>
C l u s t e r B	<p>Interacting with the SOFTWARE is often frustrating. (Die Interaktion mit SOFTWARE ist oftmals frustrierend.)</p> <p>My interaction with SOFTWARE is clear and understandable. (Die Bedienung von SOFTWARE ist klar und verständlich.)</p> <p>Interacting with SOFTWARE requires a lot of mental effort. (Die Interaktion mit SOFTWARE erfordert eine Menge geistiger Anstrengung.)</p>
C l u s t e r C	<p>It is easy for me to remember how to perform tasks using the SOFTWARE. (Ich kann mir leicht merken, wie ich Aufgaben in SOFTWARE erledigen kann.)</p> <p>I find it takes a lot of effort to become skillful at using SOFTWARE. (Ich finde es anstrengend, ein fachkundiger Benutzer von SOFTWARE zu werden.)</p> <p>Learning to operate the SOFTWARE is easy for me (Es ist leicht für mich, die Bedienung von Software zu erlernen.)</p>
u n b e w e r t e t	<p>Overall, I find the SOFTWARE easy to use. (Insgesamt finde ich SOFTWARE leicht zu benutzen.)</p>

Tab 2: PEOU-Skala für Davis erste Studie (1989)

Davis nutzte die Daten seiner ersten Studie, um die Skalen weiter zu entwickeln. Sein Hauptaugenmerk lag hierbei auf einer Verkürzung der Skalen, um sie in praxisnaher Forschung einsetzen zu können, wo lange Skalen wegen der zur Bearbeitung benötigten Zeit häufig zu Akzeptanzproblemen bei den Untersuchungsteilnehmern führen. Er analysierte die Items gemäß dem „Multi-Trait-Multi-Method“-Ansatz (MTMM-Ansatz, Mehrfache Eigenschaften, mehrfache Methoden-Ansatz, Übers. d. Verf.) von Campbell und Fiske (1959).

Dieser Ansatz sollte bei der Beurteilung der Validität und Reliabilität (siehe Kapitel Skalenentwicklung) verschiedener Messverfahren helfen. Nach Campbell und Fiske sollen die Messergebnisse verschiedener Messverfahren (z. B. Fragebogen und Verhaltensbeobachtung) für verschiedene Konstrukte (z. B. Computerkenntnisse und

Intelligenz) an verschiedenen Gruppen (z. B. Computerexperten und Computerlaien) korreliert werden. Die Messwerte der verschiedenen Messverfahren (also Fragebogen und Verhaltensbeobachtung) für dasselbe Konstrukt (z. B. Computerkenntnisse) an derselben Gruppe (z. B. Computerexperten) sollen hoch korrelieren. Eine hohe Korrelation deutet auf eine – erwünschte – Vergleichbarkeit der Messergebnisse hin. Hingegen sollen die Messergebnisse eines Messverfahrens (z. B. Fragebogen) für verschiedene Konstrukte (z. B. Computerkenntnisse und Intelligenz) an derselben Gruppe (z. B. Computerexperten) nicht hoch korrelieren. Ebenso sollen auch die Messergebnisse eines Messverfahrens (z. B. Fragebogen) für ein Konstrukt (z. B. Computerkenntnisse) über verschiedene Gruppen (z. B. Computerexperten und Computerlaien) hinweg nicht hoch korrelieren. Wären die letzteren Korrelationen hoch, würde dies auf eine unzureichende Abgrenzung der Konstrukte (im Beispiel Computerkenntnisse und Intelligenz) bzw. eine mangelnde Fähigkeit des Messverfahrens, Unterschiede der Gruppen (hier zwischen Computerexperten und Computerlaien) abzubilden.

Davis betrachtete die einzelnen Items seiner Skala als eigenständige Messverfahren im Sinne des MTMM-Ansatzes. Entsprechend überprüfte er das Muster der Interitemkorrelation in verschiedener Weise. Eine hohe Korrelation der Items innerhalb eines Konstrukts weist auf eine inhaltliche Ähnlichkeit der Items hin, was für die Homogenität einer Skala erwünscht ist. Dies entspricht der Betrachtung der Korrelationen verschiedener Messverfahren (Fragebogen und Verhaltensbeobachtung) für dasselbe Konstrukt (Computerkenntnisse) und dieselbe Gruppe (Computerexperten) des MTMM-Ansatzes. Die Items der PEOU-Skala erfüllten hinsichtlich dieser Analyse die Anforderungen: von 90 betrachteten Interitemkorrelationen waren 86 signifikant ($p < 0,05$). In einer weiteren Analyse widmete sich Davis den Korrelationen der Messwerte eines Items für eine Software (ein E-Mail-System) mit dem Messwert desselben Items für eine andere Software (ein Texteditor). Dies entspricht der Analyse der Korrelationen der Messwerte eines Messverfahrens (Fragebogen) für verschiedene Gruppen (Computerexperten und Computerlaien). Abschließend betrachtete Davis die Korrelationen des Messwerts eines Items an einer Gruppe (z. B. dem E-Mail-System) mit dem Messwert eines anderen Items für die andere Gruppe (einem Texteditor). Dies entspricht der Analyse der Korrelationen der Messwerte verschiedener Messverfahren (Fragebogen und Verhaltensbeobachtung) für verschiedene Gruppen (Computerexperten und Computerlaien). Bei den beiden zuletzt genannten Analysen sollten keine hohen Korrelationen auftreten, da solche auf Messartefakte zurückzuführen sind. Wichtiger als die absoluten Höhen der Korrelationen ist gemäß Campbell und Fiske, dass die Korrelationen der Messwerte verschiedener Messverfahren für dieselbe Gruppe höher sind als die Korrelationen desselben Messverfahrens oder unterschiedlicher Messverfahren über verschiedene Gruppen. Für Davis Items bedeutet dies, dass die Interitemkorrelationen verschiedener Items seiner Skalen für dieselbe Software höher sein sollten als die Korrelationen dieses Items mit sich selbst bei wechselnder

Software oder die Korrelation dieses Items mit einem anderen Item bei der anderen Software. Für die PEOU-Skala wurde diese Forderung bei 1800 Korrelationsvergleichen in 58 Fällen nicht erfüllt. Diese geringe Zahl von Regelabweichungen interpretiert Davis als Hinweis auf die guten Eigenschaften der Skala.

Eine genauere Analyse, welche Items die Forderungen gemäß dem MTMM-Ansatz nicht erfüllen, zeigte, dass es sich hierbei um die negativ formulierten Items handelte. Als Konsequenz eliminierte Davis die negativ formulierten Items aus der Skala, um so die Skaleneigenschaften zu verbessern. Von den in der Skala verbliebenen Items stammt je eines aus Cluster A, Cluster B und Cluster C. Eines dieser Items wurde im Hinblick auf Erlernbarkeit neu formuliert. Dieses Item wurde von Davis als zu Cluster C gehörig betrachtet. Das fünfte Item wurde bei der Clusterung nicht mitbetrachtet und wird von Davis als „overall item“ (übergreifendes Item, Übers. d. Verf.) bezeichnet. Das zu Cluster C gehörige, positiv formulierte Item

It is easy for me to remember how to perform tasks using the SOFTWARE.

(Ich kann mir leicht merken, wie ich Aufgaben mit SOFTWARE erledigen kann.)

ersetzt Davis durch eine umgepolte Version des Items

I find it takes a lot of effort to become skillful at using SOFTWARE.

(Ich finde es anstrengend, ein fachkundiger Benutzer von SOFTWARE zu werden.)

weil dieses Item explizit formuliert wurde, um Erlernbarkeit zu erfassen. Die umgepolte Version lautet

It is easy for me to become skillful at using SOFTWARE.

(Es ist leicht für mich, ein fachkundiger Benutzer von SOFTWARE zu werden.)

Damit die drei Cluster angemessen in der Skala enthalten sind, polt Davis ein Cluster A zugehöriges Item um und behält es in der Skala. Aus

The SOFTWARE is rigid and inflexible to interact with.

(SOFTWARE ist in der Interaktion starr und unflexibel.)

wird entsprechend in der umgepolten Version

I find SOFTWARE to be flexible to interact with.

(Ich finde SOFTWARE in der Interaktion flexibel.)

Die im Anhang von Davis (1989) Artikel publizierte Fassung der Skala für PEOU besteht aus den folgenden sechs Items:

- Learning to operate the SOFTWARE is easy for me.
(Es ist leicht für mich, die Bedienung von Software zu erlernen.)
- I find it easy to get the SOFTWARE to do what I want it to do.
(SOFTWARE lässt sich leicht für meine Zwecke einsetzen.)
- My interaction with SOFTWARE is clear and understandable.
(Die Bedienung von SOFTWARE ist klar und verständlich.)

- I find SOFTWARE to be flexible to interact with.
(Ich finde SOFTWARE in der Interaktion flexibel.)
- It is easy for me to become skillful at using SOFTWARE.
(Es ist leicht für mich, ein fachkundiger Benutzer von SOFTWARE zu werden.)
- I find the SOFTWARE easy to use.
(Ich finde SOFTWARE leicht zu benutzen.)

Der Prozess der Skalenentwicklung ist von Davis sehr detailliert dokumentiert und jeder seiner Schritte ist gut begründet. Dennoch hat sich der inhaltliche Fokus der Skala von der ersten Itemsammlung bis zur Endfassung immer stärker in Richtung Erlernbarkeit verschoben. Gemäß der Rangreihenfolge, wie gut die Items zur Konstrukt-Definition passen, belegten die Items des von Davis mit Erlernbarkeit bezeichneten Clusters niedrigere Plätze als die Items der beiden anderen Cluster. Zusätzlich bemerkte Davis, dass die inhaltliche Interpretation des dritten Clusters schwierig sei, die Gemeinsamkeit der Items war nicht offensichtlich. Nachdem Davis das Cluster inhaltlich auf Erlernbarkeit festgelegt hatte, formulierte er neue Items, die dieser inhaltlichen Ausrichtung stärker entsprechen sollten als die ursprünglichen Items. Obwohl bei der Analyse der Daten seiner ersten Studie eines der ursprünglichen Items des Clusters alle Bedingungen für den Verbleib in der Skala erfüllte, ersetzte Davis es durch ein Item, das erst nach der Clusterung formuliert worden war und umgepolt werden musste. Davis argumentierte damit, das Item

It is easy for me to become skillful at using SOFTWARE.

(Es ist leicht für mich, ein fachkundiger Benutzer von SOFTWARE zu werden.)

würde den Aspekt der Erlernbarkeit besser erfassen als das ursprüngliche Item

It is easy for me to remember how to perform tasks using the SOFTWARE.

(Ich kann mir leicht merken, wie ich Aufgaben mit SOFTWARE erledigen kann.)

Als Resultat sind in der letzten Fassung der PEOU-Skala noch zwei Items des Clusters A enthalten, wovon eines invertiert wurde. Aus Cluster B ist ein Item enthalten. Die Betrachtung bezüglich Cluster C ist schwierig. Von den ursprünglich als Cluster C sortierten Items ist keines in der Endfassung enthalten. Die Skala enthält stattdessen zwei neu formulierte Items, die die wahrgenommene Erlernbarkeit einer Software erfassen sollen. Zuletzt ist noch ein „übergreifendes“ Item in der Skala enthalten, welches von der Clusterung und der Sortierung in die Rangreihenfolge ausgenommen wurde und entsprechend diesbezüglich nicht eingeordnet werden kann.

Die Betrachtung des Entwicklungsprozesses der PEOU-Skala zeigt eine Veränderung der Operationalisierung hin zu einer höheren Gewichtung des Einflusses der Erlernbarkeit. Davis hat bei der Definition des Konstrukts nicht zwischen Erlernbarkeit und Benutzbarkeit unterschieden, sondern diese beiden Aspekte der Benutzung in einem Konstrukt zusammengefasst. Die Vermutung liegt nahe, dass die in dieser Hinsicht unklare Definition des Konstrukts PEOU

verantwortlich ist für die uneinheitlichen Befunde zum Einfluss dieses Konstrukts in der TAM-Forschung.

Stand der Forschung zum TAM

Seit Davis das TAM 1986 in seiner Dissertation vorgestellt hat, wurde es als theoretische Grundlage vieler Untersuchungen genutzt. Im Folgenden wird ein Überblick über wichtige Untersuchungen auf Basis des TAM gegeben. Dabei wurden die vom TAM postulierten Zusammenhänge überprüft und verschiedene Erweiterungen des TAM getestet. Die ersten fünf Studien beschäftigen sich ausschließlich mit dem TAM und bringen keine neuen Konstrukte ins Spiel. Die darauffolgenden Studien erweitern das TAM um zusätzliche Konstrukte bzw. verändern in einem Fall die Konstrukte des TAM selbst. Hervorzuheben ist die nachfolgend zusammengefasste Studie von Davis, Bagozzi und Warshaw (1989), deren Ergebnisse zu einer Vereinfachung des TAM führten, die sich durchgesetzt hat. In dieser Studie hatte das aus der Theory of Reasoned Action (TRA) stammende, vermittelnde Einstellungskonstrukt einen unbedeutenden Erklärungswert. Als Konsequenz wurde es aus der Theorie entfernt, so dass die TAM-Konstrukte PU und PEOU fortan die *Nutzungsintention* direkt beeinflussen und nicht mehr über die Einstellung vermittelt.

Experimentelle Überprüfungen zum TAM

Im Folgenden werden fünf Studien zum TAM vorgestellt. Der Schwerpunkt dieser fünf Studien lag auf der Überprüfung der Zusammenhänge des TAM selbst, nicht auf einer Erweiterung bzw. Veränderung des Modells mit Konstrukten auf derselben Hierarchieebene wie PU und PEOU oder *Nutzungsintention*. Die Studien bestätigen die Grundannahmen des TAM, weisen aber Unterschiede auf bezüglich des Zusammenhangs von PEOU und *Nutzungsintention*.

Davis, Bagozzi und Warshaw (1989) verglichen das TAM mit der TRA. In der TRA existiert neben den Einstellungen – die das TAM als PU und PEOU modelliert – noch die soziale Norm als Vorläufer einer Verhaltensintention. Davis et al. (1989) untersuchten die Einstellung von Studierenden zur Nutzung eines Textverarbeitungsprogramms im Rahmen des Studiums. Die Studierenden erhielten eine kurze Einweisung in das Programm WriteOne mit einer anschließenden schriftlichen Befragung. 14 Wochen später wurde eine zweite schriftliche Befragung durchgeführt, in der neben den unabhängigen Variablen der beiden Modelle auch das Nutzungsverhalten abgefragt wurde. Die statistische Auswertung der Daten bestätigte die gemeinsame Grundannahme beider Modelle, dass das tatsächliche Nutzungsverhalten am besten durch die *Nutzungsintention* vorhergesagt werden kann. Das Nutzungsverhalten zum zweiten Messzeitpunkt korrelierte mit 0.35 mit der *Nutzungsintention* der ersten Messung. Beide Modelle konnten einen signifikanten Anteil der Varianz der *Nutzungsintention* erklären. Für den ersten Messzeitpunkt

konnten die Konstrukte Einstellung und soziale Norm der TRA 32 % der Varianz der *Nutzungsintention* erklären. Die Konstrukte des TAM erklärten für diesen Messzeitpunkt 47 % der Varianz der *Nutzungsintention*. Für den zweiten Messzeitpunkt verschlechtert sich die Varianzaufklärung der TRA-Konstrukte auf 26 %, während die TAM-Konstrukte 51 % Varianzaufklärung erreichen.

Da beide Modelle die gleiche Anzahl von Prädiktoren verwenden, kann die erklärte Varianz direkt verglichen werden und eine höhere Varianzaufklärung deutet auf ein besseres Modell hin. In einer detaillierten Analyse der Daten zeigte die *subjektive Norm* zu keinem Messzeitpunkt einen signifikanten Einfluss auf die Verhaltensintention. Das dieser Untersuchung zugrunde liegende TAM postulierte eine Wirkung von PU und PEOU über die Einstellung auf die *Nutzungsintention*. Die vermittelnde Rolle der Einstellung konnte nicht für beide Messzeitpunkte bestätigt werden, für den zweiten Messzeitpunkt wurde der Einfluss der Einstellung auf die *Nutzungsintention* im Rahmen des TAM nicht signifikant. Der Einfluss von PEOU veränderte sich stark über die beiden Messzeitpunkte. Zum ersten Messzeitpunkt hatte PEOU einen signifikanten direkten Einfluss auf die *Nutzungsintention*, der bei der zweiten Messung verschwindet. Hier wirkt PEOU nur noch indirekt über PU auf die *Nutzungsintention*. Zusammenfassend kommen die Autoren zu einem vereinfachten Modell mit drei Komponenten: PU, PEOU und *Nutzungsintention*. Dieses Modell kann 45 % der Varianz des ersten Messzeitpunkts und 57 % der Varianz des zweiten Messzeitpunkts aufklären. Die Variablen der TRA haben keinen ausreichenden Erklärungswert. Die *subjektive Norm* kann keinen signifikanten Beitrag zur Erklärung der Varianz leisten. Die Bedeutung der Einstellung als Vermittler zwischen PU bzw. PEOU sowie der *Nutzungsintention* ist uneinheitlich und insgesamt ist der Einfluss der Einstellung zu gering, um einen Verbleib dieses Konstrukts in der Modellierung zu rechtfertigen. Interessant ist der Einfluss von PEOU auf die *Nutzungsintention*. Beim ersten Messzeitpunkt zeigt sich ein starker direkter Einfluss auf die *Nutzungsintention*, der zum zweiten Messzeitpunkt durch einen schwachen Einfluss über PU ersetzt wird. Für die Autoren erklärt sich dies durch die zunehmende Erfahrung der Teilnehmer mit dem System. Zum ersten Messzeitpunkt erfasst PEOU eine maßgeblich auf einer Selbsteinschätzung basierende Lernerfolgswahrscheinlichkeit. Mit zunehmender Systemerfahrung wird diese abgelöst durch eine Einschätzung, inwieweit der tatsächliche Aufwand der Systembenutzung die Leistungsfähigkeit beeinflusst.

Die aus dieser Studie resultierende Form eines vereinfachten TAM hat sich gegenüber der 1986 von Davis in seiner Dissertation vorgestellten durchgesetzt. In den meisten Untersuchungen, die sich auf das TAM als theoretischen Rahmen berufen, wird diese Version benutzt. Dies gilt auch für die vorliegende Arbeit.

Das methodische Vorgehen bei der Skalenentwicklung des TAM haben Segars und Grover (1993) genauer analysiert. Sie griffen hierfür auf die Daten einer Studie von Adams, Nelson und Todd (1992) zurück, die versuchten, Davis (1989)

Untersuchungen zu replizieren, dabei aber zu abweichenden Ergebnissen kamen. Segars und Grover kritisierten den Einsatz einer Faktorenanalyse zur Bestätigung der Konstrukte aus den erhobenen Daten. Nach Segars und Grover eignet sich eine Faktorenanalyse nur zur explorativen Analyse von Daten und zum Entdecken möglicher Strukturen in den Daten. Die Überprüfung des Vorhandenseins der vermuteten Strukturen kann eine Faktorenanalyse hingegen nicht leisten. Hierzu musste ein Strukturgleichungsmodell erstellt werden, was die Autoren für die Daten von Davis (1989) Untersuchung auch durchführten. Es zeigte sich eine unbefriedigende Abbildung der Daten durch die von Davis postulierten Faktoren PU und PEOU. Die Autoren veränderten die Modellierung und kamen zu einer Struktur mit 3 Faktoren. Zusätzlich zu Davis Konstrukten PU und PEOU fanden Segars und Grover einen weiteren Faktor, den sie mit „Effectiveness“ (Effektivität, Übers. d. Verf.) bezeichneten. Zwei Items von Davis konnten der neuen Faktorenstruktur nicht zugeordnet werden und wurden aussortiert. Die Zuordnung der verbliebenen Items zu den drei Skalen wurde wie folgt vorgeschlagen (SOFTWARE steht in den Itemformulierungen als Platzhalter für die konkrete Software, die Gegenstand der Befragung ist):

- Skala für PU
 - Indem ich SOFTWARE nutze, kann ich meine Aufgaben leichter erledigen.
 - Ich finde SOFTWARE nützlich.
 - Indem ich SOFTWARE nutze, habe ich bessere Möglichkeiten, meine Aufgaben zu erledigen.
- Skala für PEOU
 - SOFTWARE ist leicht zu benutzen.
 - Die Bedienung von SOFTWARE ist für mich leicht zu erlernen.
 - Es fällt mir leicht, ein erfahrener Nutzer von SOFTWARE zu werden.
- Skala für Effektivität
 - Indem ich SOFTWARE nutze, kann ich meine Aufgaben effizienter erledigen.
 - Indem ich SOFTWARE nutze, kann ich meine Aufgaben besser erledigen.

Nicht zugeordnet werden folgende beiden Items:

- Die Bedienung von SOFTWARE ist klar und verständlich.
- Indem ich SOFTWARE nutze, kann ich meine Aufgaben schneller erledigen.

Die Autoren sehen den Erklärungswert von Davis Modell nicht generell in Frage gestellt. Sie betonen die Abhängigkeit der Konstrukte von der eingesetzten Technologie, den Benutzern und dem organisatorischen Umfeld. Es fällt auf, dass die Definition der von Davis formulierten Skalen sich nicht klar in den erhobenen Messwerten wiederfindet.

Für die wissenschaftliche Evaluation der Überarbeitung einer Konfigurationssoftware nutzten Keil, Beranek und Konsynski (1995) das TAM als

theoretischen Hintergrund. Die „Config“ genannte Konfigurationssoftware sollte das Verkaufspersonal eines Computerhändlers dabei unterstützen, nur vollständig konfigurierte Systeme zu verkaufen. Kurz nach der Einführung der Software wurde eine interne Studie durchgeführt, bei der 75 % der Mitarbeiter angaben, Config zu kennen und ausprobiert zu haben. Allerdings nutzten nur 25 % der Mitarbeiter die Software regelmäßig. Einer weiteren internen Analyse zufolge war die schlechte Gestaltung der Bedienoberfläche verantwortlich für die geringe Nutzung der Software. Die Software wurde unter Beteiligung der Benutzer weiterentwickelt, wobei der Schwerpunkt auf der Neugestaltung der Bedienoberfläche lag. Die Rückmeldungen zum neuen Entwurf der Bedienoberfläche im Prototypen-Stadium durch die Benutzer waren positiv. Vor und nach der Einführung der überarbeiteten Version von Config wurden PU und PEOU mit den Skalen von Davis (1989) erhoben. Kovarianzanalysen zeigten einen signifikanten Einfluss von PU und PEOU auf die selbst berichtete Systemnutzung. Der Regressionskoeffizient zwischen PEOU und der selbst berichteten Nutzung wurde allerdings nicht signifikant. Ein Modellvergleich zwischen einem Modell mit den beiden unabhängigen Variablen PU und PEOU und einem Modell nur mit PU wurde nicht gerechnet. Interessant sind die Auswirkungen der Überarbeitung auf die beiden Konstrukte PU und PEOU. Die Autoren erwarteten keinen Einfluss der Überarbeitung auf PU, da die Funktionalität der Software nicht verändert wurde, dagegen aber eine Verbesserung von PEOU durch die Überarbeitung der Bedienoberfläche. Tatsächlich veränderte sich PU, wobei die Funktionalität als signifikant schlechter wahrgenommen wurde. Bei PEOU und der selbst berichteten Benutzung kam es nicht zu signifikanten Veränderungen der Messergebnisse durch die Überarbeitung.

Die Veränderungen von PU durch die Überarbeitung erklären die Autoren posthoc durch eine unzutreffende Vorstellung der Software-Entwickler vom Konfigurationsprozess. Die Konzeption der Software geht von einem linearen Ablauf bei der Konfiguration eines Computersystems aus. Tatsächlich ist der Verlauf iterativ mit häufigen Rücksprüngen und Veränderungen an schon konfigurierten Details. Die Überarbeitung von Config änderte nichts an der Software bezüglich der Steuerbarkeit und entsprechend war auch die überarbeitete Software für ein praxismgerechtes – und damit iteratives – Konfigurieren schlecht geeignet. Für die Posthoc-Erklärung der Befunde zu PEOU diskutieren die Autoren vier Hypothesen:

- die Verbesserungen der Bedienoberfläche wurden nicht wahrgenommen
- Lerneffekte haben die Verbesserungen überdeckt
- die Messskalen waren zu unempfindlich
- PEOU hängt von der Anpassung des Werkzeugs an die Aufgabenstellung ab

Die Autoren verwerfen die erste Hypothese als Erklärung, da viele Benutzer eine klare Präferenz für die neue Bedienoberfläche äußerten. Die zweite mögliche Erklärung erscheint den Autoren unwahrscheinlich aufgrund des späten Zeitpunkts der zweiten Befragung. Diese erfolgte erst einen Monat nach der Einführung der

überarbeiteten Variante, so dass Lernprozesse keine entscheidende Rolle mehr spielen sollten. Um die Hypothese bezüglich der unzureichenden Empfindlichkeit der Messskalen zu überprüfen, betrachteten die Autoren die Veränderung der Skalenwerte in Abhängigkeit vom Nutzungsverhalten. Für die Benutzer, die das alte System nicht benutzt hatten, die überarbeitete Version aber nutzten, zeigte sich eine signifikante Verbesserung von PEOU. Dies werten die Autoren als klaren Befund gegen die Hypothese der mangelnden Skalenempfindlichkeit.

Die Autoren präferieren die vierte Hypothese als Erklärung für ihre Befunde. Die unzureichende Eignung der Software für die Aufgaben der Vertriebsmitarbeiter hat demnach die Wahrnehmung der erzielten Verbesserungen der Bedienoberfläche verhindert. Nach Keil et al. ist die Eignung eines Werkzeugs zur Bearbeitung einer Aufgabe die fundamentalste Anforderung an Usability. PEOU ist nicht bedeutsam für die Nutzung eines Systems, solange diese grundlegende Bedingung nicht gegeben ist.

Karahanna und Straub (1999) gingen der Frage nach, wovon die beiden Einstellungskonstrukte PU und PEOU des TAM abhängen. Dazu führten sie an 180 Mitarbeitern einer Transportfirma eine Fragebogenerhebung zum E-Mail-System durch. Die Autoren leiten eine Reihe von Konstrukten, die einen Einfluss auf PU und PEOU haben könnten, aus Studien zum Kommunikationsverhalten ab. Ein Konstrukt ist die „social presence“ (*soziale Anwesenheit*, Übers. d. Verf.), das beschreibt, inwieweit eine Kommunikationsform die physische Anwesenheit eines Kommunizierenden überträgt. Ein Telefonat, bei dem Stimmlage und Betonung des Sprechenden übertragen werden, hätte hiernach eine höhere *soziale Anwesenheit* im Vergleich zu einem Brief, der solche Informationen nicht übertragen kann. Laut den Autoren soll eine hohe *soziale Anwesenheit* die PU erhöhen, ein Einfluss auf PEOU wird nicht vermutet. Ein weiteres Konstrukt ist „social influence“ (*sozialer Einfluss*, Übers. d. Verf.). Dieses Konstrukt erfasst die Abhängigkeit der Kommunikation vom sozialen Umfeld. Die Autoren berichten von Studien, in denen ein Einfluss des Umfelds auf die Einstellung zur E-Mail-Kommunikation nachgewiesen werden kann. Auch für den sozialen Einfluss wird eine Wirkung auf die PU vermutet. Ein drittes Konstrukt der Autoren zur Erklärung von PU und PEOU ist die „accessibility“ (Zugänglichkeit, Übers. d. Verf.) eines Systems. Die *Zugänglichkeit* soll einen Einfluss auf PEOU haben. Schließlich soll die „availability of user training and support“ (Verfügbarkeit von Training und Unterstützung, Übers. d. Verf.) beide Konstrukte des TAM beeinflussen. Für alle Konstrukte wurden neue Skalen formuliert und an einer kleinen Stichprobe getestet. In der Studie wurde mit stark verkürzte Versionen von Davis Skalen gemessen. Die Skala für PU besteht nur aus zwei Items, die Skala für PEOU aus einem Item. Die Items sind von den Autoren nicht ausformuliert angegeben:

- Skala für PU
 - communicate easily (Kommunikation ist einfach, Übers. d. Verf.)

- communicate quickly (Kommunikation ist schnell, Übers. d. Verf.)
- Skala für PEOU
 - comfortable using E-mail (E-Mail ist komfortabel zu benutzen, Übers. d. Verf.)

Das theoretische Modell der Autoren zum Zusammenhang der TAM-Konstrukte postulierte keinen direkten Einfluss von PEOU auf PU. Zur statistischen Absicherung verglichen die Autoren mehrere Strukturgleichungsmodelle mit unterschiedlichen Abhängigkeiten zwischen den Konstrukten. PEOU hatte nur einen geringen, nicht signifikanten direkten Einfluss auf die Systemnutzung. PU hatte in allen Modellierungen einen signifikanten Einfluss auf die Systemnutzung. Wie von den Autoren vermutet, hatte PEOU einen signifikanten Einfluss auf PU, also einen indirekten Einfluss, wie ihn auch Davis (1989) gefunden hat. Von den neuen Konstrukten der Autoren hat die Verfügbarkeit von Training und Unterstützung weder einen signifikanten Einfluss auf PU noch auf PEOU. Die übrigen neuen Konstrukte beeinflussen PU und PEOU wie erwartet. Die *soziale Anwesenheit* und der *soziale Einfluss* wirken auf die PU. PEOU hängt von der *Zugänglichkeit* ab.

Die Ergebnisse von Karahanna und Straub bezüglich der TAM-Struktur sind wegen der gravierenden Veränderungen der Skalen schwer zu interpretieren. Es wird augenscheinlich nicht klar, warum die Autoren „communicate easily“ als PU-Item betrachten, und nicht, wie der Wortlaut nahe legt, als PEOU zugehörig. Insgesamt wird ein im Vergleich zu PU schwächerer, indirekter Einfluss von PEOU auf die Systemnutzung festgestellt. Dies steht im Einklang mit den Befunden von Davis (1989) Untersuchung. Die neuen Konstrukte sind sehr auf den Anwendungsfall eines Kommunikationsmediums zugeschnitten. Die Konstrukte *soziale Anwesenheit* und *sozialer Einfluss* können nicht auf andere Bereiche wie z. B. eine Textverarbeitung übertragen werden. Der Einfluss der *Zugänglichkeit* auf PEOU ist zwar statistisch signifikant, mit 7.8 % aufgeklärter Varianz allerdings nicht hoch.

Venkatesh und Davis (2000) erweiterten das TAM in ähnlicher Weise wie Karahanna und Straub. Venkatesh und Davis wollten die Vorbedingungen des TAM-Konstrukts PU untersuchen und erweiterten hierzu das TAM zu einem von ihnen TAM2 genannten Modell. Die Mehrheit der zusätzlichen Variablen des TAM2 sollen eine Vorhersage des Konstrukts PU ermöglichen, modellieren also einen vom TAM nicht abgedeckten Bereich. Diese neuen Konstrukte wirken von außerhalb des TAM direkt auf PU. Sie sind hinsichtlich einer Beurteilung des TAM nicht von Interesse, da sie das TAM selbst nicht verändern. Einzig für die aus der TRA stammende *subjektive Norm* vermuten die Autoren einen Einfluss sowohl auf PU wie auch auf die *Nutzungsintention*. Erst die Aufnahme dieses Konstrukts verändert die TAM-Struktur selbst, weil eine Wirkung parallel zu PU und PEOU auf die *Nutzungsintention* vermutet wird. Venkatesh und Davis unterstellen eine Abhängigkeit dieser Wirkung von der *Freiwilligkeit der Benutzung*. Erfolgt die Benutzung freiwillig, so erwarten sie keinen direkten Einfluss der *subjektiven Norm* auf die

Nutzungsintention. Besteht eine Nutzungsverpflichtung, so soll sich die *subjektive Norm* – man könnte in diesem Fall auch von Zwang sprechen – auf die *Nutzungsintention* auswirken.

Die Autoren erhoben Daten an vier verschiedenen Gruppen, von denen je zwei eine neue Software freiwillig nutzen konnten und die anderen beiden dazu verpflichtet waren. Für alle vier Gruppen wurden zu je drei Messzeitpunkten Daten erhoben. Die Messung erfolgt mit bewährten Skalen, für die TAM-Konstrukte greifen die Autoren auf die Skalen von Davis (1989) und Davis, Bagozzi & Warshaw (1989) zurück. Die Messungen erfolgten unmittelbar vor der System Einführung sowie einen und drei Monate später. Insgesamt wurden demnach pro Gruppe drei Messungen durchgeführt. Die Auswertung der Daten bestätigte frühere Ergebnisse zum TAM und zeigte über alle Gruppen und Messzeitpunkte hinweg einen deutlichen Einfluss von PU und einen schwächeren Einfluss von PEOU auf die *Nutzungsintention*. Der Einfluss der *subjektiven Norm* fällt wie erwartet aus: bei verpflichtender Systemnutzung hat die *subjektive Norm* einen direkten Einfluss auf die *Nutzungsintention*, der bei freiwilliger Nutzung nicht auftritt. Auch in dieser Untersuchung finden die Autoren eine indirekte Wirkung von PEOU auf die *Nutzungsintention* über das Konstrukt PU.

Die fünf betrachteten Studien basieren auf dem theoretischen Grundgerüst des TAM. Dieses wurde in den bisher betrachteten Studien nicht grundlegend verändert und maximal moderat erweitert. Bei der Auswertung der Daten wurden die vom TAM postulierten Zusammenhänge geprüft. Dabei zeigt sich ein einheitliches Bild für den Zusammenhang zwischen PU und der *Nutzungsintention*. Dieser Zusammenhang ist ausgeprägt und über verschiedene Rahmenbedingungen hinweg zu beobachten. Unklarer ist die Situation bezüglich des Konstrukts PEOU, dessen Einfluss auf die *Nutzungsintention* unterschiedlich stark ausfällt. Zusätzlich wirkt PEOU nicht ausschließlich direkt auf die *Nutzungsintention* sondern auch vermittelt über PU. Von den zusätzlich aufgenommen Variablen hat nur die *subjektive Norm* einen direkten Einfluss auf die *Nutzungsintention* und dies auch nur dann, wenn die Nutzung eines Systems vorgeschrieben ist.

Erweiterungen des TAM

Die nachfolgend beschriebenen Studien untersuchen das TAM teilweise für Anwendungen außerhalb der Arbeitswelt (meistens im Zusammenhang mit dem Internet) und verändern oder erweitern das theoretische Modell des TAM. Die Veränderungen des TAM sollen dessen Vorhersagekraft für die Nutzung moderner Informationsverarbeitungssysteme erhöhen.

Zusätzlich zu den TAM-Konstrukten PU und PEOU nehmen Davis, Bagozzi und Warshaw (1992) noch das Konstrukt „Enjoyment“ (*Vergnügen*, Übers. d. Verf.) aus dem Bereich der Motivation in ein theoretisches Modell der Technologieakzeptanz auf. Sie begründen die Aufnahme des neuen Konstrukts mit Erkenntnissen der

Motivationsforschung, die häufig zwischen „extrinsischer“ und „intrinsischer“ Motivation unterscheidet. Als extrinsisch wird Motivation dann verstanden, wenn sie auf einen Umgebungsfaktor aufbaut, also dem Erhalt einer Belohnung für ein gezeigtes Verhalten. Benutzt ein Büroangestellter ein E-Mail-Programm, weil er sich davon verbesserte Beförderungschancen verspricht, wäre dieses Verhalten entsprechend extrinsisch motiviert. Intrinsische Motivation liegt nach Davis et al. vor, wenn keine erkennbare äußere Belohnung existiert, das Ausführen der Tätigkeit also so angenehm ist, dass es selbst als Belohnung fungiert. Die Autoren setzen diese Dichotomie in die beiden Konstrukte PU und *Enjoyment* um. PU soll die extrinsische Motivation erfassen und *Enjoyment* die intrinsische. Einflüsse des Arbeitsumfelds sollen zwei weitere Konstrukte erfassen: die „Output Quality“ (*Qualität der Ausgabe*, Übers. d. Verf.) und „Task Importance“ (*Bedeutsamkeit der Aufgabe*, Übers. d. Verf.).

Die Autoren vermuten einen direkten Einfluss von PU und *Enjoyment* auf die *Nutzungsintention*. Der Einfluss von PEOU soll sowohl über PU als auch über *Enjoyment* vermittelt werden. Für den Einfluss von PEOU auf PU wird noch eine Abhängigkeit von der *Task Importance* vermutet: je höher die *Task Importance*, desto ausgeprägter der Einfluss von PEOU auf PU. Für das neue Konstrukt *Output Quality* werden die gleichen Einflusswege wie für PEOU erwartet. *Output Quality* soll also in Abhängigkeit von der *Task Importance* über PU auf die *Nutzungsintention* wirken und darüber hinaus immer auf das *Enjoyment*.

Die Datenerhebung erfolgte an zwei Studentengruppen, die in unterschiedliche Software eingewiesen wurden. Die erste Gruppe durchlief einen Einführungskurs für Textverarbeitung im Rahmen des Studiums und die zweite Gruppe wurde für den Versuch in der Benutzung einer Software zum Erstellen von Diagrammen unterrichtet. Für die erste Gruppe wurde das tatsächliche Nutzungsverhalten zusammen mit PU und PEOU am Ende des ersten Semesters gemessen. Eine Messung des Konstrukts *Task Importance* erfolgte nur für die zweite Gruppe, bei der allerdings die tatsächliche Nutzung nicht erhoben wurde. Die Daten konnten die von den Autoren aufgestellten Hypothesen im Wesentlichen bestätigen. Die Interpretation der Ergebnisse ist dabei schwierig, weil die Autoren mehrere Regressionsmodelle verwendet haben, allerdings keine Kennwerte für die Güte der Modellanpassung angeben. In einem Modell mit den vier Konstrukten PU, PEOU, *Enjoyment* und *Output Quality* werden in der ersten Studie nur die Regressionskoeffizienten zwischen PU bzw. *Enjoyment* und der *Nutzungsintention* signifikant. In der zweiten Studie wird nur die Regression von PU auf die *Nutzungsintention* signifikant. Die Autoren analysieren den Zusammenhang zwischen PU und PEOU bzw. *Output Quality* in weiteren Kovarianzanalysen. Dabei werden die Regressionskoeffizienten zwischen PU und PEOU bzw. *Output Quality* in beiden Studien signifikant. Die Regressionskoeffizienten zwischen *Enjoyment* und PEOU bzw. *Output Quality* erreichen in der ersten Studie alle ein signifikantes Niveau. In der zweiten Studie wird – abweichend von der ersten Studie – der Regressionskoeffizient

zwischen *Enjoyment* und *Output Quality* nicht signifikant.

Die Ergebnisse dieser Untersuchung sind bei näherer Betrachtung schwer zu interpretieren. Eine inhaltlich begründete Abgrenzung zwischen PU und *Output Quality* ist nicht ersichtlich. Die Ausgabemöglichkeiten eines Programms (*Output Quality*) gehören eindeutig zu dessen Funktionalität und sollten von Benutzern auch so wahrgenommen werden. Weshalb ein eigenständiges Konstrukt hierfür eingeführt wird, ist nicht nachvollziehbar. In den beiden Studien ergibt sich für die zusätzlichen Konstrukte kein klar erkennbarer Zusatzwert. Ein Vergleich zum klassischen TAM mit den beiden Konstrukten PU und PEOU findet in der Auswertung nicht statt. Ob der Erklärungszugewinn die zusätzliche Aufnahme der Variablen ins Modell rechtfertigt, bleibt offen. Für die Konstrukte PEOU und *Output Quality* zeigt sich keine eindeutige Befundlage. Die Ergebnisse der beiden Studien sind nicht einheitlich, was eine Interpretation erschwert. Die schwankende Bedeutung des TAM-Konstrukts PEOU erklären die Autoren mit einer unterschiedlichen Bedeutung der Erlernbarkeit in den beiden Studien. Sie vermuten eine Abhängigkeit des Einflusses von PEOU vom Lernprozess der Benutzer:

Future research should attempt to tease apart ease of use from ease of learning. Since our studies measured ease of use in the early phase of adoption, our ease of use measure may primarily be tapping ease of learning. However, it is unclear whether ease of use and ease of learning can be empirically distinguished, even long after adoption.

(Weitere Forschung sollte versuchen, eine Trennung zwischen Leichtigkeit der Benutzung und Leichtigkeit des Erlernens zu erreichen. Da unsere Studien Leichtigkeit der Benutzung in einer frühen Phase des Einsatzes erhoben, könnte unser Messinstrument für Leichtigkeit der Benutzung hauptsächlich Leichtigkeit des Erlernens erfasst haben. Allerdings ist es unklar, ob Leichtigkeit der Benutzung empirisch von Leichtigkeit des Erlernens unterschieden werden kann, insbesondere lange nach der Produkteinführung, Davis et al, 1992, Übers. d. Verf.)

Eine Untersuchung der Nutzung des Internets im theoretischen Rahmen des TAM stammt von Moon und Kim (2001). Da das Internet neben Informationen auch ein reichhaltiges Angebot an Unterhaltung bietet, erweitern die Autoren das theoretische Modell um ein Konstrukt, das den Einfluss der Unterhaltsamkeit auf die Einstellung erfassen soll. Das zusätzliche Konstrukt heißt „Perceived Playfulness“ (wahrgenommene Verspieltheit, Übers. d. Verf.) und soll wie auch PU und PEOU auf die *Nutzungsintention* wirken. Die Wirkung auf die *Nutzungsintention* wird nach Moon und Kim wieder über ein Einstellungskonstrukt vermittelt. Damit kehren die Autoren zur Form des TAM vor der Vereinfachung durch Davis, Bagozzi und Warshaw (1989) zurück, ohne dies allerdings inhaltlich zu begründen. Wie Davis, Bagozzi und Warshaw (1992) ihr Konstrukt *Enjoyment* verstehen die Autoren ihr Konstrukt *Perceived Playfulness* als Maß der intrinsischen Motivation.

Die Datenerhebung erfolgte mittels Fragebogen an Studenten, die das Internet für ihre Arbeit oder zur Unterhaltung nutzen konnten. Dabei kamen für PU und

PEOU Skalen zum Einsatz, die sich verbal an denen von Davis (1989) orientierten, dabei aber neu formulierte Items enthielten. Die Skala für Perceived Playfulness wurde komplett neu entwickelt und unvalidiert eingesetzt. Die Auswertung der Ergebnisse erfolgte durch mehrere Kovarianzanalysen. Die Autoren überprüften den direkten Zusammenhang zwischen PU bzw. Perceived Playfulness und der *Nutzungsintention*. Zusätzlich überprüften sie die Zusammenhänge zwischen PU, PEOU bzw. Perceived Playfulness und der Einstellung. Diese Überprüfungen ergaben durchgängig signifikante Regressionen. Eine Modellierung ohne das Konstrukt Einstellung, wie sie Davis, Bagozzi und Warshaw (1992) auf der Grundlage ihrer Befunde vorschlagen, haben die Autoren dieser Studie nicht berechnet. Die Aufnahme des Konstrukts Perceived Playfulness erhöht die Aufklärung der Varianz der abhängigen Variablen durch die unabhängigen von 35 % auf 39 %. Eine statistische Absicherung der Verbesserung des erweiterten Modells gegenüber dem originalen TAM wurde nicht durchgeführt. Es zeigte sich ein deutlicher Zusammenhang zwischen PEOU und der Einstellung in dieser Untersuchung. Die Regressionskoeffizienten zwischen PEOU und der Einstellung waren für beide Modellierungen (das erweiterte TAM und das originale TAM) höher als die der anderen Konstrukte (PU und – im erweiterten TAM – Perceived Playfulness). Die Autoren führten abschließend eine Analyse des Zusammenhangs von PU bzw. Perceived Playfulness und der *Nutzungsintention* durch. Hierbei unterschieden sie zwischen zwei vorgefundenen Nutzergruppen. Die erste Gruppe benutzte das Internet ausschließlich zu Unterhaltungszwecken, die zweite Gruppe nur zur Arbeit. Der Einfluss von PU und Perceived Playfulness hing deutlich vom Nutzungszweck ab. Für die Benutzer, die das Internet zur Unterhaltung nutzten, war PU nicht bedeutsam. Für die andere Gruppe waren beide Regressionskoeffizienten zwar signifikant, zwischen Perceived Playfulness und der *Nutzungsintention* aber deutlich geringer als zwischen PU und der *Nutzungsintention*.

Venkatesh, Morris, Davis und Davis (2003) nutzen die Daten aus vier Studien, um ein stark vom TAM abweichendes Modell der Technologieakzeptanz zu entwickeln. Die Autoren analysierten neben dem TAM sieben weitere Modelle zur Vorhersage der *Nutzungsintention* moderner Informationsverarbeitungstechnologie. Die Konstrukte der acht Modelle wurden auf der Grundlage von vier verschiedenen Erhebungen bezüglich ihres Erklärungswertes verglichen. In allen vier Erhebungen wurden die Skalen zu drei verschiedenen Zeitpunkten eingesetzt. Die erste Messung erfolgte direkt im Anschluss an ein Einführungstraining, die zweite Messung einen Monat später und die dritte Messung weitere drei Monate später. Für die gepoolten Daten erreicht das TAM die höchste Varianzaufklärung mit 53 %. Die theoretische Analyse der acht Modelle führte zu sieben Konstrukten, die über alle Modelle hinweg insgesamt zur Erklärung der *Nutzungsintention* dienten. Die sieben Konstrukte sind:

- Performance Expectancy (Leistungserwartung, Übers. d. Verf.)
- Effort Expectancy (Aufwandserwartung, Übers. d. Verf.)
- Social Influence (Sozialer Einfluss, Übers. d. Verf.)
- Facilitating Conditions (Erleichternde Umstände, Übers. d. Ver.)
- Computer self-efficacy (Computer-Selbstwirksamkeit, Übers. d. Verf.)
- Computer anxiety (Computer-Ängstlichkeit, Übers. d. Verf.)
- Attitude (Einstellung, Übers. d. Verf.)

Die Autoren erwarten aufgrund theoretischer Überlegungen einen direkten Einfluss der ersten drei Konstrukte auf die *Nutzungsintention*. Für *Erleichternde Umstände* wird kein Einfluss auf die *Nutzungsintention*, sondern eine direkte Wirkung auf die Nutzung erwartet. Für die letzten drei Konstrukte erwarten die Autoren keinen direkten Einfluss auf die *Nutzungsintention*. *Computer-Selbstwirksamkeit* und *Computer-Ängstlichkeit* beeinflussen die *Aufwandserwartung*. Es sind also Vorläufer der *Aufwandserwartung*, deren Einfluss durch die *Aufwandserwartung* erschöpfend abgebildet wird. Anders verhält es sich bezüglich der Einstellung. Das allgemeinere Konstrukt *Einstellung* umfasst zwar die Effekte der *Leistungserwartung* und der *Aufwandserwartung*, ist allerdings zu unspezifisch. Entsprechend sollen die enger definierten Konstrukte *Leistungserwartung* und *Aufwandserwartung* einen höheren Erklärungswert im Modell haben.

Zusätzlich definieren die Autoren noch eine Reihe von Zusatzkonstrukten, die den Einfluss der Hauptkonstrukte modellieren sollen. Dies sind das *Alter der Benutzer*, ihr *Geschlecht*, ihre *Erfahrung mit Computern* und ob die Nutzung des Systems freiwillig oder unfreiwillig erfolgt. Das neue Modell zur Erklärung der *Nutzungsintention* nennen die Autoren Unified Theory of Acceptance and Use of Technology (UTAUT, Einheitliche Theorie der Akzeptanz und Benutzung von Technologie, Übers. d. Verf.).

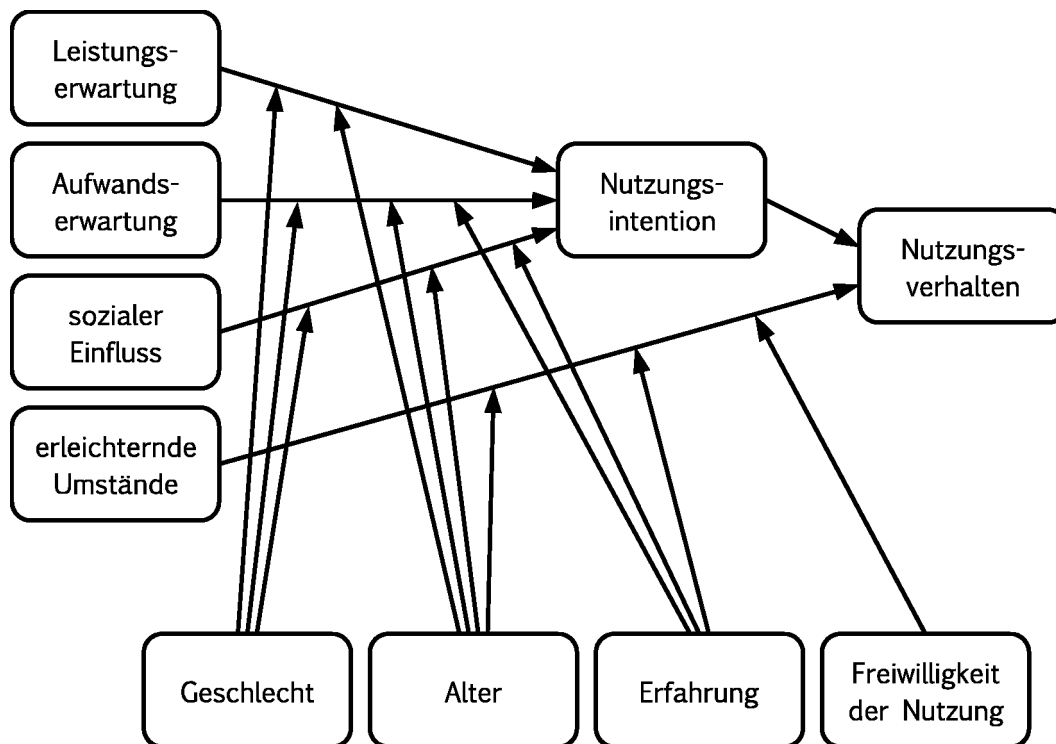


Abb. 4: Unified Theory of Acceptance and Use of Technology (UTAUT, nach Venkatesh, Morris, Davis & Davis, 2003)

Für die Entwicklung der Skalen der neu definierten Konstrukte griffen die Autoren auf Items der Skalen der acht betrachteten Modelle zurück. Eine faktoranalytische Betrachtung bestätigte die von Venkatesh et al. vermuteten sieben Konstrukte. Die Items wurden zu sieben neuen Skalen mit brauchbaren Skaleneigenschaften kombiniert. Insgesamt wurden in der Vorstudie 86 Items getestet und die Skalen bestanden aus bis zu 24 Items. Die Autoren reduzierten die Itemanzahl pro Skala auf vier Items, wobei die Items mit den höchsten Faktorenladungen ausgewählt wurden. Sieben der acht Items der resultierenden Skalen wurden ursprünglich von Davis (1989) für die Konstrukte PU und PEOU formuliert. Allerdings wurden drei dieser Items in Davis Arbeit aus der finalen Skala ausgeschlossen und fanden ihren Weg in diese Untersuchung über eine Arbeit von Moore und Benbasat (1991). Moore und Benbasat nutzten darin die Itemformulierungen von Davis, um Skalen für ihr Modell der „Innovation Diffusion Theorie“ (Theorie der Ausbreitung von Neuerungen, Übers. d. Verf.) zu entwickeln. Die Theorie der Ausbreitung von Neuerungen ist eins der von Venkatesh et al. betrachteten Modelle, und so wurden die von Davis für PU bzw. PEOU formulierten Items in diese Studie als Items für Konstrukte der Theorie der Ausbreitung von Neuerungen aufgenommen.

Zur statistischen Auswertung ihrer Daten berechnen die Autoren Kovarianzanalysen. Die verwendeten Modelle entsprechen nicht dem postulierten Modell der Einheitlichen Theorie der Akzeptanz und Benutzung von Technologie (UTAUT). Statt dessen wird für jeden Messzeitpunkt je ein Modell nur mit

Haupteffekten und ein Modell mit allen Interaktionseffekten bis zur vierten Ordnung berechnet. Bei der Modellierung nur mit Haupteffekten wird der Regressionskoeffizient von *Leistungserwartung* signifikant, zusätzlich zum ersten Messzeitpunkt der Regressionskoeffizient von *Aufwandserwartung*. Dieselben Ergebnisse finden sich in der Auswertung eines weiteren Datensatzes als Kreuzvalidierung. Die einfachen Modelle ohne Interaktion – also mit sieben Haupteffekten – erklärten zwischen 35 % und 38 % der Varianz der *Nutzungsintention*. Das volle Modell mit Interaktionseffekten bis zur vierten Ordnung und entsprechend 38 schätzbaren Parametern erklärte zwischen 45 % und 48 % der Varianz. Für eine abschließende Analyse poolten die Autoren die Daten über alle Messzeitpunkte, berücksichtigten bei der Modellierung aber nicht, dass pro Versuchsperson nun drei Messzeitpunkte im Datensatz enthalten waren und somit systematische Abhängigkeiten zwischen den Daten bestanden. In dieser Auswertung konnte das UTAUT 69 % bzw. 70 % der Varianz erklären, was die Autoren als Bestätigung ihrer theoretischen Ansätze interpretieren. Von den 38 Regressionskoeffizienten werden hierbei nur vier für den ersten Datensatz und zwei für den zweiten Datensatz signifikant. Darunter ist nur ein Regressionskoeffizient für einen Haupteffekt beim ersten Datensatz: die Wirkung der *Leistungserwartung* auf die *Nutzungsintention*. Alle anderen signifikanten Regressionen sind Wechselwirkungen zweiter oder höherer Ordnung, die sich inhaltlich nicht interpretieren lassen.

Aufgrund der Ergebnisse kann keine klare Überlegenheit des neuen UTAUT über das TAM festgestellt werden. Ein hierfür notwendiger Modellvergleich wurde nicht durchgeführt. Das neue Modell ohne Interaktionen entspricht weitgehend dem TAM. Die Skalen der neuen UTAUT-Konstrukte bestehen weitgehend aus PU- und PEOU-Items. In der Kovarianzanalyse ohne Interaktionen sind nur die Regressionskoeffizienten der PU und PEOU entsprechenden Konstrukte *Leistungserwartung* und *Aufwandserwartung* signifikant geworden. Interessant ist die Veränderung des Einflusses der *Aufwandserwartung*: für den ersten Messzeitpunkt ist die Regression signifikant, für die späteren Messungen nicht mehr. Der Einfluss der *Aufwandserwartung* auf die *Nutzungsintention* scheint also vom Lernfortschritt der Benutzer abhängig zu sein.

Für die Analyse der Akzeptanz eines internet-basierten Informationssystems für Studenten erweiterten Yi und Hwang (2003) das TAM um drei zusätzliche Konstrukte. Die Autoren greifen auf das bereits von Davis et al. (1992) eingeführte Konstrukt „Enjoyment“ (*Vergnügen*, Übers. d. Verf.) zurück. Zusätzlich erweitern sie das TAM um das aus Banduras (1977) sozial-kognitiver Theorie stammende Konstrukt der *Selbstwirksamkeit*. *Selbstwirksamkeit* ist die Überzeugung, eine zielführende Handlung erfolgreich umsetzen zu können. Das dritte Konstrukt stammt aus der Lernforschung und ist die *Zielorientierung* bei der Handlung. Dweck und Leggett (1988) unterschieden zwei mögliche Zielorientierungen: eine, die die eigene

Leistungsfähigkeit gegenüber anderen demonstrieren soll und eine, die an der möglichen Lernerfahrung interessiert ist.

Die Autoren postulieren eine Wirkung von *Enjoyment* auf PEOU. Dies steht der Annahme von Davis et al. (1992) diametral entgegen, die eine Wirkung von PEOU auf *Enjoyment* vermuteten. Auch die zur Begründung zitierte (und oben besprochene) Studie von Moon und Kim (2003) unterstützt diese These nicht. Für die *Selbstwirksamkeit* vermuten die Autoren eine direkte Wirkung sowohl auf PEOU wie auch auf die tatsächliche Systemnutzung. Die *Zielorientierung* soll sich auf die *Selbstwirksamkeit* auswirken, wobei eine höhere *Lernorientierung* zu einer höheren *Selbstwirksamkeit* führen soll.

Die Daten wurden an 109 Studenten erhoben, die ein internet-basiertes Informationssystem zur Verfügung hatten. Die Autoren führten eine Pfadanalyse durch, um die Daten ihren Hypothesen entsprechend zu modellieren. Alle Regressionskoeffizienten wurden den Hypothesen der Autoren entsprechend signifikant. PU und PEOU konnten 29 % der Varianz der *Nutzungsintention* erklären. Die neuen Konstrukte *Vergnügen* und *Selbstwirksamkeit* erklärten die Varianz von PEOU zu 51 %. Der Regressionskoeffizient zwischen *Selbstwirksamkeit* und tatsächlicher Nutzung lag mit 0.30 höher als derjenige zwischen *Nutzungsintention* und tatsächlicher Nutzung (0.20). Damit erklärte der nicht im TAM enthaltene Einfluss der *Selbstwirksamkeit* die tatsächliche Nutzung besser als das TAM-Konstrukt der *Nutzungsintention*. Insgesamt konnten diese beiden Konstrukte jedoch nur 14 % der Varianz der tatsächlichen Nutzung aufklären. Ein statistischer Modellvergleich mit dem TAM wurde von den Autoren nicht durchgeführt, so dass nicht beurteilt werden kann, ob der zusätzliche Erklärungswert der neuen Konstrukte deren Aufnahme in das Modell rechtfertigt.

Ausgehend vom TAM haben Calisir und Calisir (2004) nicht die *Nutzungsintention* als abhängiges Konstrukt, sondern die Zufriedenheit untersucht. Die Untersuchung war als Fragebogenstudie angelegt, in der Fragebögen in 24 Unternehmen verteilt wurden, die zentrale Warenwirtschaftssysteme benutzen. Zusätzlich zu den TAM-Konstrukten PU und PEOU wurden sechs weitere Konstrukte als mögliche Prädiktoren der Zufriedenheit der Benutzer mit dem Warenwirtschaftssystem erhoben. Die neuen Konstrukte sollen die Eigenschaften der Bedienoberfläche erfassen. Die Forschungsfrage der Autoren hierzu lautet:

Which, if any, interface usability characteristics influence end-user satisfaction with ERP systems?

(Welche Usability-Eigenschaften der Bedienoberfläche beeinflussen, falls überhaupt ein Zusammenhang besteht, die Zufriedenheit mit Warenwirtschaftssystemen?
Übers. d. Verf.)

Als Usability-Kriterien in diesem Sinne wurden „System capability“ (Leistungsfähigkeit des Systems, Übers. d. Verf.), „Compatibility“ (Kompatibilität, Übers. d. Verf.), „Flexibility“ (Flexibilität, Übers. d. Verf.), „User guidance“

(Benutzerführung, Übers. d. Verf.), „Learnability“ (Erlernbarkeit, Übers. d. Verf.) und „Minimal memory load“ (geringstmögliche Gedächtnisbelastung, Übers. d. Verf.) definiert.

Die Entwicklung der zugehörigen Skalen wurde in der Arbeit nicht näher beschrieben. Insgesamt wurden 21 Fragebögen ausgewertet. Die Auswertung erfolgte durch Kovarianzanalysen. PU und die Erlernbarkeit hatten eine signifikante Regression auf die Zufriedenheit der Endnutzer. Zusammen erklären sie 48 % der Varianz der Zufriedenheit der Endnutzer. Von den übrigen Konstrukten sind nur PEOU, die Leistungsfähigkeit des Systems und die Benutzerführung in der Modellierung enthalten geblieben. Der Regressionskoeffizient zwischen PEOU und der Zufriedenheit wurde von den Autoren nicht angegeben, erreichte aber augenscheinlich keine Signifikanz. PEOU und die Leistungsfähigkeit des Systems haben einen signifikanten Zusammenhang mit PU. Die Benutzerführung hat sowohl Einfluss auf PU als auch auf die Erlernbarkeit.

Zufriedenheit kann als Einstellung bezüglich des Warenwirtschaftssystems verstanden werden und wäre somit ein Vorläufer der *Nutzungsintention* nach dem ursprünglichen, stark an die TRA angelehnten TAM. Die Autoren erklären selbst nicht, warum sie nicht die *Nutzungsintention* bzw. die tatsächliche Systemnutzung erheben, sondern statt dessen die Zufriedenheit. Interessant ist der bedeutende Einfluss der Erlernbarkeit, die als zweites Konstrukt neben PU einen Einfluss auf die Zufriedenheit hat. Dabei wird in dieser Untersuchung Erlernbarkeit von PEOU abgegrenzt betrachtet. Dieser Befund stützt die in dieser Arbeit untersuchte Annahme, dass Erlernbarkeit als Einflussfaktor von PEOU eine besondere Rolle zukommt.

Klopping und McKinney (2004) untersuchten die Nutzung von Internetshops. Dabei erweiterten sie das TAM um das Konstrukt des Task-Technology Fit (Anpassung der Technologie an die Aufgabe, Übers. d. Verf.), welches 1995 von Goodhue vorgestellt wurde. Nach Goodhue sind drei Faktoren für die Einschätzung der Leistungsfähigkeit einer Technologie in einer bestimmten Situation verantwortlich. Die geschätzte Leistungsfähigkeit bestimmt die Einstellung zur entsprechenden Technologie. Die bestimmenden Faktoren sind die Aufgabe, die Technik und das Individuum. Der Faktor Aufgabe beinhaltet den Schwierigkeitsgrad und die Vielfältigkeit der Aufgabe. Unter dem Faktor Technologie werden Eigenschaften des Informationssystems erfasst. Der Faktor Individuum berücksichtigt die Fähigkeiten der Benutzer eines Systems. Die Autoren erwarten Einflüsse des Konstrukts Task-Technology Fit auf PU, PEOU und die *Nutzungsintention*.

Für die Überprüfung der Hypothesen wurde eine Fragebogenstudie an 429 Studenten durchgeführt. Die Skalen des Fragebogens wurden für PU, PEOU und die *Nutzungsintention* von Davis übernommen, Task-Technology Fit wurde mittels der von Goodhue (1995) hierfür entwickelten Skala erfasst. Zur Auswertung der Daten wurden Kovarianzanalysen durchgeführt. Obwohl die Regressionskoeffizienten des

zusätzlich ins Modell aufgenommen Konstrukts Task-Technology Fit alle entsprechend der Hypothesen signifikant werden, ist der zusätzliche Erklärungswert dieses Konstrukts gering. Die aufgeklärte Varianz der Systemnutzung erhöht sich nicht und bleibt bei 36 %. Die aufgeklärte Varianz des Konstrukts *Nutzungsintention* steigt von 47 % auf 52 % an. Die Autoren geben verschiedene Maße für die Modellgüte an, allerdings sind diese nicht klar zu interpretieren, da die im Text berichteten Werte von denen in einer Tabelle abweichen. Nach den Maßen für die Güte der Modellanpassung aus der Tabelle zeigt sich keine Verbesserung der Modellierung durch die Aufnahme des zusätzlichen Konstrukts. Dies entspricht auch den Erwartungen aus der nicht erhöhten Varianzaufklärung des erweiterten Modells. Die Zusammenhänge innerhalb des TAM zeigen einen starken Einfluss von PU auf die *Nutzungsintention* und auch einen direkten Einfluss von PU auf die tatsächliche Systemnutzung. Der Einfluss von PEOU auf die *Nutzungsintention* fällt geringer aus, der Regressionskoeffizient beträgt 0.15 in der Modellierung ohne Task-Technology Fit und 0.12 in der Modellierung mit Task-Technology Fit. Beide Regressionskoeffizienten sind statistisch signifikant. Zusammenhänge zwischen PU und PEOU wurden von Klopffing und McKinney nicht postuliert und auch nicht ins Modell aufgenommen.

Die berufliche Nutzung des Internets war der Forschungsgegenstand in einer Studie von Shih (2004), die das TAM als theoretischen Hintergrund nutzte. Der Autor entwickelte ein theoretisches Modell zur Vorhersage der Perceived Performance (wahrgenommene Leistung, Übers. d. Verf.), dem das TAM in seiner ursprünglichen Form, also mit dem Konstrukt Einstellung, zugrunde lag. Shih's Modell erklärt allerdings nicht die *Nutzungsintention* und die tatsächliche Nutzung, sondern die wahrgenommene Leistung. Zusätzlich wurde die vom Benutzer eingeschätzte Relevanz der im Internet abrufbaren Informationen als Konstrukt in das Modell aufgenommen. Die Relevanz soll sich auf die TAM-Konstrukte PU und PEOU sowie die Einstellung und die wahrgenommene Leistung auswirken. Das Forschungsmodell von Shih ist in Abb. 5 dargestellt.

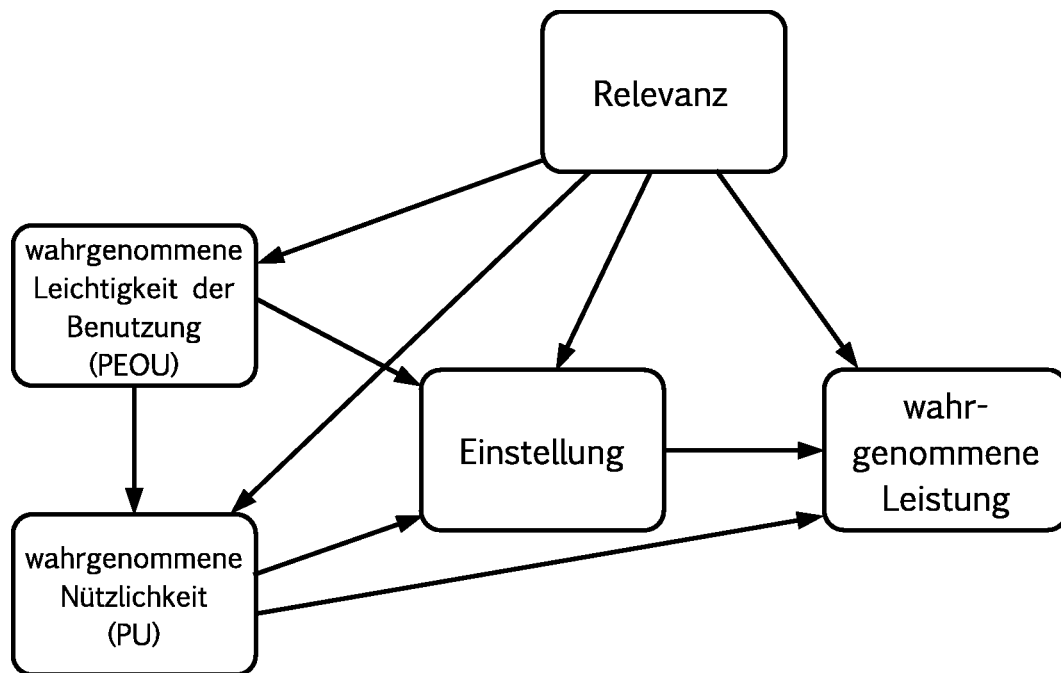


Abb. 5: Modifiziertes TAM (nach Shih, 2004)

Zur Überprüfung des Modells führte Shih eine Fragebogenerhebung in 10 taiwanesischen Firmen durch. Insgesamt konnten 203 Fragebogen ausgewertet werden. Shih benutzte für die Skalen für PU und PEOU eine nicht näher begründete Auswahl von Items aus Davis (1992) Untersuchung. Auch für die Skala für Einstellung griff er auf die Arbeiten von Davis zurück. Die Items der Skalen für Relevanz und wahrgenommene Leistung wurden vom Autor formuliert. Zur Überprüfung der Skalen wurde eine Faktorenanalyse über alle Items gerechnet. Die Ladungsstruktur der Items entsprach den vermuteten Konstrukten. Zur Überprüfung des Modells wurde eine Kovarianzanalyse durchgeführt. Ein Vergleich verschiedener Modelle wurde nicht vorgenommen. Alle Regressionskoeffizienten des Forschungsmodells erreichten statistische Signifikanz. Relevanz hatte den stärksten Einfluss auf die wahrgenommene Leistung, die Einstellung den schwächsten. PEOU beeinflusste die Einstellung am stärksten, der Regressionskoeffizient war mit 0.48 hochsignifikant. Der Einfluss von PU auf die Einstellung war mit einem Regressionskoeffizienten von 0.16 zwar auch signifikant, aber deutlich geringer. Das neue Konstrukt Relevanz zeigte signifikante Zusammenhänge mit allen anderen Konstrukten. Obwohl das Modell keine direkte Verbindung von PEOU zur wahrgenommenen Leistung enthält, interpretiert der Autor die Ergebnisse als Beleg für die Bedeutung von PEOU für das Internet. Leider wurde in dieser Untersuchung mit wahrgenommener Leistung eine vom TAM abweichende abhängige Variable gemessen, was die Vergleichbarkeit mit anderen TAM-Studien einschränkt.

Wu und Wang (2004) untersuchen die Akzeptanz der mobilen – also via Mobiltelefon oder PDA durchgeführten – Benutzung von Internetshops. Dabei nutzen sie das TAM als theoretische Grundlage, erweitern es aber um das Konstrukt

„*Compatibility*“ (Kompatibilität, Übers. d. Verf.) aus der „*Innovation Diffusion Theorie*“ (Theorie der Ausbreitung von Neuerungen, Übers. d. Verf.) und die Konstrukte „*Cost*“ (Kosten, Übers. d. Verf.) und „*Perceived risk*“ (wahrgenommenes Risiko, Übers. d. Verf.). Die zusätzlich in die Modellierung aufgenommenen Konstrukte sollen sich alle auf die *Nutzungsintention* auswirken. Außerdem wird ein Einfluss des Konstrukts *Kompatibilität* auf PU erwartet.

Die Skalen für die Messung der Konstrukte wurden auf Grundlage bewährter Skalen entwickelt. Die Autoren passten die Formulierungen an den Kontext ihrer Studie an und schlossen nicht näher spezifizierte Items von den Skalen aus. Für die Untersuchung wurden 850 Fragebögen über acht Anbieter entsprechender mobiler Dienste an deren Kunden verschickt. Aus dem Rücklauf konnten 310 Fragebögen ausgewertet werden. Eine Faktorenanalyse bestätigte die den Fragebogenitems zugrundeliegende Skalenstruktur. Die statistische Auswertung erfolgte durch Kovarianzanalysen der Skalenwerte. Bis auf die direkte Wirkung von PEOU auf die *Nutzungsintention* wurden die Regressionskoeffizienten aller postulierten Zusammenhänge signifikant. Allerdings wirkt das wahrgenommene Risiko genau entgegen der Erwartung der Autoren: entsprechend der Daten führt ein höheres wahrgenommenes Risiko von mobilen Transaktionen zu einer höheren *Nutzungsintention*. Die Autoren geben verschiedene Parameter für die Güte ihrer Modellierung an, vergleichen ihre Modellierung aber nicht mit alternativen Modellierungen wie beispielsweise dem normalen TAM. Die angegebenen Parameter der Modellgüte zeigen eine gute Anpassung des Modells an die Daten. Die aufgeklärte Varianz der *Nutzungsintention* oder des tatsächlichen Nutzungsverhaltens wurde von den Autoren nicht angegeben. Den stärksten Einfluss auf die *Nutzungsintention* hat in dieser Studie das Konstrukt *Kompatibilität* mit einem Regressionskoeffizienten von 0.51, gefolgt von PU mit einem Regressionskoeffizienten von 0.33. PEOU hat keinen signifikanten Einfluss auf die *Nutzungsintention*, wirkt allerdings auf PU, und somit indirekt auf die *Nutzungsintention*.

Im Rahmen eines Universitätsseminars untersuchten Saadé und Bahli (2004) die Akzeptanz einer Online-Lernsoftware. Hierfür erweiterten sie das TAM um das Konstrukt „*Cognitive absorption*“ (Kognitive Vereinnahmung, Übers. d. Verf.). Dieses neue Konstrukt soll die Eindrücke des Benutzers bei der Nutzung eines Informationsverarbeitungssystems erfassen. Das neue Konstrukt *Kognitive Vereinnahmung* soll sich auf die TAM-Konstrukte PU und PEOU auswirken.

Das Konstrukt *Kognitive Vereinnahmung* soll sich aus den drei Aspekten „*Temporal dissociation*“ (*Verlust des Zeitgefühls*, Übers. d. Verf.), „*Focused immersion*“ (*Konzentriertes Eintauchen*, Übers. d. Verf.) und „*Heightened enjoyment*“ (*Erhöhtes Vergnügen*, Übers. d. Verf.) zusammensetzen, für die die Autoren drei neue Skalen formulieren. Die Erhebung von Daten erfolgte im Rahmen des Universitätsseminars mittels Fragebögen. 103 Fragebögen konnten ausgewertet

werden. In einem ersten Schritt wurde die Qualität der Skalen getestet. Dabei wurde allerdings nicht beschrieben, wie die drei Skalen *Verlust des Zeitgefühls*, *Konzentriertes Eintauchen* und *Erhöhtes Vergnügen* in eine Skala für Kognitive Vereinnahmung überführt wurden. Die Autoren geben lediglich die Eigenschaften der Skala für Kognitive Vereinnahmung an, deren Reliabilität mit Chronbach's Alpha von 0.52 schwach ausfällt. Statistische Kenngrößen zu den drei zugrundeliegenden Skalen wurden von den Autoren nicht beschrieben.

Die Überprüfung der Hypothesen erfolgte durch eine Kovarianzanalyse. Es zeigten sich – wie von den Autoren vermutet – statistisch signifikante Zusammenhänge zwischen dem neuen Konstrukt Kognitive Vereinnahmung und den TAM-Konstrukten PU und PEOU. Auch die Regressionskoeffizienten zwischen PU bzw. PEOU und der *Nutzungsintention* wurden statistisch signifikant. Der Zusammenhang zwischen PU und PEOU war ebenfalls signifikant und wurde von den Autoren als Wirkung von PEOU auf PU interpretiert. Die beiden TAM-Konstrukte erklärten gemeinsam 26 % der Varianz der *Nutzungsintention*. Die tatsächliche Nutzung wurde in dieser Studie nicht erhoben. In einer alternativen Modellierung überprüften die Autoren einen möglichen direkten Einfluss des Konstrukts Kognitive Vereinnahmung auf die *Nutzungsintention*. Dieser Zusammenhang war statistisch hochsignifikant. Das neue Konstrukt allein erklärte 29 % der Varianz der *Nutzungsintention*. Diesen hohen Erklärungswert des neuen Konstrukts gegenüber den TAM-Konstrukten begründen die Autoren nicht. Eine Modellierung, bei der sowohl die TAM-Konstrukte wie auch das neue Konstrukt zur Vorhersage der *Nutzungsintention* herangezogen wurden, wurde nicht berichtet. Das neue Konstrukt erscheint vielversprechend. Es bleibt allerdings unklar, wie es sich im direkten Vergleich zu den TAM-Konstrukten verhält. Von den TAM-Konstrukten hatte PU einen höheren Einfluss auf die *Nutzungsintention* als PEOU, wobei beide statistisch signifikant waren.

In den neun betrachteten Studien wurde das TAM in unterschiedlichen Kontexten eingesetzt und konzeptionell verändert. Die Veränderungen erhöhten meistens den Anteil der durch das Modell aufgeklärten Varianz der Versuchsdaten. Dieser Umstand legt eine Verbesserung der Anpassungsgüte des Modells nahe, ist aber kein eindeutiger Beleg hierfür. Die Aufnahme eines zusätzlichen Prädiktors in ein Modell verbessert die Modellanpassung immer, weshalb der zusätzliche Erklärungswert des neuen Parameters statistisch zusätzlich abgesichert werden sollte. In den meisten Studien wurden keine hierfür geeigneten statistischen Kenngrößen angegeben. Von den zusätzlichen Konstrukten hat sich keines als Verbesserung des TAM etablieren können. In den neun Studien war meistens PU der stärkste Prädiktor für die *Nutzungsintention*, während der Einfluss von PEOU auf die *Nutzungsintention* in den Untersuchungen starke Unterschiede aufweist. In drei der neun Studien war der Einfluss von PEOU auf die *Nutzungsintention* stärker als der Einfluss von PU. In der Untersuchung von Venkatesh, Morris, Davis und Davis (2003) zeigte sich eine Veränderung des Zusammenhangs zwischen dem PEOU-

entsprechenden Konstrukt *Aufwandserwartung* und der *Nutzungsintention* abhängig von der Nutzungsdauer. Die Autoren benutzten in dieser Studie das vom TAM abweichende Konstrukt der *Aufwandserwartung*, um die subjektive Einschätzung der Bedienbarkeit zu erfassen. Die Operationalisierung dieses Konstrukts unterscheidet sich allerdings nur marginal vom TAM-Konstrukt PEOU, woraus eine Vergleichbarkeit der Konstrukte resultiert. Am Anfang der Benutzung war der Einfluss der *Aufwandserwartung* stark und mit zunehmender Erfahrung der Benutzer wurde er schwächer.

Zusammenfassung

In den betrachteten 14 Studien mit dem TAM als theoretischem Grundgerüst zeigen die Konstrukte PU und PEOU mehrheitlich einen signifikanten Zusammenhang mit den abhängigen Konstrukten *Nutzungsintention* bzw. tatsächlicher Nutzung. In 12 Studien werden die Regressionskoeffizienten zwischen PU bzw. PEOU und der *Nutzungsintention* oder einer vergleichbaren Variablen angegeben.

Venkatesh, Morris, Davis und Davis (2003) formulierten die beiden neuen Konstrukte *Leistungserwartung* und *Aufwandserwartung*, die im wesentlichen PU und PEOU entsprechen. Die Regressionskoeffizienten wurden in dieser Untersuchung angegeben. Alle Regressionkoeffizienten zwischen PU und der *Nutzungsintention* oder einer vergleichbaren Variablen wurden statistisch signifikant. Damit erweist sich das Konstrukt PU als verlässlicher Prädiktor für die Systemakzeptanz.

Unklarer ist der Zusammenhang zwischen dem zweiten TAM-Konstrukt PEOU und anderen Variablen. In sechs Untersuchungen war der Regressionskoeffizient zwischen PEOU und der *Nutzungsintention* oder einer vergleichbaren Variablen statistisch signifikant. In fünf Studien wurde die direkte Korrelation nicht signifikant, allerdings war hier meist ein signifikanter Zusammenhang zwischen PEOU und PU zu beobachten. Diesen Zusammenhang interpretierte Davis (1992) als indirekte Wirkung von PEOU über PU auf die *Nutzungsintention*. In zwei Studien waren die Ergebnisse bezüglich der Regression von PEOU nicht eindeutig. Hier war zum Beginn der Systemnutzung ein signifikanter Zusammenhang zwischen PEOU und der *Nutzungsintention* zu beobachten, der sich bei nachfolgenden Messungen abschwächte und statistisch unbedeutsam wurde.

Das TAM-Konstrukt PEOU entspricht der Wahrnehmung der Bedienbarkeit durch den Benutzer eines Systems. Es ist somit das Wahrnehmungskorrelat der Systemeigenschaften, die unter dem Begriff „Usability“ zusammengefasst werden. Vor dem Hintergrund der Bedeutung, die Usability-Aspekten bei der Systemgestaltung derzeit beigemessen werden, überrascht der schwankende Einfluss von PEOU auf die *Nutzungsintention*. Es stellt sich die Frage, ob die Usability von Systemen überschätzt wird, oder der Einfluss von Usability auf die *Nutzungsintention* komplexer ist, als im TAM modelliert. Als Hinweis hierfür kann auch die Studie von Segars und Grover (1983) interpretiert werden, die eine

komplexere Struktur hinter den Items der Skalen gegenüber dem TAM annimmt. Im Gegensatz zum TAM postulieren die Autoren drei zugrunde liegende Faktoren als Ergebnis einer faktoranalytischen Analyse der Daten.

In dieser Arbeit wird untersucht, ob eine Aufgliederung des Konstrukts PEOU in eine erfahrungsabhängige Komponente der wahrgenommenen Erlernbarkeit und eine stabile Komponente der allgemeinen Effizienz eine geeignete Weiterentwicklung des TAM darstellen. Für die weitere Betrachtung dieser Frage soll zunächst ein theoretischer Rahmen des Begriffs Usability definiert werden.

Kognitionspsychologische Grundlagen

Die folgenden Theorien und Begriffe gehören zum Gebiet der Kognitionspsychologie. Im Folgenden wird anhand Neissers (1976) Wahrnehmungszyklus eine fundamentale Annahme kognitionspsychologischer Theorien knapp erläutert. Demnach wird die Wahrnehmung nicht ausschließlich von den wahrgenommenen Reizen determiniert, sondern der Prozess der Wahrnehmung erfolgt unter aktiver Mitwirkung des Wissens, der Erwartungen und der Vorstellungen des Wahrnehmenden. Anschließend erfolgt eine detaillierte Darstellung der Idee eines mentalen Modells, welches die für das Individuum handlungsrelevante Umwelt repräsentiert. Die Vorstellung modellhafter mentaler Repräsentationen als Handlungsgrundlage ist älter als die Kognitionspsychologie, wurde in deren Rahmen aber wiederentdeckt und weiterentwickelt. Abschließend wird das Konzept eines mentalen Modells im Rahmen der Interaktion mit komplexen Systemen sowie dem Lernen betrachtet. Diese beiden Bereiche sind für die vorliegende Arbeit von zentraler Bedeutung.

Die Psychologie hat seit Beginn der sechziger Jahre des vorigen Jahrhunderts einen wichtigen Paradigmenwechsel vollzogen. Während zuvor der Forschungsschwerpunkt auf offenem, beobachtbarem Verhalten lag, wanderten mit der kognitiven Wende nicht direkt beobachtbare Informationsverarbeitungsprozesse in den Aufmerksamkeitsfokus der Forschung (Anderson, 1985). Die Kognitionspsychologie sieht den Menschen als aktives Wesen, das Informationen über seine Umwelt aktiv wahrnimmt und verarbeitet, statt lediglich passiv auf Reize zu reagieren. Neissers (1976) Wahrnehmungszyklus veranschaulicht das Zusammenwirken zwischen äußeren und inneren Prozessen (Abb. 6).

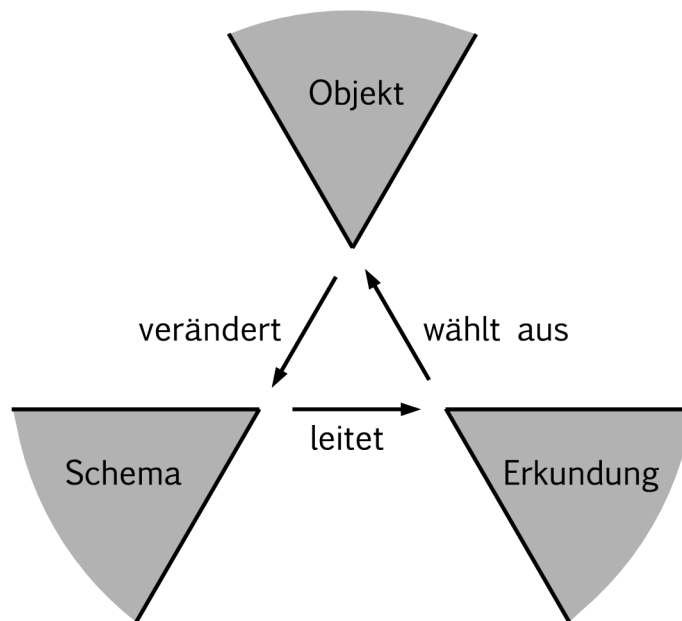


Abb. 6: Neissers Wahrnehmungszyklus (aus Neisser, 1979)

Nach Neisser ist Wahrnehmung ein konstruktiver Prozess. Der Wahrnehmende bildet aufgrund seiner aktivierten Schemata Erwartungen über die Realität aus, die seine Informationsaufnahme lenken. Die über die Objekte aufgenommenen Informationen aktualisieren die aktivierten Schemata und hierüber sein weiteres Erkundungsverhalten. Zur Veranschaulichung eines solchen Wahrnehmungsprozesses soll das automatische Speichern des Textverarbeitungsprogramms, das zum Verfassen dieser Arbeit genutzt wurde, als Beispiel dienen. In regelmäßigen Abständen öffnet sich ein Dialogfenster (Abb. 7), in dem nach Zustimmung zur automatischen Speicherung gefragt wird.



Abb. 7: Dialogfenster zum automatischen Speichern von OpenOffice

Das Intervall für die automatische Speicherung ist mit 15 Minuten lange genug gesetzt, dass der genaue Zeitpunkt nicht exakt vorhersehbar ist. Das plötzliche Erscheinen des Objekts „Dialogfenster“ löst entsprechend eine bottom-up-Verarbeitung aus. Sobald der Wahrnehmungsprozess das Objekt als „Dialogfenster“ erkannt hat, wird auf der Grundlage des Schemas „automatische Speicherung alle 15 Minuten“ die Vermutung gebildet, dass seit der letzten Speicherung 15 Minuten vergangen sind und es sich um den Dialog zum automatischen Speichern handelt. Nun wird gezielt nach Informationen gesucht, die diese Vermutung bestätigen.

Werden innerhalb des Fensters Worte wie „automatisches“, „Änderungen“, „speichern“ oder „gespeichert“ entdeckt, wird die Hypothese als bestätigt betrachtet und der Dialog mit der gewünschten Auswahl geschlossen.

Rumelhart (1980) differenziert je nach aktuell vorhandenem Realitätsbezug zwischen den grundsätzlich vorhandenen Schemata und einem „instanziierten“ Schema. Schemata repräsentieren verallgemeinerte Konzepte im Langzeitgedächtnis. Ein instanziiertes Schema existiert, wenn durch Wahrnehmung oder Vorstellung ein Schema geistig benutzt wird. Wir haben immer ein Schema „Stuhl“ im Gedächtnis verfügbar. Erst wenn wir an einen Stuhl denken – was momentan der Fall sein dürfte – gibt es ein instanziiertes Schema „Stuhl“ in unserer Vorstellung. Rumelhart benutzt die Analogie eines Theaterstücks, dessen Ablauf vorgegeben ist und bei dem trotzdem keine Aufführung der anderen gleicht. Das Stück selbst entspricht dem Schema, jede Aufführung wäre eine Instanzierung desselben. Rumelharts instanziiertes Schema entspricht Neissers aktiviertem Schema. Die Summe aller instanziierten Schemata „formt unser inneres Modell der Situation, mit der wir konfrontiert sind“ (Rumelhart, 1980, S. 37, Übers. d. Verf.). Diese Vorstellung einer mentalen Abbildung taucht erstmals in Craiks (1943) Monographie „The Nature of Explanation“ auf. Sie wird im Folgenden als „mentales Modell“ bezeichnet.

Ein mentales Modell erlaubt uns, mit einer gewissen Zuverlässigkeit Prognosen für unbeobachtete Ereignisse zu bilden. Das Konstrukt eines mentalen Modells ermöglicht es, komplizierte kognitive Prozesse beim Wahrnehmen und beim Problemlösen zu verstehen (Manktelow & Jones, 1987). Damit bietet es sich als kognitionspsychologische Grundlage für das Verstehen der Benutzung von Informationssystemen an. Im Folgenden soll das Konstrukt zunächst abstrakt dargestellt und dann auf die Mensch-Maschine-Interaktion bezogen werden.

Mentale Modelle

Der Modell-Begriff

Ein Modell im Allgemeinen ist die Abbildung eines Originals. Das Modell enthält eine Auswahl von Elementen des Originals und eine Auswahl von Bezügen zwischen den Elementen. Es stellt das Original vereinfacht dar. Trotzdem kann ein Modell auch Elemente enthalten, die im Original nicht vorhanden sind. Das elektrische Schaltbild einer Spannungsquelle als Modell für dieselbe kann in Reihe zur Spannungsquelle das Schaltzeichen für einen Widerstand beinhalten, wenn der innere Widerstand der Spannungsquelle wichtig ist und modelliert werden muss. Im Original sind Spannungsquelle und Widerstand eine untrennbare Einheit, die Modellierung beinhaltet in diesem Fall also mehr Elemente als das Original. Die Auswahl der abgebildeten Elemente orientiert sich am Zweck der Modellierung. Der Zweck von Modellen kann auf allgemeiner Ebene als „erkenntnisgewinnend und kommunikativ bezeichnet werden“ (Dutke, 1994). So kann der Prototyp eines

Automobils als Modell im Windkanal Aufschluss über die aerodynamischen Eigenschaften des geplanten Originals liefern. Dasselbe Modell kann der Kommunikation des geplanten Designs gegenüber Entscheidungsträgern dienen.

Analogiehaftigkeit mentaler Modelle

Eine besondere Form der Original-Modell-Beziehung ist die Analogie (Stachowiak, 1973). Hierbei unterscheiden sich die Elemente des Originals und des Modells, wobei sich die Relationen zwischen den unterschiedlichen Elementen gleichen. Als Beispiel soll die Analogie zwischen einem Wasserkreislauf und einem elektrischen Stromkreislauf dienen. Obwohl Wasser und elektrischer Strom unterschiedliche Dinge sind, verhalten sie sich in vielerlei Hinsicht gleich. So sind die Fallhöhe im Wasserkreislauf und die Spannung im Stromkreislauf Maße für die gespeicherte Energie pro Mengeneinheit (Masse beim Wasser bzw. Anzahl der Elektronen beim Strom). Werden diese erhöht, steigt in beiden Systemen die Durchströmung eines gegebenen Querschnitts. Eine Analogie ermöglicht den Transfer von Wissen aus einem Bereich in einen anderen, falls die strukturelle Vergleichbarkeit erkannt wird. So sind Vorhersagen über das Verhalten eines Stromkreislaufs möglich, wenn die Gesetzmäßigkeiten in Wasserkreisläufen bekannt sind und die Vergleichbarkeit der entsprechenden Parameter aus beiden Bereichen erkannt wird. Der Wissensgewinn durch eine solche Strukturübertragung ist von den Elementen der Modelle unabhängig. Allerdings werden Analogien für ähnliche Elemente leichter entdeckt (Dutke, 1994).

Schema und mentale Modelle

Die Einführung des Schemabegriffs wird häufig Bartlett (1932) zugeschrieben (Rumelhart, 1980). Bartlett beschäftigte sich mit der Frage, welche Prozesse am Erinnern beteiligt sind. In seinen Experimenten benutzte er eine Version des Kinderspiels „Stille Post“. Dabei wird eine Nachricht über eine Reihe von Personen übertragen. Dabei erfolgt die Weitergabe der Information flüsternd, also immer nur von einer Person zur nächsten. In der Regel werden die Informationen bei dieser Art der Weitergabe zunehmend verändert bzw. verfälscht. Als Material dienten Bartlett Sagen aus einem fremden Kulturkreis, er benutzte Erzählungen amerikanischer Indianer mit englischen Studenten als Versuchspersonen. Im Prozess des Wiedergebens veränderten sich die Geschichten nach einem wiederkehrenden Muster. Die Geschichten wurden verflacht, Namen von Protagonisten wurden weggelassen und stilistische Besonderheiten gingen verloren. Einige Details blieben erhalten und wurden häufig überspitzt. Die Geschichten wurden in den Nacherzählungen verständlicher für den Kulturkreis der Versuchspersonen. Inhalte der fremden Kultur – in den Originalgeschichten waren z. B. die Geister Verstorbener Protagonisten – verschwanden aus den Nacherzählungen. Diese Befunde führte Bartlett auf zwei grundlegende Zusammenhänge bei der menschlichen Informationsverarbeitung zurück:

Lernen und Gedächtnis: das Verstehen neuer Informationen erfordert deren Interpretation im Rahmen des vorhandenen Wissens. Die neuen Informationen müssen in bestehende Konzepte oder Schemata eingeordnet werden. Was gelernt wird, hängt also auch von schon vorhandenem Wissen ab. Die neuen Informationen werden verändert und verlieren in diesem Prozess an Details, die schlecht zu existierenden Schemata passen. Diese werden dabei aber für das Individuum stimmiger (Mayer, 1991).

Erinnern und Gedächtnis: Erinnern ist ein aktiver Prozess der Rekonstruktion. Beim Erinnern werden Details aus existierenden Schemata abgeleitet. Die im Gedächtnis gespeicherten Informationen sind detailarm und schematisch.

Rumelhart (1980) postuliert folgende Merkmale und Eigenschaften von Schemata:

- sie haben Variablen
- sie können verschachtelt sein
- sie stellen Wissen auf unterschiedlichem Abstraktionsgrad dar
- sie stellen eher Wissen als Definitionen dar
- sie sind aktive Prozesse
- sie sind an der Wahrnehmung beteiligt, indem sie ihre eigene Abbildungsgüte überprüfen

Schemata sind somit Strukturen allgemeinen Wissens. Sie enthalten abstrahiertes Wissen über typische Zusammenhänge eines Realitätsbereichs. Sie sind nicht auf Objekte begrenzt, sondern können sich auch auf Handlungsabläufe und Situationen beziehen. Sie sind der Hintergrund für die Bildung eines konkreteren mentalen Modells. (Marks, 1990). Dutke (1994, S. 28) beschreibt dies so: „Mentale Modelle werden auf der Grundlage schematischen Wissens konstruiert, sind selbst jedoch konkrete Instanzierungen eines oder mehrerer Schemata.“

Der Bezug zwischen Schemata und mentalen Modellen gleicht dem zwischen den Klassen und deren Instanzierungen in der objektorientierten Programmierung. Die sog. Klassen sind Vorlagen zur Verknüpfung von Datenstrukturen und Verarbeitungsregeln für diese Daten (Methoden). Sie fassen wiederkehrende und zusammenhängende Datenmanipulationen zu sinnvollen Einheiten zusammen und kapseln dabei die Detailebene der Datenmanipulation. Damit bilden sie ein Werkzeug zur Abbildung von komplexen Strukturen und Prozessen auf höherer Ebene, genau wie Schemata versuchen, Einzelereignisse der Welt aufgrund wiederkehrender Muster zu gliedern. Sowohl Klassen als auch Schemata werden allerdings nicht selbst benutzt, sondern stellen nur Regelwerke zur Verfügung. Diese Regelwerke kommen zur Anwendung, wenn in der objektorientierten Programmierung Instanzen einer Klasse gebildet, also die Datenstrukturen mit Daten gefüllt werden, um fortan mit den vordefinierten Methoden manipuliert zu werden. Mental werden die Schemata genutzt, um zu einem konkreten Modell der Situation oder eines Problems zu kommen. Das schematische Wissen definiert, wie sich Elemente des konkreten

mentalen Modells verhalten und miteinander wechselwirken.

Inhaltsabhängigkeit von schlussfolgerndem Denken

Eine simple Betrachtung versteht schlussfolgerndes Denken als die Anwendung formal-logischer Regeln auf gegebene Probleme. Die Ergebnisse psychologischer Experimente zu menschlichem Problemlösen widerlegen diese These. Mayer (1991) berichtet von Experimenten, die zum folgenden Schluss führen „formal logic and individual mental or psycho-logic are not necessarily the same“ (formale Logik und individuelle geistige bzw. psychologische Logik sind nicht zwingend dasselbe, Übers. d. Verf.). Ein vielzitiertes Experiment hierzu ist Wasons (1966) Selektionsaufgabe. Das Entscheidende an Wasons Experiment ist, dass die zugrundeliegende Aufgabe in unterschiedlichen thematischen Rahmen dargeboten werden kann. In der abstrakten Form der Aufgabe benutzte Wason Karten mit Buchstaben auf der einen und Ziffern auf der anderen Seite. Zusätzlich bekommen die Versuchspersonen eine Aussage genannt, deren Gültigkeit sie überprüfen sollen. Abb. 8 veranschaulicht die Aufgabenstellung. Die Aussage kann beispielsweise lauten: „Jede Karte, die ein D auf der einen Seite hat, hat eine 3 auf der anderen Seite.“

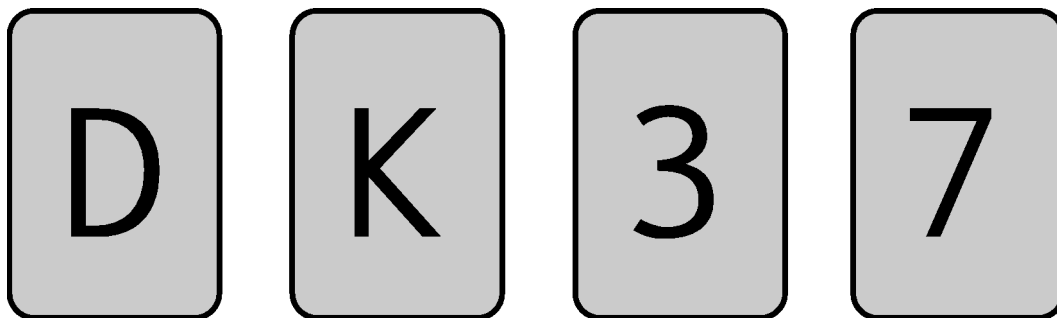


Abb. 8: Wasons Selektionsaufgabe (abstrakt; nach Wason & Shapiro, 1971)

Die Versuchsperson soll angeben, welche Karten umgedreht werden müssen, um die Gültigkeit der Aussage zu überprüfen. Es sollen nur die Karten gedreht werden, die zur Überprüfung beitragen können. Dabei zeigt sich über verschiedene Experimente hinweg eine niedrige Lösungsquote. Rips (1990) gibt eine typische Lösungsquote von unter 20 % an. Interessanterweise steigt die Lösungsquote dramatisch an, wenn die Aufgabe in einer alltagsnäheren Form präsentiert wird. Hasebrook (1995) gibt eine Lösungsquote von über 90 % für eine Version der Selektionsaufgabe mit Briefumschlägen an (Abb. 9). Hierzu lautet die Aufgabenstellung:

Sie arbeiten in einer Poststation und prüfen das Briefporto der eingehenden Briefe: Geschlossene Briefe kosten 1 DM, offene Briefe kosten 80 Pfennig. Überprüfen Sie, ob alle Briefe ausreichend frankiert sind.

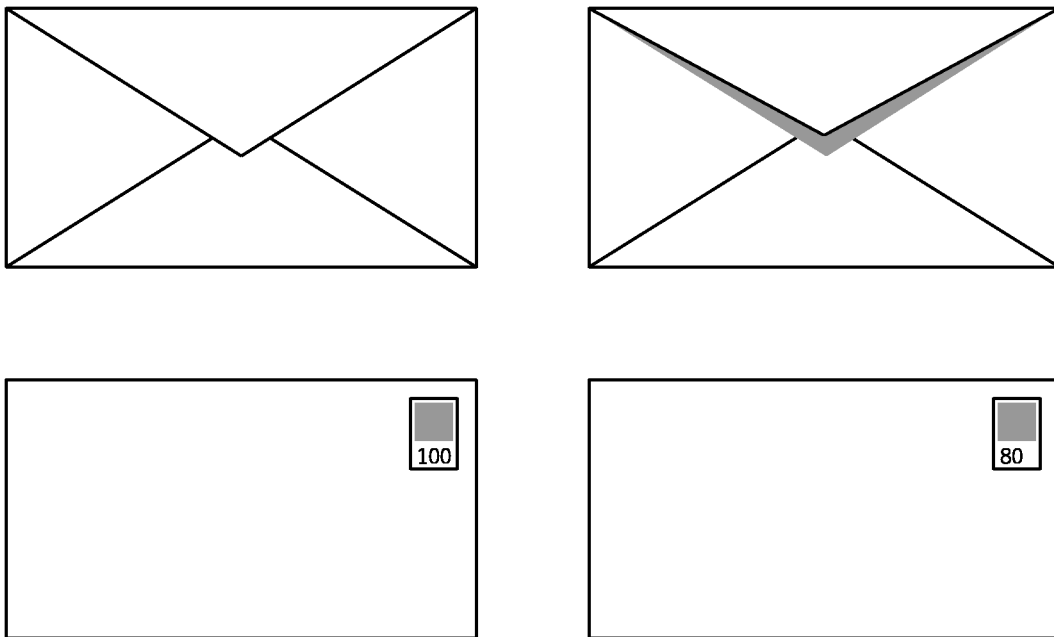


Abb. 9: Wasons Selektionsaufgabe (alltagsnah; nach Hasebrook, 1995)

Formal-logisch betrachtet sind die beiden Aufgaben identisch. Zwei Aussagen p und q werden zu „wenn p , dann q “ miteinander verknüpft. Diese Verknüpfungsregel ist dann verletzt, wenn auf p „Nicht- q “ folgt. Es müssen also die Fälle überprüft werden, die eine der beiden Aussagen enthalten: p oder „Nicht- q “. Für das abstrakte Beispiel wären dies die Karten mit einem D (p) oder Karten mit einer von 3 verschiedenen Ziffer („Nicht- q “), für das alltagsnahe Beispiel geschlossene Briefe (p) oder mit 80 Pfennig frankierte („Nicht- q “).

Manktelow und Jones (1987) betrachten unterschiedliche Untersuchungen zur Version der Selektionsaufgabe mit den offenen und geschlossenen Briefumschlägen (Abb. 9) genauer. Dabei war nicht immer eine erhöhte Lösungsquote für die Briefaufgabe gegenüber einer abstrakten Aufgabenstellung zu beobachten. Sie fassen die Ergebnisse einiger Studien hierzu zusammen: der erleichternde Effekt war nur zu beobachten, wenn die Versuchspersonen direkte Erfahrung mit einer entsprechenden Regelung hatten. Eine solche Regelung gab es tatsächlich in Großbritannien, und Personen, die entsprechend ihres Alters diese Regelung miterlebt hatten, zeigten eine bessere Lösungsquote. Für jüngere englische Probanden und US-amerikanische Versuchsteilnehmer zeigte sich in verschiedenen Studien kein verbessernder Effekt der Alltagsnähe auf die Lösungsquote. Manktelow und Jones schließen, dass der leistungsverbessernde Effekt bei der Selektionsaufgabe nur dann zu beobachten ist, wenn vorhandenes schematisches Wissen zur Konstruktion eines geeigneten mentalen Modells führt. Eine Erklärung der Problemlösung über das Konstrukt eines mentalen Modells erweist sich gegenüber der Annahme, Lösungen würden durch die Anwendung formal-logischer Regeln abgeleitet, als überlegen. Mentale Modelle können erklären, wie Alltagswissen menschliches Problemlösen beeinflusst.

Mentale Modelle als kognitive Simulationen

Mentale Modelle ermöglichen die Simulation von Vorgängen der Außenwelt (Dutke, 1994). Die Simulationsfähigkeit eines Modells hängt von seinem Abstraktionsgrad ab. Eine stärkere Abstraktion des Modells geht einher mit einer besseren Simulationsfähigkeit für komplexe Vorgänge. Viele Modelltheorien unterscheiden zwischen wahrnehmungsnahen und abstrakten Modellen. Johnson-Laird (1983) nennt die wahrnehmungsnahen Modelle „physical models“ (physische Modelle, Übers. d. Verf.) und die abstrakten „conceptual models“ (konzeptionelle Modelle, Übers. d. Verf.). De Kleer und Brown (1983) beziehen sich auf mechanistische Modelle und bezeichnen unterschiedlich abstrakte Modelle als „device topology“ (Baugruppen-Struktur, Übers. d. Verf.) und „causal model“ (Kausalmodell, Übers. d. Verf.). Sie veranschaulichen am Beispiel des elektrischen Stromkreises einer Haustürklingel, wie sich ein simulationsfähiges kognitives Modell entwickelt. Am Anfang steht eine einfache Repräsentation der Einzelkomponenten (Klingel, Schalter, Stromquelle und Leitungen), die „device topology“. Diese wird durch einen Veranschaulichungsprozess, den die Autoren als „envisioning“ (Vergegenwärtigen, Übers. d. Verf.) bezeichnen, in ein Kausalmodell überführt. Dabei wird aus dem Zustand einer Systemkomponente zunächst auf den Zustand einer anderen geschlossen. Dieser Prozess wird fortgeführt, bis die Kausalrelationen aller Einzelkomponenten so integriert sind, dass sie das Verhalten des Gesamtsystems determinieren. Das so gebildete Kausalsystem ermöglicht die Simulation spezifischer Vorgänge („running the causal model“, de Kleer & Brown, S. 160).

Die Art des mentalen Modells bestimmt die Qualität der Aussagen über das Verhalten des Zielsystems. Gentner und Gentner (1983) untersuchten, wie die Modellvorstellung über Elektrizität die Vorhersagen über einen Stromkreislauf beeinflussen. Sie erhoben, ob sich ihre Versuchspersonen elektrischen Strom als Wasserkreislauf oder als Menschenmenge in einem Gangsystem vorstellen. Beide Gruppen sollten vorhersagen, wie sich Strom und Spannung in einem Stromkreislauf verändern, wenn mehrere Batterien bzw. mehrere Widerstände parallel oder in Reihe geschaltet werden. Es zeigte sich eine Überlegenheit des Wassermodells für die Verschaltung von Batterien und des Menschenmengenmodells für die Verschaltung von Widerständen. Am Beispiel dieser beiden Vorstellungen zeigt sich eine weitere Eigenheit von mentalen Modellen: sie sind unvollständig, unscharf und teilweise widersprüchlich (Norman, 1983).

Seel (2003) beschreibt, wie die kognitive Simulation mit mentalen Modellen Vorhersagen über bisher unbeobachtete Ereignisse ermöglicht. Dabei wird das mentale Modell in die Zukunft projiziert und der veränderte Zustand wieder in das Modell integriert.

That means that a "mental simulation runs" envisioning the events in imagination that would take place in the world if a particular action were to be performed. (Dies bedeutet, dass eine „mentale Simulation abläuft“, die die Ereignisse veranschaulicht, die real passieren würden wenn eine bestimmte Handlung ausgeführt würde, Seel, 2003, S. 64, Übers. d. Verf.)

Die Möglichkeit, Prognosen über Handlungsergebnisse zu bilden, ist eine zentrale Voraussetzung für geplantes Handeln. Historisch betrachtet war die Simulationsfähigkeit schon bei Craik (1943) ein zentraler Aspekt seines „small-scale model“ (verkleinerten Modells, Übers. d. Verf.).

Mentale Modelle und komplexe Systeme

Neben dem Bereich der Kognitionspsychologie findet sich die Idee mentaler Modelle auch in der Forschungsrichtung zur Prozesskontrollsteuerung (Dutke, 1994). Rouse und Morris leiten 1986 einen Artikel über mentale Modelle ein mit:

It is a common assertion that humans have "mental models" of the system with which they interact. In fact, it is hard to explain many aspects of human behavior without resorting to a construct such as mental models.

(Es ist eine gängige Annahme, dass Menschen „mentale Modelle“ über die Systeme haben, mit denen sie interagieren. Tatsächlich ist es schwierig, viele Aspekte menschlichen Verhaltens ohne Rückgriff auf ein Konstrukt wie ein mentales Modell zu erklären, Übers. d. Verf.)

Veldhuyzen und Stassen (1976) berichten in einem Übersichtsartikel über eine weite Verbreitung der Vorstellung eines inneren Modells als Grundlage einer Regelungstätigkeit. Sie differenzieren (1977) folgende Bereiche dieser Modelle:

- Wissen über das zu regelnde System
- Wissen über Störbedingungen, die auf das System einwirken können
- Wissen über die auszuführende Aufgabe (Sollwerte, Randbedingungen, Signalbedeutungen)

Hacker (1998) benutzt in der Handlungsregulationstheorie den Begriff des „operativen Abbilds“. Dieses umfasst die Handlungsziele, die Ausführungsbedingungen und das Wissen über die Transformationsmaßnahmen, um den Ist- in den erwünschten Sollzustand zu überführen. Hacker berücksichtigt bei seiner Konzeption des operativen Abbilds die gedächtnispsychologische Differenzierung in Langzeit- und Arbeitsgedächtnis:

Desweiteren ist für die Tätigkeitsregulation der Unterschied zwischen Langzeit- und Arbeitsgedächtnis zu berücksichtigen. Im Langzeitgedächtnis werden schematisierte, rahmenhafte Abbilder für die Tätigkeitsregulation aufbewahrt. Hingegen sind die nur für die Dauer das Ausführens einer Tätigkeit unveränderten Sollwertabbildungen Inhalt des Arbeitsgedächtnisses. Sie sind gleichsam „Arbeitskopien“. Sie werden aus den Rahmenabbildern des Langzeitgedächtnisses abgeleitet und aktuell auf den neusten Stand gebracht. (Hacker, 1998, S. 189)

Die Übereinstimmung zwischen Hackers Begriff eines operativen Abbilds und dem Konstrukt eines mentalen Modells ist offensichtlich. Hackers Bezug auf Konstrukte der Gedächtnispsychologie schlägt eine Brücke zum nächsten Kapitel, in dem eine Betrachtung des Lernprozesses vor dem Hintergrund mentaler Modelle erfolgt.

Lernen und mentale Modelle

Das Konstrukt eines mentalen Modells hat Erklärungswert für die Interaktion eines Benutzers mit einem komplexen System. Ein erfahrener Benutzer unterscheidet sich durch Vorhandensein eines elaborierten mentalen Modells von einem unerfahrenen Anfänger. Insofern kann das Erlernen der Benutzung eines Systems – beispielsweise eines Computerprogramms – als Aufbau eines geeigneten mentalen Modells bzw. geeigneter Schemata verstanden werden (Seel, 2003).

De Kleer und Brown (1983) führen anhand eines überschaubaren Systems – ein elektronischer Summer – aus, welche Prozesse am Aufbau eines mentalen Modells beteiligt sind. In einem ersten Schritt werden einfache Annahmen über das System gemacht. Ein zweiter Schritt besteht darin, implizite Vermutungen über das Zusammenwirken von Modellkomponenten explizit zu machen. Dies ermöglicht die Bildung von Hypothesen über das Systemverhalten, die überprüft werden können. Falls sich Vermutungen nicht bestätigen, wird das Modell verändert, bis ein Modell entsteht, welches beobachtbares Systemverhalten in einer mentalen Simulation korrekt vorhersagen kann. Dieses Modell wird schematisch im Langzeitgedächtnis abgespeichert. Diese Ablaufbeschreibung deckt sich mit Berichten von Carroll und Mack (1984), die beobachteten, wie naive Anwender die Nutzung eines Textverarbeitungsprogramms erlernen. Dabei ist auffällig, wie gerade Fehler zu einem potentiellen Verständniszuwachs führen:

Ad hoc "theories" are hastily assembled out these odds and ends of partially relevant and partially extraneous generalization. And these "theories" are used for further prediction. Whatever initial confusions get into such as process, it is easy to see that they are at the mercy of an at least partially negative feedback loop: things quite often get worse before they get better.

Ad-hoc-„Theorien“ werden vorschnell aus diesen Überbleibseln teils relevanter, teils irrelevanter Verallgemeinerungen aufgestellt. Und diese „Theorien“ werden dann zu weiteren Vorhersagen herangezogen. Ganz gleich welche Fehlannahmen sich ursprünglich in diesen Prozess eingeschlichen haben, liegt auf der Hand, dass diese Theorien mit einer – zumindest teilweise – negativen Resonanzschleife stehen und fallen: die Dinge werden häufig zunächst schlimmer, bevor sie besser werden. (Carroll & Mack, 1984, S. 16, Übers. d. Verf.)

Mentale Modelle sind unvollständig und unscharf (Norman, 1983). Fehler, die aus der Anwendung eines mentalen Modells folgen, geben Hinweise auf Veränderungs-/Erweiterungsbedarf des Modells bzw. deren zugrundeliegenden Schemata (Senghaas, 2004). Dabei werden mentale Modelle allerdings nicht

zwingend verfeinert: Norman (1983) berichtet von Studien zur Benutzung von Taschenrechnern, in denen die meisten Versuchspersonen die „Clear“-Taste mehrfach drückten, um vorherige Ergebnisse zu löschen. Die Versuchspersonen waren sich im Klaren darüber, dass dieses Verhalten suboptimal war. Allerdings führt es für verschiedene Taschenrechnerversionen zum gewünschten Ergebnis. Mentale Modelle müssen zuverlässige Aussagen über die Konsequenzen von Handlungen ermöglichen, also funktional und nicht elegant sein.

Zusammenfassung

Dutke (1994) fasst zusammen, dass das Konzept mentaler Modelle sich in unterschiedlichen Ansätzen findet, die durch eine gemeinsame Perspektive auf problemlösendes Denken verknüpft sind. Das Wesentliche dieser Perspektive ist der Prozesscharakter menschlichen Wahrnehmens und Denkens, das nicht an einem Startpunkt beginnt und linear zu einem Endpunkt verläuft, sondern fortlaufend und rückgekoppelt stattfindet. Das Konstrukt eines mentalen Modells vermittelt hierbei zwischen Wahrnehmungs- und Gedächtnisprozessen und trägt zum Verständnis menschlichen Problemlösens bei. Folgende Eigenschaften sind charakteristisch für mentale Modelle:

- Mentale Modelle sind funktionale Abbildungen der Umwelt:
Mentale Modelle dienen als Grundlage der Handlungssteuerung oder ermöglichen das Verstehen von Problemen in konkreten Situationen. Sie dienen nicht der langfristigen Speicherung von Wissen im Gedächtnis, sondern existieren im Arbeitsgedächtnis, solange sie für die Wahrnehmung und Handlungssteuerung nützlich sind. Schematisches Wissen aus dem Langzeitgedächtnis ermöglicht die Bildung eines mentalen Modells einer Situation.
- Mentale Modelle werden situationsbezogen erzeugt:
Als Inhalt des Arbeitsgedächtnisses haben mentale Modelle immer einen Bezug zur aktuellen Situation, sie sind aber nicht von direkter Beobachtung abhängig. Auch eine vorgestellte Situation kann zur Bildung eines mentalen Modells führen bzw. die Vorstellung an sich ist schon das mentale Modell.
- Der Abstraktionsgrad – die Abbildungsregeln – sind variabel:
Je nach Einsatzzweck und Expertise werden unterschiedliche mentale Modelle gebildet. Mentale Modelle sind vereinfachte und bewährte Abbildungen ihres Originals. Es werden nur die relevanten Merkmale und die relevanten funktionalen Zusammenhänge im Modell abgebildet. Für die Bedienung eines Systems im Rahmen des normalen Betriebs reicht entsprechend ein einfacheres Modell im Vergleich zur Fehlerdiagnose am selben System.
- Mentale Modelle ermöglichen Prognosen:
Durch die Simulationsfähigkeit mentaler Modelle können Hypothesen über bisher nicht beobachtbares Verhalten komplexer Systeme gebildet werden. Die simulierten Erkenntnisse über einen zu erwartenden Endzustand von Aktionen

können zur Überprüfung des Modells genutzt werden. Das kognitive Experimentieren mit Systemverhalten kann Widersprüche und Lücken innerhalb des Modells aufdecken und zu dessen Weiterentwicklung beitragen.

- Mentale Modelle werden beim Lernen angepasst oder neu gebildet:
Lernen kann als Modifikation und Neubildung von mentalen Modellen bzw. deren zugrundeliegenden Schemata verstanden werden. In der Regel wird zunächst eine Modellierung auf Grundlage bestehender Schemata versucht und diese gegebenenfalls angepasst. Falls dies nicht möglich ist, werden „ad hoc“ Zusammenhänge vermutet und die so entstehende Modellierung aufgrund ihrer Abbildungsgüte der Realität bewertet und weiterentwickelt.

Insgesamt ermöglicht das Konstrukt mentaler Modelle eine einheitliche Perspektive auf verschiedene Prozesse, die bei der Nutzung komplexer Systeme und dem zugehörigen Problemlösen eine Rolle spielen. Die Stärke besteht dabei darin, dass die Zusammenhänge zwischen den einzelnen Prozessen und die vorhandenen Rückkopplungen vor dem Hintergrund mentaler Modelle besser sichtbar werden.

Usability

Ursprung des Begriffs

„Usability“ ist ein Begriff, der häufig im Kontext der Gestaltung von Bedienoberflächen, insbesondere von Computerprogrammen oder Webseiten, anzutreffen ist. Dabei wird der Begriff sehr flexibel genutzt, was eine Definition schwierig macht. Eine gängige deutsche Übersetzung für Usability ist „Benutzerfreundlichkeit“. Diese wird allerdings von Experten eher abgelehnt, da Gebrauchsgegenstände nicht „freundlich“ sein müssen, sondern funktional. Entsprechend übersetzt die EN ISO 9241 Usability mit „Gebrauchstauglichkeit“ und definiert diese als „das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.“ Im Folgenden werden der Ursprung des Konzepts sowie verschiedene Konzepte von Usability betrachtet und ihre Übereinstimmungen und Unterschiede herausgearbeitet.

Mit der raschen Ausbreitung des Computers in der Arbeitswelt begannen verschiedene Fachrichtungen das Nutzungsverhalten des Computers als Forschungsgegenstand zu entdecken. 1980 veröffentlichte Ben Shneiderman „Software psychology“ und wird seither als einer der Begründer des Forschungsgebiets der Softwarepsychologie angesehen.

Im Kapitel „Software Quality Evaluation“ beschreibt Shneiderman (1980) einen Ansatz von Boehm, Brown und Lipow (1976), die drei grundlegende Anforderungen an Software-Pakete definieren. Neben dem Aspekt der Nutzung spielen die Administrierbarkeit und die Portierbarkeit auf andere Betriebssysteme eine Rolle für die Qualitätseinschätzung. Sie benutzen den Begriff „Usability“ synonym zu „As-is

utility“, also der Nützlichkeit der Software für Endbenutzer. Die Beschränkung der Betrachtung auf die Endbenutzer ist typisch für Usability und klammert Fragen nach der Wartbarkeit des Programmcodes oder dessen Portierbarkeit in andere Programmiersprachen oder auf andere Betriebssysteme aus. Die As-is utility setzt sich aus den drei Bestandteilen Reliability (Zuverlässigkeit), Efficiency (Effizienz) und „Human engineering“ zusammen. Human engineering definieren Boehm, Brown und Lipow als „extent to which code fulfills its purpose without wasting users' time and energy or degrading their morale“ (Ausmaß in dem der Programmcode seinen Zweck erfüllt, ohne die Zeit und Energie des Benutzers zu verschwenden oder ihn zu demoralisieren, Boehm, Brown & Lipow, 1976, zitiert nach Shneiderman, Übers. d. Verf.).

Für die Betrachtung des Forschungsgebiets zu Usability ist die 1982 erfolgte Gründung der „Special Interest Group on Computer-Human Interaction“ innerhalb der Association for Computing Machinery (ACM) wichtig. Diese Gruppe der ACM befasst sich schwerpunktmäßig mit Fragen der Mensch-Maschine-Interaktion, also auch Usability (Grudin, 1992). Die Betrachtung der Interaktion zwischen Mensch und Computer wurde dabei vom Inhalt der Interaktion losgelöst betrachtet – soweit dies eben möglich ist. Dass eine solche Trennung nicht zweckdienlich sein kann, stellte Grudin (1992) sehr früh fest. Er erläutert, wie es zu dieser Trennung kam: hauptsächlich macht er den Prozess der Softwareentwicklung verantwortlich, in dem zunächst die gewünschte Funktionalität einer neuen Software definiert wurde. Erst später, nach Beginn der Produktentwicklung wurde mit der Implementierung der Bedienoberfläche begonnen. Die an der Entwicklung beteiligten Usability-Spezialisten hatten entsprechend keinen Gestaltungsspielraum bezüglich grundlegender Funktionalität und konzentrierten sich in ihrer Arbeit auf die Bedienoberfläche.

Die DIN 66234 „Grundsätze der Dialoggestaltung“ von 1988, ein direkter Vorläufer der besser bekannten EN ISO 9241-10 „Grundsätze der Dialoggestaltung“², folgt dieser auf die Bedienoberfläche beschränkten Betrachtung von Usability. Wie schon die DIN 66234 beinhaltet auch die EN ISO 9241-10 folgende Gestaltungsgrundsätze für Bedienoberflächen:

- Aufgabenangemessenheit
- Selbstbeschreibungsfähigkeit
- Steuerbarkeit
- Erwartungskonformität
- Fehlertoleranz
- Individualisierbarkeit
- Lernförderlichkeit

2 Die EN ISO 9241-10 wurde kürzlich zur EN ISO 9241-110 überarbeitet, dabei wurden allerdings keine grundlegenden Änderungen an der Norm vorgenommen.

Nur der erste Gestaltungsgrundsatz „Aufgabenangemessenheit“ stellt einen Bezug zum Nutzungskontext des Informationssystems her. Die Norm liefert folgende Definition von Aufgabenangemessenheit:

Ein Dialog ist aufgabenangemessen, wenn er den Benutzer unterstützt, seine Arbeitsaufgabe effektiv und effizient zu erledigen (EN ISO 9241-10, 1996).

Die aus dem EU-Forschungsprojekt MUSiC (Measuring Usability of Systems in Context) hervorgegangene EN ISO 9241-11 „Anforderungen an die Gebrauchstauglichkeit – Leitsätze“ rückt den Zusammenhang von Usability und Anwendungskontext in den Vordergrund. Nach diesem aktuelleren Ansatz – die Norm trat 1998 in Kraft – ist Usability keine Produkteigenschaft an sich, sondern entsteht erst im Wechselspiel zwischen Informationstechnologie und Nutzungskontext einschließlich des Benutzers.

Dieser neuere Blickwinkel auf Usability unterstützt den Grundgedanken dieser Arbeit, das Erfahrungsniveau eines Benutzers als wesentlichen Faktor in die Modellierung von Technologieakzeptanz aufzunehmen.

Usability in der EN ISO 9241-10

Dzida stellte 1983 das IFIP-Modell (International Federation of Information Processing Model) der Benutzerschnittstelle vor (Abb. 10).

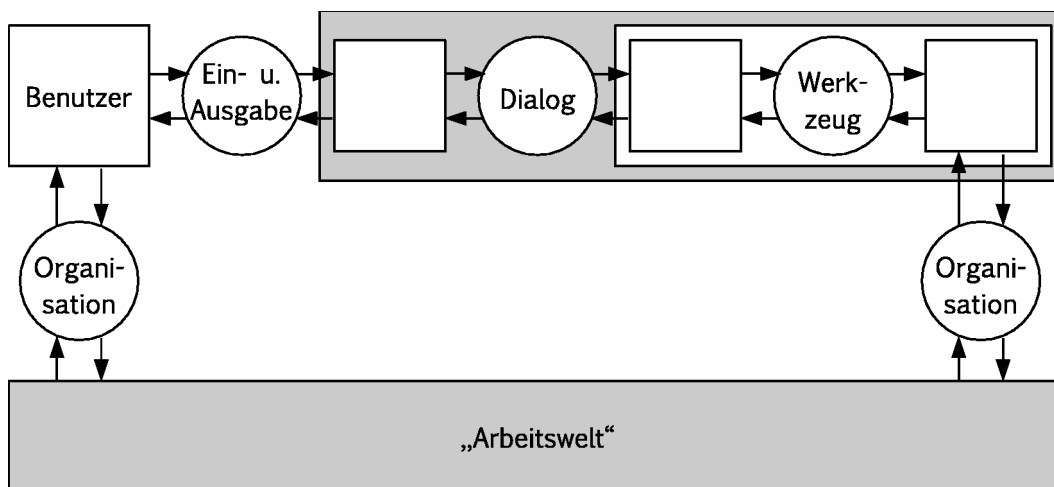


Abb. 10: IFIP-Modell (nach Dzida, 1983)

Grundlegender Gedanke des Modells war es, die Schnittstellen von Software zu standardisieren und abgekoppelt von konkreten Softwareanwendungen zu entwickeln. Das IFIP-Modell schlug eine Differenzierung der Nutzerschnittstelle in vier Teilbereiche vor. Die Ein- und Ausgabeschnittstelle regelt auf einfachster Ebene wie Daten eingegeben werden (z. B. über Tastatur und Maus) und ausgegeben werden (z. B. formatierte Bildschirmausgabe). In der Dialogschnittstelle wird spezifiziert, wie die Interaktion zwischen Mensch und Maschine ablaufen soll (z. B. wie Bearbeitungsaufträge ausgelöst werden oder Hilfe angefordert werden kann). Die

Werkzeugschnittstelle erlaubt es, Veränderungen an den Verarbeitungsregeln der Software vorzunehmen. Ein Beispiel für Werkzeugeinstellungen wären die Standardeinstellungen eines Internetbrowsers oder die Standardvorlage für ein neues Dokument in einer Textverarbeitungssoftware. Modifikationen hieran (z. B. Schriftart und Schriftgröße) würden gemäß dem IFIP-Modell über die Werkzeugschnittstelle erfolgen. Die vierte Schnittstelle bezeichnet das Modell als Organisationsschnittstelle. Sie beschreibt das Zusammenspiel zwischen den Funktionen der Software und den betrieblichen Bedürfnissen einer Organisation. Hierzu gehören Arbeitsteilung, Dienstweg und die Regeln für innerbetriebliche Kooperation. Allerdings schränkte Dzida die Trennbarkeit der Schnittstellen selbst ein:

Die drei Benutzerschnittstellen (Ein/Ausgabe-Schnittstelle, Dialog- und Werkzeugschnittstelle) können nicht völlig unabhängig voneinander realisiert und bewertet werden.

Dzida stellte Bewertungskriterien für die verschiedenen Schnittstellen auf. Die Bewertungskriterien für die Dialogschnittstelle leitet er aus den Ergebnissen einer Befragung von 233 Computerexperten ab (Dzida, Herda, Itzfeldt, 1978). In dieser Befragung wurde die Wichtigkeit von Aussagen wie:

Ein Dialogsystem sollte leicht erlernbar sein.

auf einer siebenstufigen Skala bewertet. Der Fragebogen umfasste 100 solcher Items, von denen 47 aus der Analyse ausgeschlossen wurden, weil sie unklar oder zu abstrakt formuliert waren, sich mit anderen Items überschneiden oder keine interpretierbare Varianz aufwiesen. Für die übrigen 53 Items wurde eine Faktorenanalyse durchgeführt, um eine zugrundeliegende Struktur der Antworten aufzudecken. Die gefundenen Faktoren bezeichneten Dzida et al. mit Verlässlichkeit, Flexibilität und Fehlertoleranz. Diese Kriterienauflistung war die Grundlage der DIN 66234, dem direkten Vorläufer der EN ISO 9241-10 „Grundsätze der Dialoggestaltung“. Die Norm listet die folgenden sieben Gestaltungsgrundsätze auf:

- **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 Benutzer auf Anfrage erklärt wird.
- **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.
- **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.

- Fehlertoleranz

Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebn trotz erkennbarer fehlerhafter Eingaben entweder mit keinem oder minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann.

- Individualisierbarkeit

Ein Dialogsystem ist individualisierbar, wenn das Dialogsystem Anpassungen an die Erfordernisse der Arbeitsaufgabe sowie an die individuellen Fähigkeiten und Vorlieben des Benutzers zulässt.

- Lernförderlichkeit

Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen des Dialogsystems unterstützt und anleitet.

Gemäß Dzidas (1983) Ansatz sieht die EN ISO 9241-10 die Usability als Aspekt der Dialogschnittstelle, die unabhängig von der Anwendung an sich entwickelt werden kann. Usability hängt in dieser Betrachtung nicht von der Einbettung eines Systems in einen Nutzungskontext ab, sondern ist eine Eigenschaft eines Systems an sich. Die sieben Kriterien basieren nicht auf einem theoretischen Fundament, sondern wurden aus empirischen Daten extrahiert. Die Bezeichnung „Gestaltungsgrundsätze“ verdeutlicht, dass bei der Formulierung der Schwerpunkt auf praktischer Anwendbarkeit und nicht auf der Theoriebildung lag.

Usability nach Ulich

Ulich (1986) entwickelte aus der Perspektive eines Arbeitspsychologen Kriterien zur Beurteilung der Usability von Informationsverarbeitungssystemen. Ulichs Kriterien wurden „in einem langjährigen Forschungsprozess durch Feldstudien und experimentelle Untersuchungen überprüft, präzisiert, strukturiert und in ein Gesamtkonzept der Benutzerfreundlichkeit integriert“ (Spinas, Waeber & Strohm, 1990). Ulichs Konzept der Gebrauchstauglichkeit umfasst folgende Kriterien:

- Transparenz
- Konsistenz
- Toleranz
- Kompatibilität
- Unterstützung
- Flexibilität/Individualisierbarkeit
- Partizipation

Ulich liefert 1986 keine Definitionen für die Kriterien, sondern illustriert sie mit Beispielen:

- Transparenz

-Der Benutzer sollte erkennen können, ob ein eingegebener Befehl behandelt wird oder ob das System auf weitere Eingaben wartet.

- Bei längeren Vorgängen sollte das System Zwischenstandsmeldungen abgeben können.
- Konsistenz
 - Die Antwortzeiten des Systems sollten möglichst wenig variieren; wichtiger als kurze Antwortzeiten sind regelmäßige und damit kalkulierbare Intervalle.
 - Das System sowie dessen Antwortverhalten sollten für den Benutzer transparent und konsistent sein; ähnliche Aktionen sollten ähnliche Ausführungen bewirken, andernfalls muss dies durchschaubar gemacht werden.
- Toleranz
 - Die Ausgabe sollte – insbesondere bei einem größeren Umfang – ohne Verlust anhaltbar und fortsetzbar sein.
 - Der Benutzer sollte den Ablauf unterbrechen und Eingaben nachträglich ändern können; an solchen Unterbrechungspunkten, die vom System her fix gewählt sein können, sollte der Dialog in veränderter Form wieder aufgenommen bzw. fortgesetzt werden können.
- Unterstützung
 - Dialoghilfen sowohl zu inhaltsbezogenen wie zu vorgehensbezogenen Aspekten sollten vom Benutzer während des Dialogs jederzeit abgerufen werden können; das Betätigen der allfälligen Help-Taste sollte gegenüber andern Befehlen einen Sonderstatus einnehmen.
 - Das System sollte eine Rückfragemöglichkeit derart bereitstellen, dass auf eine Aufforderung durch den Benutzer hin ggf. ausführlichere Antworten abgegeben werden.
- Kompatibilität
 - Bei der Darstellungsform für Einzelinformationen sollte ebenso wie für ganze Bilder ggf. auf Übereinstimmung mit entsprechenden gedruckten Vorlagen oder Unterlagen geachtet werden.
 - Sprache und begriffliche Komplexität des Dialogs sollten an den Gepflogenheiten und Kenntnissen des spezifischen Benutzerkreises orientiert sein; anstelle von EDV-Kürzeln sollte mit den jeweils fachspezifischen Begriffen der Benutzer gearbeitet werden können.
- Flexibilität / Individualisierbarkeit
 - Zum Zweck der Individualisierungsmöglichkeit von Dialogabläufen sollte die Software unterschiedliche Vorgehensweisen offenlassen; die Abfolge einzelner Arbeitsschritte sollte so wenig wie möglich vorgegeben sein.
 - Ebenso wie der inhaltliche Dialogablauf sollte auch die Geschwindigkeit der Interaktion vom Benutzer bestimmt werden können; Aufforderungen zur Eingabe (z. B. durch Blinken von Zeichen) sollten möglichst unaufdringlich und abschaltbar sein.
- Partizipation
 - Den Mitarbeitern soll durch die Beteiligung am Veränderungsprozess Gelegenheit

zur Beeinflussung der Veränderung geboten werden. Dies trägt zum Abbau von Widerstand bei bzw. ermöglicht dessen Umsetzung in konkrete (Alternativ-)Vorschläge.

- Der Einbezug der Mitarbeiter sollte schon in der Vorbereitungs- und Planungsphase erfolgen, damit ihre Bedürfnisse bei der Planung berücksichtigt und ihr Fachwissen genutzt werden kann. Dies reduziert auch die Wahrscheinlichkeit, dass nachträgliche Korrekturen notwendig werden.

(Ulich, 1986, Hervorhebungen im Original)

Ulichs Kriterien basieren auf arbeitspsychologischen Analysen der Mensch-Maschine-Interaktion. Dabei versucht auch Ulich, die Dialogschnittstelle eines Informationsverarbeitungssystems von dessen Funktionalität und dem Nutzungskontext entkoppelt zu betrachten. Wie auch bei den Kriterien der EN ISO 9241-10 fehlt ein theoretischer Unterbau, der die Grundlagen der Interaktion zwischen Mensch und System modelliert. Vielmehr sollen die Kriterien eine praktische Bewertung von Systemen ermöglichen und nicht primär die Theorieentwicklung vorantreiben.

Usability nach Nielsen

Nielsen (1993) beschreibt detailliert, wie die „Usability“ eines Systems durch iteratives Vorgehen während der Produktentwicklung sichergestellt werden kann. Er empfiehlt frühzeitige und häufige Tests mit zukünftigen Nutzern oder die Durchführung von Expertenbefragungen, um Probleme schon während der Systementwicklung aufzudecken. Die Lösungsvorschläge sollen in das System eingearbeitet und anschließend erneut überprüft werden. Dieses Wechselspiel zwischen Test und Überarbeitung soll möglichst den gesamten Prozess einer Systemimplementierung begleiten.

Nielsen definiert Usability als einen Teilaspekt der Systemakzeptanz, genauer der „Usefulness“ (Zweckmäßigkeit, Übers. d. Verf.) eines Systems. Nielsen unterteilt die Usefulness in „Utility“ (Brauchbarkeit, Übers. d. Verf.) und Usability. Nielsen beruft sich auf Grudin (1992), ignoriert aber, dass Grudin diese Aufteilung zwar feststellt, aber keinesfalls als angemessen bewertet. Abb. 11 stellt Niensens Konzeption von Usefulness mit den beiden Hauptkomponenten Utility und Usability dar.

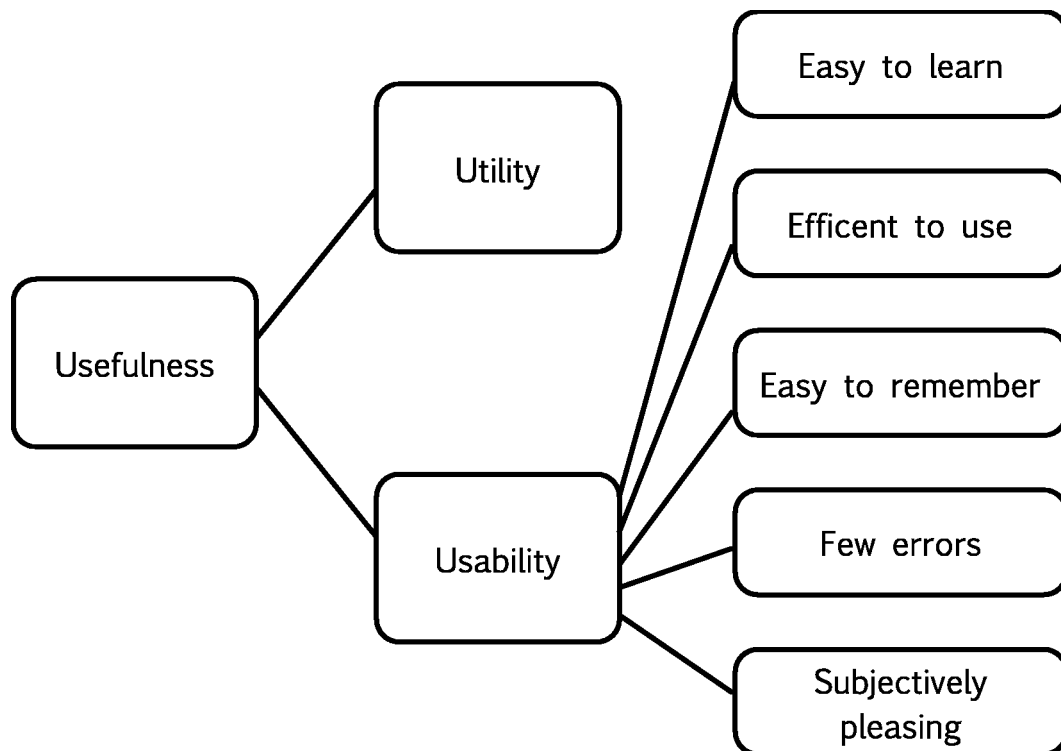


Abb. 11: Nielsens Konzeption von Usefulness (Nielsen, 1993)

Nielsen führt aus, daß Usability kein eindimensionales Konstrukt ist und sich aus fünf Aspekten zusammensetzt:

- Easy to learn (Leicht zu erlernen, Übers. d. Verf.)
- Efficient to use (Effizient zu benutzen, Übers. d. Verf.)
- Easy to remember (Leicht zu merken, Übers. d. Verf.)
- Few errors (Wenig Fehler, Übers. d. Verf.)
- Subjectively pleasing (Angenehm für den Benutzer, Übers. d. Verf.)

Nielsen bezeichnet die Erlernbarkeit als fundamentalsten Aspekt des Konstrukts Usability. Nielsen führt an, dass der erste Kontakt mit einem neuen System immer durch das Erlernen geprägt wird. Die Effizienz der Benutzung zeigt sich gemäß Nielsen für erfahrene Benutzer. Über die Bearbeitungszeit von Testaufgaben durch erfahrene Benutzer lässt sich die Effizienz messen. Der dritte Aspekt von Usability ist relevant für Gelegenheitsnutzer. Für diese Gruppe ist es wichtig, die Bedienelemente nach einer gewissen Pause erfolgreich wiederzuerkennen. Hinsichtlich der Benutzerfehler differenziert Nielsen zwischen solchen, die unmittelbar vom Benutzer korrigiert werden und solchen, deren Korrektur nicht möglich ist. Für die praktische Bewertung der Qualität eines Systems empfiehlt er eine gesonderte Betrachtung der nicht korrigierbaren Fehler, da diese den Anwender erheblicher beeinträchtigen als korrigierbare Fehler. Der letzte Aspekt ist die subjektive Einschätzung der Benutzung eines Systems. Hierbei geht es nicht um eine generelle Einstellung Computern und moderner Technologie gegenüber, obwohl sich diese

nach Nielson sicherlich auf die Einschätzung eines konkreten Systems auswirkt.

Nielsens Definition stammt aus einem anwendungsorientierten Ansatz. Er erläutert die einzelnen Aspekte anhand möglicher Erhebungsverfahren und mit Praxisbeispielen. Es fehlt ein theoretischer Hintergrund, entsprechend ergeben sich Schwierigkeiten, die Aspekte gegeneinander abzugrenzen: der Aspekt der Erlernbarkeit lässt sich nur schwer vom Wiedererkennen durch einen Gelegenheitsnutzer abgrenzen. Bezüglich der Effizienz schränkt Nielsen selbst ein, dass nicht alle Benutzer ein entsprechendes Lernniveau erreichen, damit dieser Aspekt relevant und erfassbar wird. Kleinere Fehler wirken sich auf die Arbeitsgeschwindigkeit aus und beeinflussen somit die Effizienz. Erlernbarkeit und Effizienz haben einen Einfluss auf die subjektive Einschätzung des Systems.

Für die angewandte Arbeit ist Nielsens Konzeption sicherlich hilfreich, weil sie den Blick des Praktikers auf verschiedene Problemfelder lenkt. Diese beschränken sich nicht auf die Usability einer Systemgestaltung sondern umschließen auch methodische Fallstricke bei der Operationalisierung von Usability. So eignen sich erfahrene Computernutzer nicht für die Erhebung der Erlernbarkeit, weil sie Wissen von anderen Bedienoberflächen übertragen können und somit eher die Wiedererkennbarkeit erfasst wird. Insgesamt bleibt Nielsens Usability Konzept allerdings unscharf, es umfasst alles, was die Erlernbarkeit verbessert und die Bedienung schneller und zufriedenstellender macht.

Usability in der EN ISO 9241-11

Die EN ISO 9241-11 „Anforderungen an die Gebrauchstauglichkeit – Leitsätze“ ist 1998 in Kraft getreten, ist also 2 Jahre jünger als die EN ISO 9241-10. Die Norm baut auf Ergebnisse des EU-Forschungsprojekt MUSiC (Measuring Usability of Systems in Context) auf (Bevan & Macleod, 1994). Der dominierende Gedanke der Norm wie auch des Projekts ist, dass die Gebrauchstauglichkeit eines Werkzeugs immer nur im größeren Rahmen seines Nutzungseinsatzes bewertet werden kann. Entsprechend definiert die Norm Gebrauchstauglichkeit:

Das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen (EN ISO 9241-11, 1998).

Abb. 12 zeigt wie sich der Nutzungskontext gemäß der Norm auf die Maße der Gebrauchstauglichkeit auswirkt.

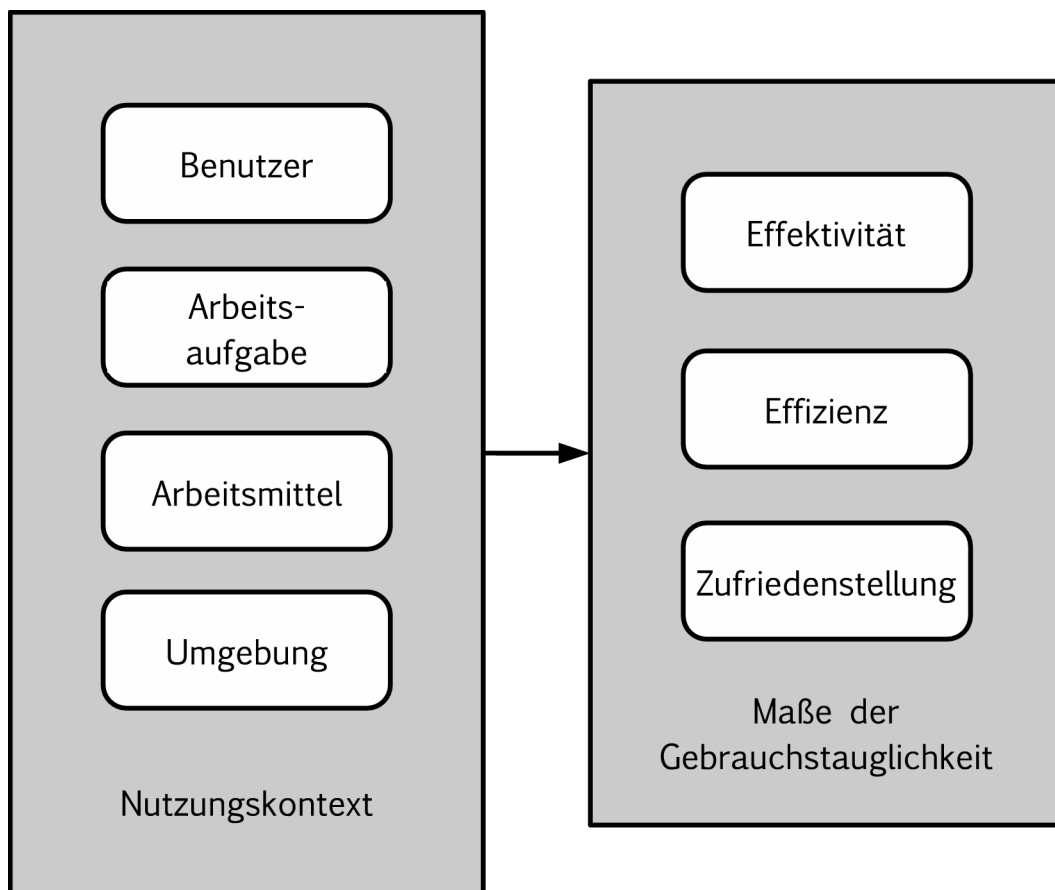


Abb. 12: Nutzungskontext und Usability gemäß EN ISO 9241-11 (1998)

Der Nutzungskontext umfasst den Benutzer, dessen Arbeitsaufgabe, das Arbeitsmittel und die Umgebung, in der die Arbeit erledigt wird. All diese Faktoren wirken sich auf die Gebrauchstauglichkeit aus. Für die Gebrauchstauglichkeit existieren drei relevante Maße: die Effektivität, die Effizienz und die Zufriedenstellung. Effektivität definiert die Norm als Grad, zu dem das angestrebte Ziel erreicht wird. Effizienz betrachtet den zur Zielerreichung eingesetzten Arbeitseinsatz bzw. Aufwand. Zufriedenstellung ist ein subjektives Maß, welches die erlebte Zufriedenheit des Benutzers mit dem System erfasst.

Die EN ISO 9241-11 stellt das Konzept Usability in einen breiteren Kontext. Sie weicht von den anderen Konzeptionen ab, indem sie ihren Fokus auf die Messung von Usability und nicht auf Gestaltungsgrundsätze zur Erreichung derselben legt. Usability wird nicht mehr auf die Benutzerschnittstelle beschränkt, sondern umfasst das ganze Informationssystem, einschließlich dessen Funktionalität, in seinem Nutzungskontext. Die Norm betrachtet Informationssysteme als Werkzeuge, die der Erreichung eines vom Benutzer definierten Ziels dienen. Sie operationalisiert Usability als Kombination aus Zielerreichung an sich (Effektivität), Aufwand, der zur Zielerreichung eingesetzt werden muss (Effizienz) und der subjektiven Zufriedenheit, die die Benutzer über die Benutzung angeben (Zufriedenstellung). Sie betrachtet die einzelnen Aspekte der anderen Konzepte wie z. B. Erlernbarkeit nicht generell, sondern in Abhängigkeit vom Nutzungskontext als Bestandteile von Usability. Der

Einfluss einzelner Aspekte wird durch den Aufgabenkontext und die Benutzergruppe bestimmt. Für unerfahrene Benutzer spielt die Erlernbarkeit eine wichtige Rolle, für erfahrene Benutzer wird sie irrelevant. Im Anhang B der Norm werden Vorschläge für die Aspekte Effizienz, Effektivität und subjektive Zufriedenheit für verschiedene Benutzergruppen und Einsatzzwecke aufgelistet. Hier finden sich die Aspekte der anderen Konzepte wieder und die Norm verweist explizit auf die EN ISO 9241-10. Als Rahmen für die Operationalisierung von Usability erscheint die EN ISO 9241-11 sinnvoll und kann bei der praktischen Entscheidung zwischen konkurrierenden Lösungen helfen.

Für die theoretische Konzeption des Konstrukts Gebrauchstauglichkeit ist die Norm nicht hilfreich, da sie keine Aussagen über die Voraussetzungen für Gebrauchstauglichkeit macht.

Zusammenfassung

Usability ist ein junger Begriff, der im Bereich des Zusammenwirkens von Endbenutzer und moderner Informationstechnologie angesiedelt ist. Die deutsche Übersetzung von Usability nach der EN ISO 9241-11 ist Gebrauchstauglichkeit, umgangssprachlich wird Usability häufig mit Benutzerfreundlichkeit übersetzt (u. a. dict.leo.org).

Verschiedene Definitionen von Usability unterscheiden sich erheblich in ihrer Betrachtungsweise der Mensch-Maschine-Interaktion. Dzida (1983) differenzierte verschiedene Bereiche der Interaktion mit einem Informationsverarbeitungssystem und führte entsprechend verschiedene Schnittstellen zwischen Benutzer und Informationssystem ein. Hier hat die Idee der Dialogschnittstelle ihren Ursprung. Gemäß Dzida kann diese von der Anwendungsart und dem Nutzungskontext unabhängig gestaltet werden. Eine solche Konzeption findet sich auch bei Ulich (1986) und Nielsen (1993), die die Bedienung eines Systems von inhaltlichen Aspekten der Interaktion zu trennen versuchen. Damit wird Usability auf die Gestaltung der Dialogoberfläche begrenzt. Diese Betrachtungsweise mündete in die EN ISO 92431-10 „Grundsätze der Dialoggestaltung.“ Eine andere Perspektive liegt der EN ISO 9241-11 „Anforderungen an die Gebrauchstauglichkeit – Leitsätze“ zugrunde. Der Computer wird als Werkzeug verstanden, das von einer Nutzergruppe zur Bearbeitung einer Aufgabe eingesetzt wird. Wie erfolgreich der Einsatz einer Software ist, hängt somit von ihrer Passung zu den Benutzern und deren Aufgabe, sowie ihrem Einsatzumfeld allgemein ab. Usability ist hier nicht eine Eigenschaft eines Informationsverarbeitungssystems an sich, sondern ist von der konkreten Einsatzsituation eines Systems abhängig.

Tab 3 zeigt eine Gegenüberstellung der einzelnen Aspekte verschiedener Usability-Konzeptionen.

EN ISO 9241-10	Ulich (1986)	Nielsen (1993)	EN ISO 9241-11
Aufgabenangemessenheit	Kompatibilität		
Selbstbeschreibungsfähigkeit	Unterstützung		
Lernförderlichkeit		Leicht zu erlernen	
Steuerbarkeit	Toleranz		
	Flexibilität		
Erwartungskonformität	Konsistenz		
Fehlertoleranz		Wenig Fehler	
Individualisierbarkeit	Individualisierbarkeit		
	Transparenz		
	Partizipation		
		Effizient zu benutzen	Effizienz
		Leicht zu merken	
		Angenehm für den Benutzer	Zufriedenstellung
			Effektivität

Tab 3: Vergleich verschiedener Beurteilungskategorien

Die EN ISO 9241-10 und Ulichs Konzeption zeigen ein hohes Maß an Übereinstimmung. Beide stellen einen schwachen Bezug zum Arbeitsumfeld her, in dem ein System eingesetzt wird. Die EN ISO 9241-10 nennt diesen Bezug *Aufgabenangemessenheit*, bei Ulich heißt dieses Kriterium *Kompatibilität*. Die EN ISO 9241-10 nennt die beiden Kriterien *Selbstbeschreibungsfähigkeit* und *Lernförderlichkeit*, die sich beide in Ulichs *Unterstützung* wiederfinden. Nielsens Beurteilungskriterium „*Leicht zu erlernen*“ entspricht der *Lernförderlichkeit* und einem Teil von Ulichs *Unterstützung*. Die *Steuerbarkeit* der EN ISO 9241 differenziert Ulich in *Toleranz* und *Flexibilität*, wobei sich *Toleranz* eher auf die Steuerbarkeit der Ergebnisausgabe und *Flexibilität* auf die Eingabemöglichkeiten des Benutzers bezieht. Die *Erwartungskonformität* der EN ISO 9241-10 umfasst Ulichs *Konsistenz*, dehnt das Konzept allerdings über die Dialogschnittstelle hinaus auf den Kenntnisstand des Benutzers aus, an dem sich ein Dialogsystem orientieren soll. Bei Ulich findet sich die Passung zur Arbeits- und Erfahrungswelt des Benutzers im Aspekt der *Kompatibilität* (der hier der *Aufgabenangemessenheit* gegenübergestellt

wurde). Die EN ISO 9241-10 berücksichtigt für die Bewertung der Usability die *Fehlertoleranz* eines Systems. Nach Nielsens Konzeption sollten Benutzerfehler möglichst nicht auftreten. Nielsen betrachtet unter *Wenig Fehler* die nicht-korrigierbaren Fehler gesondert. Nielsens *Wenig Fehler* entspricht insofern genau der *Fehlertoleranz* der EN ISO 9241-10. Die *Anpassbarkeit* des Dialogs an die Bedürfnisse des Benutzers ist als *Individualisierbarkeit* ein Kriterium der Usability in der EN ISO 9241-10 und bei Ulich. Wie schon bei *Erwartungskonformität/Konsistenz* definiert die EN ISO 9241-10 dieses Kriterium breiter als Ulich und bezieht im Gegensatz zu Ulich Aspekte der Arbeitsaufgabe des Benutzers mit ein. In Ulichs Ansatz sind direkte Rückmeldungen zum System-Status wichtig und werden unter dem Aspekt der *Transparenz* erfasst. Ulich sieht eine größtmögliche Nutzerbeteiligung am Entwicklungsprozess als grundlegende Voraussetzung für eine gelungene Schnittstellengestaltung. Entsprechend ist für ihn das Ausmaß der Partizipation ein Beurteilungskriterium der Dialoggestaltung. Nielsen und die EN ISO 9241-11 betrachten die *Effizienz*, mit der eine Aufgabenbearbeitung möglich ist, als Kriterium der Usability. Nielsen räumt in seiner Konzeption den Gelegenheitsbenutzern, die ein System nur sporadisch nutzen, besonderen Stellenwert ein und berücksichtigt deren Belange unter dem Aspekt *Leicht zu merken*. Die *Zufriedenheit* der Benutzer ist für Nielsen und die EN ISO 9241-11 ein Aspekt der Usability eines Systems. Die EN ISO 9241-11 betrachtet noch die *Effektivität* eines System als Indikator für Gebrauchstauglichkeit.

Alle betrachteten Usability-Konzepte haben ihre Ursprünge in anwendungsorientierter Forschung. Der Grundgedanke einer von der Anwendungssoftware losgelösten Dialogschnittstelle, wie in der EN ISO 9241-10, bei Nielsen und bei Ulich, ist vor dem Hintergrund damaliger Betriebssysteme ohne einheitliche Benutzeroberflächen nachvollziehbar, heute aber nicht mehr zielführend. Innerhalb der jeweiligen Konzeptionen sind die Aspekte von Usability nur unzureichend gegeneinander abgegrenzt. In Dzidas (Dzida et al., 1978) faktorenanalytischer Auswertung, die die Grundlage der EN ISO 9241-10 legte, luden verschiedene Items ähnlich hoch auf mehreren Faktoren. Die undeutliche Abgrenzung wird auch im Vergleich der Kriterien sichtbar: innerhalb des einen Konzepts wird zwischen Aspekten unterschieden, die in einem anderen Konzept zu einem Aspekt vereinigt sind. Die *Selbstbeschreibungsfähigkeit* der EN ISO 9241-10 beispielsweise beinhaltet Rückmeldungen zum Systemstatus, denen Ulich exklusiv das Kriterium der *Transparenz* zuordnet. Die EN ISO 9241-10 differenziert zwischen *Selbstbeschreibungsfähigkeit* und *Lernförderlichkeit*, obwohl Teile der Definition zur *Selbstbeschreibungsfähigkeit*, „wenn jeder einzelne Dialogschritt durch Rückmeldung des Dialogsystems unmittelbar verständlich ist oder dem Benutzer auf Anfrage erklärt wird“, auch als Definition von *Lernförderlichkeit* vorstellbar wären. Fehler spielen beim menschlichen Lernen eine wichtige Rolle (Senghaas, 2004). Die EN ISO 9241-10 listet *Fehlertoleranz* als Kriterium für die Usability eines Systems auf. Nielsen unterscheidet zwischen leicht korrigierbaren und schwerwiegenden Fehlern. In

Nielsens Konzeption wirken sich die leicht korrigierbaren Fehler auf die *Effizienz* aus und müssen nicht gesondert erfasst werden. Schwerwiegende Fehler sollten möglichst nicht auftreten. Beide Ansätze übersehen das Wechselspiel zwischen dem Auftreten von Fehlern und Lernfortschritt, obwohl Carroll und Mack dies bereits 1984 für Benutzer einer Textverarbeitung als charakteristisch für die Lernphase beschrieben.

Das Konzept Usability hat sich aus der Notwendigkeit heraus entwickelt, Computer angesichts ihrer zunehmenden Verfügbarkeit und Verbreitung für eine breite Benutzerschicht zugänglich zu machen. Die existierenden Konzeptionen sind für den praktischen Einsatz entwickelt und offenbaren bei genauerer Betrachtung theoretische Unklarheiten.

Usability vor dem Hintergrund mentaler Modelle

Im Folgenden wird überprüft, inwieweit sich die Gestaltungsregeln und Bewertungskriterien von Usability komplexer Systeme durch eine Theorie mentaler Modelle erklären lassen. Bei Erlernen der Benutzung spielt der Aufbau geeigneter mentaler Modelle – häufig über Analogien und Metaphern – eine wichtige Rolle. Für den erfahrenen Benutzer ist die entscheidende Frage, ob er funktional günstige mentale Modelle gebildet hat, die für ihn wichtige Funktionen eines Systems repräsentieren.

Mentale Modelle und das Erlernen der Benutzung

Dem Gebrauch neuer Informationstechnologie ist immer das Erlernen der Benutzung vorgeschaltet. Die Lernphase eines unerfahrenen Benutzers ist im Rahmen einer Theorie mentaler Modelle die Zeit, in der neue Modelle für das zu erlernende System gebildet und überprüft werden. Im Rahmen dieses Vorgangs werden existierende Schemata gegebenenfalls modifiziert oder es müssen neue gebildet werden. In der Praxis gibt es keine abgeschlossene Lernphase. Auch erfahrene Benutzer können noch weiter lernen, allerdings verschiebt sich der Schwerpunkt mit zunehmender Systemerfahrung von der Bildung geeigneter mentaler Modelle hin zu deren Nutzung.

Bis auf die EN ISO 9241-11 beinhalten alle Konzeptionen von Usability Aspekte der Erlernbarkeit. Die EN ISO 9241-10 fordert *Lernförderlichkeit*, Ulich *Unterstützung* und *Transparenz* und Nielsen, dass Systeme leicht zu erlernen sind. Das Fehlen der Erlernbarkeit in der EN ISO 9241-11 kann durch deren Orientierung auf den Einsatz von Software erklärt werden. Die Norm betrachtet die *Effizienz* und *Effektivität* von Software in einem vorgegebenen Nutzungskontext, zu dem auch die Benutzer gehören. In einem konkreten Nutzungskontext mit unerfahrenen Benutzern, also beispielsweise der Benutzung eines Informationssystems für Museumsgäste, würde die Erlernbarkeit sich unmittelbar auf Effizienz, Effektivität und Zufriedenheit auswirken. Ein solcher Audioguide für Museumsgäste bräuchte nur wenige

Funktionen (Lautstärkeregelung und die Anwahl von zum Exponat passenden Erläuterungen mit der Möglichkeit, das Abspielen zu unterbrechen oder zu beenden), sollte aber von verschiedensten Benutzern mit unterschiedlichem Vorwissen bedienbar sein. Insofern berücksichtigt auch die EN ISO 9241-11 die Erlernbarkeit, allerdings variabel und abhängig von der Benutzergruppe.

Die Bildung mentaler Modelle bzw. neuer Schemata erfolgt nicht aus dem Nichts. Menschen bringen ihr vorhandenes Wissen in neue Situationen mit und probieren, es dort zu nutzen. Dabei wird versucht, schematisches Wissen aus einem bekannten Bereich über eine Analogiebeziehung auf einen neuen Bereich zu übertragen. Dies nutzten erste graphische Bedienoberflächen wie Xerox-Star oder LISA und Macintosh von Apple gezielt aus und gestalteten die Bedienoberfläche wie einen Schreibtisch (Spinas, Waeber & Strohm, 1990). Dieses Konzept hat sich durchgesetzt und findet sich auch bei heutigen Betriebssystemen.

Für neue Informationstechnologien kann in hohem Maße eine generelle Vorerfahrung der Benutzer mit graphischen Bedienoberflächen vorausgesetzt werden. Eine Bericht von TNS-Infratest (2007) weist für Deutschland 60,2 % Internetbenutzer aus, die entsprechend über allgemeine Computererfahrung verfügen. Auch aus diesem Bereich wird schematisches Wissen auf neue Bedienoberflächen übertragen. Seit der Einführung graphischer Benutzeroberflächen (GUI, vom englischen Graphical User Interface) hat sich eine Reihe von Bedienelementen als Standard etabliert. Für die Betriebssysteme regeln sogenannte Styleguides, wie Schaltflächen, andere Kontrollelemente und Menüs auszusehen und sich zu verhalten haben. Eine aktuelle Linksammlung zu Styleguides verbreiteter Betriebssysteme und Programmiersprachen findet sich beispielsweise unter <http://developer.gnome.org/projects/gup/references.html>. Der Grundgedanke einer von der konkreten Anwendung unabhängigen Bedienoberfläche, wie er von Dzida (1983) vorgeschlagen wird, ist in den graphischen Oberflächen heutiger Betriebssysteme von Apple, Microsoft sowie Linux und Unix weitgehend verwirklicht. Für alle Systeme finden sich Fenster, in denen Anwendungen laufen, deren Steuerung über Menüs und Schaltflächen erfolgt, die überall gleich reagieren. Reicht der Bildschirm für die Darstellung nicht aus, kann der Bildschirminhalt über eine Scroll-Leiste verschoben werden. Je nach allgemeiner Erfahrung mit einem Computer sind bei einem Anwender also bereits eine Vielzahl von Schemata vorhanden, wie Bedienelemente von GUIs funktionieren. Diese werden für die Bildung neuer mentaler Modelle unbekannter Systeme herangezogen.

Der Einsatz von moderner Informationsverarbeitungstechnologie erfolgt meist zweckmäßig, die Systeme sollen bei der Erledigung von Aufgaben helfen. Das Wissen über die Aufgabe sowie deren Umfeld ist eine zweite Quelle für Schemata, die bei der Erzeugung neuer mentaler Modelle genutzt wird. Eine Bürokräft, die das Erstellen von Texten mittels Textverarbeitung lernt, wird versuchen, ihr vorhandenes Wissen über das Schreiben mit einer Schreibmaschine auf die neue Situation zu

übertragen (Douglas & Moran, 1984). Das Übertragen von vorhandenem Wissen kann zu Problemen führen, wenn sich das neue System abweichend von den bekannten Schemata verhält (Douglas & Mouran, 1984; Dutke, 1994).

Dutke (1994) nennt zwei prinzipielle Möglichkeiten, wie die Bildung mentaler Modelle von außen beeinflusst werden kann. Dem Benutzer können Analogien aufgezeigt werden oder ihm wird ein konzeptionelles Modell des Systems vermittelt.

Eine Analogie kann innerhalb einer Schulung vermittelt werden, über Begleitmaterial wie Handbücher oder durch die Gestaltung des Systems selbst. Abb. 13 zeigt, wie durch die Gestaltung der Nutzungsoberfläche die Analogie zwischen einem Taschenrechner und einem entsprechenden Computerprogramm hervorgehoben wird. Ein anderes Beispiel, in dem eine Analogie zu einem Taschenrechner hergestellt wird, stammt aus einem Handbuch zu dem Statistikprogramm, mit dem die Auswertungen dieser Arbeit gerechnet wurden: Dalgaard (2002) bezeichnet in einer Kapitelüberschrift das Statistikpaket R als „an overgrown calculator“ (ein überdimensionierter Taschenrechner, Übers. d. Verf.).

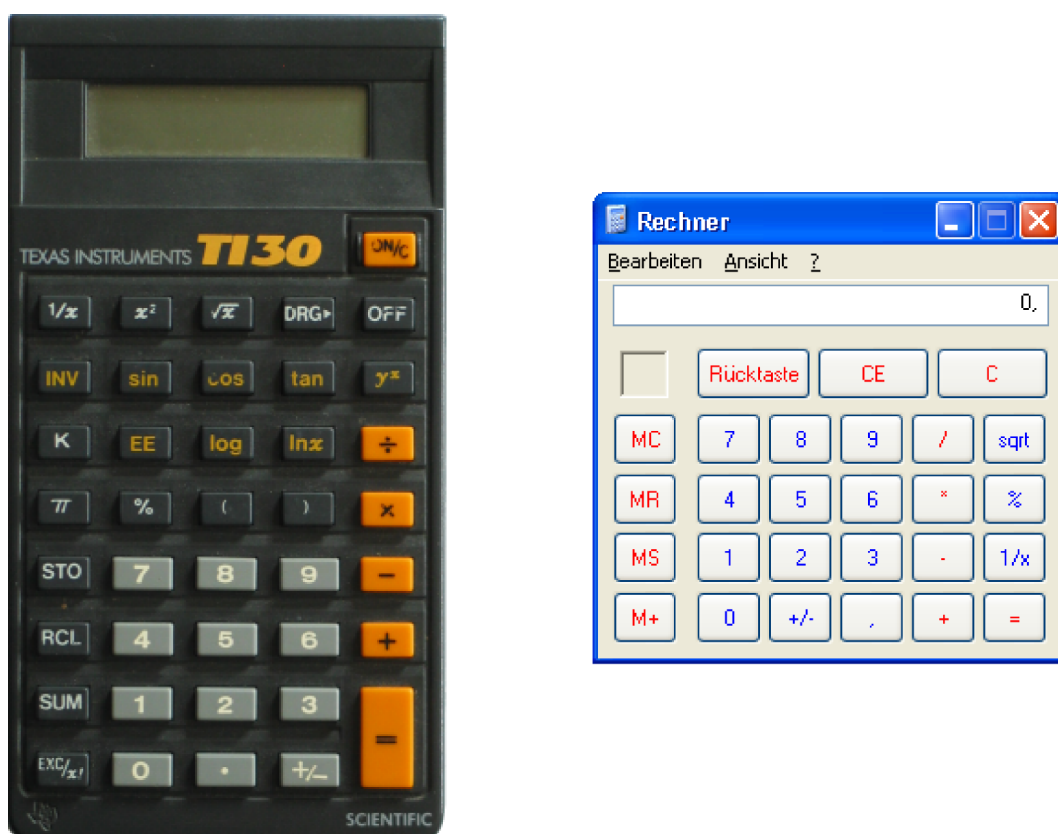


Abb. 13: Bedienoberfläche eines Taschenrechners und einer analog gestalteten Software

Vermittlung konzeptioneller Modelle

Je nach Art der Analogie, die vermittelt wird, unterscheiden sich die ausgebildeten mentalen Modelle.

Neu gebildete mentale Modelle werden fortlaufend überprüft (Johnson-Laird, 1983). Abb. 14 zeigt, wie eine Modellbewertung auf Grundlage der Modellanwendung zu Veränderungen am Modell führen kann. Wird das Modell durch seine Anwendung bestätigt, so verfestigen sich die Modellannahmen. Zeigen sich Widersprüche, so wird versucht, die Modelle so umzugestalten, dass die Widersprüche aufgelöst werden (Penner, 2001). Bei entsprechend großen Widersprüchen kann schließlich eine Neubildung des Modells stattfinden.

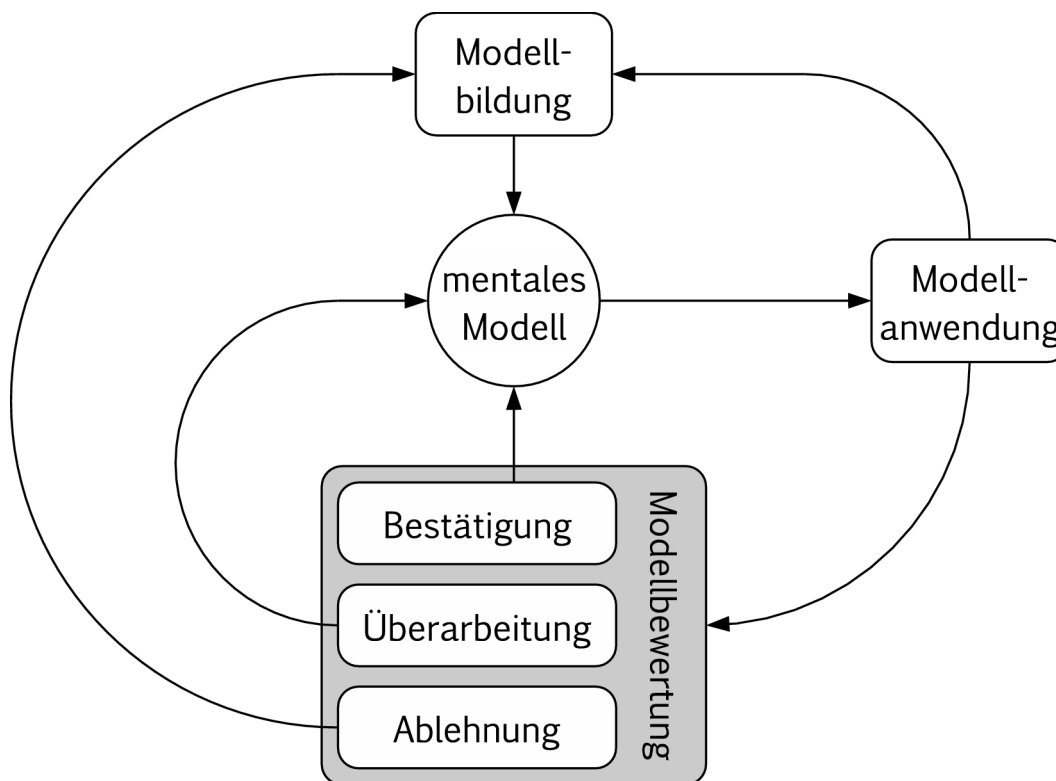


Abb. 14: Prozesse bei der Entwicklung eines mentalen Modells (nach Seel, 2003)

Von zentraler Bedeutung für die Weiterentwicklung mentaler Modelle ist das Auftreten von Fehlern als Rückmeldung über mangelnde Modellgüte. Erweisen sich die Prognosen, die aufgrund der benutzten mentalen Modelle gebildet wurden, als falsch, müssen die Modelle verändert werden. Carroll und Mack (1984) beschreiben diesen Prozess am Beispiel des Erlernens der Benutzung einer Textverarbeitungssoftware. Die angehenden Benutzer bildeten Erwartungen über das Verhalten des Systems aufgrund rudimentärer mentaler Modelle. Häufig waren diese Erwartungen – dem Entwicklungsstand der mentalen Modelle entsprechend – falsch und führten zu Fehlern. Dies gehört zu einem normalen Lernprozess. An den in der Studie benutzten Textverarbeitungsprogrammen kritisierten Carroll und Mack, dass die Korrektur von Fehlern häufig nicht einfach und problemlos möglich war. Sie fordern Programme, in denen sich Fehler sicher, schnell und einfach rückgängig machen lassen, damit Lernen aus Fehlern problemlos möglich ist.

Zusammenfassend kann Lernen als Entwicklung neuer mentaler Modelle bzw.

als Anpassung und Entwicklung zugrunde liegender Schemata verstanden werden. Dieser Prozess wird erleichtert, wenn vorhandenes schematisches Wissen sich auf den neuen Bereich übertragen lässt, sich also beispielsweise ein Computer ähnlich wie ein bekanntes Objekt (z. B. eine Schreibmaschine) verhält. Die neuen Modelle werden im Hinblick auf ihre Brauchbarkeit überprüft. Diese Bewertung hängt von der Rückmeldung ab. Treten Fehler auf, muss das Modell verändert werden, um das unerwartete Verhalten zukünftig zu erklären. Das Auftreten von Fehlern und die daraus resultierenden Modifikationen sind also für die Entwicklung mentaler Modelle essentiell.

Mentale Modelle bei der Nutzung moderner Informationsverarbeitung

Die Bearbeitung von Aufgabenstellungen mit einem Informationsverarbeitungssystem erfolgt anhand eines Handlungsplans, der auf der Grundlage des mentalen Modells des Systems entwickelt wurde. Probleme lassen sich dann effizient lösen, wenn das System dafür benötigte Funktionen bereitstellt und das mentale Modell des Benutzers diese Funktionen abbildet. Anders gesagt können Probleme beim Arbeiten mit einem Informationsverarbeitungssystem am System selbst liegen oder durch ein unpassendes mentales Modell verursacht werden.

Vicente (1997) hinterfragt die Designempfehlung, die Bedienoberfläche eines Systems an das mentale Modell der Benutzer anzupassen. Er berichtet von der Gestaltung des Kontrollraums eines Kernkraftwerks unter Mitwirkung eines erfahrenen Schichtleiters. Die fertige Gestaltung des Kontrollraums wurde von anderen Schichtleitern kritisiert. In einer anschließenden Analyse zeigte sich, dass die mentalen Modelle aller Schichtleiter im Vergleich zu den physikalischen Zusammenhängen im Kraftwerk unvollständig waren und Fehler enthielten. Orientiert sich das Design eines Kontrollstands nicht an den tatsächlichen Gegebenheiten eines technischen Systems, sondern an den mentalen Modellen seiner Benutzer, so verstärkt diese Gestaltung die falschen Annahmen. Dies gilt nicht nur für komplexe System wie den von Vicente betrachteten Kontrollstand, sondern entsprechend auch für einfachere Bedienoberflächen. Die graphische Darstellung der Möglichkeit, Dateien zu löschen, in Form eines Papierkorbs kann falsche Vorstellungen beim Benutzer erwecken. In einer realen Büroumgebung wird ein Papierkorb vom entsprechenden Personal regelmäßig – häufig täglich – geleert. Dokumente, die in den Papierkorb geworfen wurden, sind entsprechend nach kurzer Zeit nicht mehr verfügbar. Der virtuelle Papierkorb graphischer Benutzeroberflächen verhält sich meist anders: er hat ein bestimmtes Volumen und kein fixes Leerintervall. Dateien werden erst gelöscht, wenn der Papierkorb voll ist. Tritt dieser Fall ein, werden so viele der ältesten Dateien aus dem Papierkorb gelöscht bis die neu gelöschten Dateien in den Papierkorb passen. Hieraus ergibt sich eine stark schwankende Verweildauer der gelöschten Objekte im Papierkorb, die nicht vorhergesagt werden kann. Eine andere Gestaltung, die dem Benutzer das Konzept des festgelegten

Volumens vermitteln würde, würde zu einem passenderen mentalen Modell führen.

Ein anderes Problem ist die mangelnde Ausbildung mentaler Modelle, die vorhandene Funktionalität des Systems unter Umständen nicht abbilden. Das mentale Modell eines Benutzers muss nicht alle Funktionen eines Programms umfassen. Heutige Anwendungsprogramme sind so umfangreich, dass die meisten Benutzer nur einen Bruchteil der zur Verfügung stehenden Funktionen nutzen (Draper, 1984). Ein angemessenes mentales Modell sollte allerdings alle für die typischen Arbeiten des Benutzers nützlichen Funktionsweisen der Software enthalten.

Die mentalen Modelle sind häufig lückenhaft, wie Untersuchungen von Norman (1983) zur Benutzung von Taschenrechnern zeigten. Die Versuchspersonen hatten meist keine klaren Vorstellungen zu den Speichermöglichkeiten der Taschenrechner und wie diese zu handhaben sind. Zwischenergebnisse wurden teilweise auf Papier notiert, anstatt sie im Speicher des Rechners abzulegen. Um den Speicher zu löschen wurden die verschiedenen Lösch Tasten mehrmals betätigt. Die Versuchspersonen waren sich allerdings zumeist bewusst, dass dies keine optimale Benutzungsstrategie darstellt. Das gezeigte Verhalten hat den Vorteil über verschiedene Taschenrechnermodelle hinweg zum gewünschten Ergebnis zu gelangen. Es ist nicht elegant im Sinne einer optimalen Lösung, sondern im Anwendungsalltag robust. Vor diesem Hintergrund ist die Frage nach einem optimalen mentalen Modell eines erfahrenen Benutzers schwierig. Umwege, die robust zu einer Problemlösung führen, können nicht ohne weiteres auf unpassende mentale Modelle zurückgeführt werden.

Das mentale Modell eines Benutzers bildet die Grundlage seiner Interaktion mit einem komplexen System. Schlechte Lösungsstrategien können auf unvollständige bzw. unpassende Modelle zurückgeführt werden. In einem unvollständigen Modell sind nicht alle Funktionen eines Systems abgebildet. In einem unpassenden Modell sind Funktionen falsch abgebildet. Das mentale Modell eines Benutzers kann nicht direkt beobachtet werden. Als Vorstellung hat das mentale Modell eines Benutzers erheblichen Erklärungswert, indem es komplexes Benutzerverhalten nachvollziehbar macht.

Vergleich einer Theorie mentaler Modelle mit den Usability-Konzepten

Im Folgenden werden die Aspekte der verschiedenen Usability-Konzepte vor dem Hintergrund der Theorie mentaler Modelle analysiert. Dabei wird betrachtet, inwieweit diese Aspekte mit einem theoretischen Grundgerüst mentaler Modelle in Einklang stehen oder ob gravierende Abweichungen festzustellen sind. Abschließend wird beurteilt, inwiefern sich der Ansatz mentaler Modelle zur theoretischen Weiterentwicklung des Konstrukts Usability eignet.

Die meisten der betrachteten Usability-Konzepte stellen einen Bezug zum schematischen Wissen über das Arbeitsgebiet der Benutzer her. Für das Kriterium der *Aufgabenangemessenheit* der EN ISO 9241-10 muss eine Bedienoberfläche „der

Komplexität der Arbeitsaufgabe unter Berücksichtigung der Fertigkeiten und Fähigkeiten des Benutzers Rechnung“ tragen. Hinsichtlich der *Selbstbeschreibungsfähigkeit* soll die Bedienoberfläche eine Sprache benutzen, „die sich aus dem Arbeitsgebiet ableitet“. Für *Erwartungskonformität* müssen „Fachausdrücke [...] die gleichen [sein], wie sie im Bereich der Arbeitsaufgabe des Benutzers tatsächlich verwendet werden.“ Ähnliche Gestaltungsregeln formuliert Ulich (1986) hinsichtlich seines Konzepts *Kompatibilität*. Die „Sprache und begriffliche Komplexität des Dialogs sollten an den Gepflogenheiten und Kenntnissen des spezifischen Benutzerkreises orientiert sein.“ Soweit vorhanden sollten Formulare oder andere gedruckte Unterlagen als Vorbilder für die graphische Gestaltung dienen. Für Ulich ist die *Partizipation* der Benutzer an der Systementwicklung die geeignete Methode dies sicherzustellen. Nielsen (1993) ist bezüglich des vorhandenen schematischen Wissens am deutlichsten und fordert:

The user's model of the task should also be identified, since it can be used as a source for metaphors for the user interface.

(Das Modell der Aufgabe des Benutzers sollte bestimmt werden, weil es als Quelle für Metaphern der Benutzerschnittstelle dienen kann, Übers. d. Verf.)

Die Usability eines Systems profitiert also davon, wenn sich die Bedienung an Abläufen der Arbeitsaufgabe orientiert und Wissen über die Arbeitsaufgabe berücksichtigt. Dies deckt sich mit den angestellten Überlegungen zum Lernen als Bildung neuer mentaler Modelle. Das vorhandene Wissen kann in Form von Schemata die Entwicklung neuer mentaler Modelle erleichtern.

Ein Großteil der Aspekte von Usability in den verschiedenen Konzeptionen betrachtet die innere Schlüssigkeit der Bedienoberfläche. So formuliert die EN ISO 9241-10 zu *Erwartungskonformität*:

Ein Dialog ist erwartungskonform, wenn er konsistent ist [...].

Die Norm listet Beispiele für Gestaltungen auf, die sich innerhalb der Bedienoberfläche immer gleich verhalten bzw. auch immer gleich gestaltet sein sollen, wenn sie dem selben Zweck dienen, z. B.:

Um in einem Dialogsystem mit mehreren Anwendungsprogrammen eines davon zu starten, muss der Benutzer stets das betreffende Symbol doppelklicken.

oder:

Verschiebbare Fenster werden stets durch Ziehen eines bestimmten Bereiches in der Nähe der Oberkante des Fensters bewegt.

Hinsichtlich *Lernförderlichkeit* formuliert die Norm:

Regeln und zugrundeliegende Konzepte, die für das Erlernen nützlich sind, sollten dem Benutzer zugänglich gemacht werden, damit dieser sich eigene Ordnungsschemata und Merkgeln aufbauen kann.

Ulichs Aspekt der *Konsistenz* betrachtet die Konsistenz der Bedienoberfläche. Dabei sollen „ähnliche Aktionen [...] ähnliche Ausführungen bewirken.“ Insgesamt

sollte das System „transparent und konsistent“ sein. Nielsen leitet ein Kapitel über Konsistenz ein:

Consistency is one of the most basic usability principles. If users know that the same command or the same action will always have the same effect, they will feel more confident in using the system, and they will be encouraged to try out exploratory learning strategies because they will already have part of the knowledge needed to operate new parts of the system.

Konsistenz ist eines der grundlegendsten Usability-Prinzipien. Wenn die Benutzer wissen, dass derselbe Befehl bzw. dieselbe Aktion immer dasselbe bewirkt, steigt ihr Zutrauen und sie werden ermutigt, Dinge auszuprobieren, weil sie teilweise schon wissen, wie sie mit neuen Teilen des Systems umgehen müssen. (Nielsen, 1993, S. 132, Übers. d. Verf.)

Damit ist für Nielsen *Konsistenz* ein zentraler Aspekt der Erlernbarkeit.

Für den Lernprozess ist das Entdecken von Regelmäßigkeiten und Mustern ein essentieller Schritt. Wissen über Regelmäßigkeiten und Muster der Umwelt wird im Gedächtnis als schematisches Wissen abgespeichert. Damit schematisches Wissen erlernt werden kann, sollten leicht entdeckbare Muster vorhanden sein bzw. das Systemverhalten nach Regeln konzipiert werden. Die Regeln an sich können in vielerlei Hinsicht variabel gestaltet werden. Es ist prinzipiell unerheblich, ob die Schaltfläche zum Schließen eines Fensters rechts oder links oben angeordnet wird. Wichtig ist, dass sie immer an der gleichen Stelle angeordnet ist, also entweder immer links oder – wie sich in der Praxis durchgesetzt hat – immer rechts. Dies zu erreichen ist die Funktion der bereits erwähnten Styleguides, die entsprechende Gestaltungsregeln festlegen. Wird die Regelmäßigkeit durch ein entsprechendes Design eines Systems gezielt erhöht, verbessert sich die Erlernbarkeit. Entdeckte Regeln werden als schematisches Wissen im Gedächtnis gespeichert. Bei der aktuellen Benutzung wird das mentale Modell des Systems basierend auf diesen Regeln aufgebaut. Eine möglichst weitreichende Gültigkeit dieser Regeln erleichtert die Bildung passender mentaler Modelle. Erlerntes schematisches Wissen kann für die Bildung eines mentalen Modells einer neuen Anwendung überall dort genutzt werden, wo diese Anwendung auf Standardbedienelemente zurückgreift, die der Benutzer aus anderen Anwendungen schon kennt. Die von den verschiedenen Konzepten für Usability geforderte Konsistenz von Bedienoberflächen steht daher im Einklang mit einer Betrachtung vor dem Hintergrund mentaler Modelle.

Gegenüber den unterschiedlichen Usability-Konzepten schafft der theoretische Rahmen mentaler Modelle zusätzliche Klarheit. Die in allen Usability-Konzepten geforderte Ähnlichkeit und Konsistenz der Bedienoberfläche kann sich generell auf zwei unterschiedliche Bereiche beziehen:

- das Arbeitsumfeld bzw. die Arbeitsaufgabe (wie beispielsweise eine Schreibmaschine oder der Papierkorb im Büro)
- eine dem Benutzer vertraute Bedienoberfläche (wie beispielsweise die graphische

Benutzeroberfläche von Windows-Betriebssystemen)

Entsprechende Gestaltungsregeln sind an verschiedenen Stellen in den verschiedenen Usability-Konzepten zu finden. Allerdings wird dort der übergeordnete Bezug zu vorhandenem schematischen Wissen des Benutzers nicht herausgearbeitet. Die beiden geeigneten Quellen schematischen Wissens – das Arbeitsumfeld des Benutzers und seine Kenntnisse über Bedienoberflächen – werden ebenfalls nicht deutlich genannt und teilweise in der Darstellung vermengt.

Die Bedeutung der Erlernbarkeit eines Systems für dessen Usability wird von den verschiedenen Usability-Konzepten nicht einheitlich bewertet. Während die EN ISO 9241-10, Ulich und Nielsen darin übereinstimmen, dass Erlernbarkeit wichtig für Usability ist, macht die EN ISO 9241-11 dies von der Benutzergruppe und dem Nutzungskontext ab. Nach der EN ISO 9241-11 ist die Erlernbarkeit für unerfahrene Benutzer und Gelegenheitsnutzer wichtig, aber weniger für erfahrene Benutzer. Auch Nielsen unterscheidet hinsichtlich der Erlernbarkeit zwischen unerfahrenen und erfahrenen Benutzern, differenziert jedoch noch weiter. Generelle Erfahrungen mit Computern spielen eine Rolle wie auch die Kenntnisse über das Aufgabengebiet. Nielsen kommt mit seiner Analyse des Vorwissens potenzieller Benutzer zu vergleichbaren Quellen nützlicher Vorerfahrung wie die Betrachtung eines Lernprozesses basierend auf vorhandenem schematischen Wissen.

Vor dem Hintergrund mentaler Modelle ist der Lernprozess durch die Bildung neuer mentaler Modelle gekennzeichnet. Der erfahrene Benutzer hat stabile Modelle ausgebildet, die ihm eine zielgerichtete Nutzung des Systems ermöglichen. Für die Effizienz der Systembenutzung ist – neben den objektiven Möglichkeiten des Systems – die Güte des Modells entscheidend. Ein brauchbares Modell bildet für den Benutzer nützliche Systemfunktionalität in einer angemessenen Weise ab.

Eine Betrachtung vor dem Hintergrund mentaler Modelle legt also eine Unterscheidung zwischen unerfahrenen und erfahrenen Nutzern nahe. Bei den unerfahrenen Benutzern liegt der Schwerpunkt auf der Bildung geeigneter mentaler Modelle. Für die erfahrenen Benutzer ist die Güte der gebildeten Modelle maßgebend. Sie entscheidet über die Effizienz der Systemnutzung eines erfahrenen Benutzers. So betrachtet kann eine unbefriedigende Leistung erfahrener Benutzer zwei verschiedene Gründe haben:

- die Funktionalität des Systems ist für die Aufgabe unpassend
- im mentalen Modell erfahrener Benutzer ist entsprechende Funktionalität des Systems nicht oder unpassend abgebildet

Das mentale Modell eines erfahrenen Benutzers ist das Ergebnis von Lernprozessen. Eine Abgrenzung zwischen der Lernphase und einer Anwendungsphase ist entsprechend dieser Abhängigkeit schwierig. Unpassende mentale Modelle können – per Definition von Erlernbarkeit im Kontext mentaler Modelle – auf Mängel der Erlernbarkeit zurückgeführt werden, die ja die Ausbildung

eines passenden Modells fördern soll. Der theoretische Hintergrund mentaler Modelle bildet die zugrunde liegenden Prozesse ab und kann so zu einer differenzierten Betrachtung unerfahrener und erfahrener Benutzer verhelfen.

Sowohl in der EN ISO 9241-10 als auch von Ulich wird eine Steuerung der Interaktion durch den Benutzer gefordert. Die Erledigung einer Aufgabe sollte an beliebiger Stelle unterbrochen und wieder fortgesetzt werden können. Die Eingabereihenfolge von Formulardaten sollte frei wählbar sein.

Falls die *Steuerbarkeit* eines Systems eingeschränkt ist, muss der Benutzer diese Einschränkungen erlernen, er muss sie in sein mentales Modell integrieren. So muss jeder Computerbenutzer lernen, dass er ein Dokument sinnvollerweise erst abspeichert, nachdem er Daten eingegeben hat. Falls das System eine unnötig rigide Handlungsabfolge vorgibt, muss der Benutzer diese ebenfalls erlernen. Das zur Bedienung des Systems notwendige mentale Modell muss komplizierter werden und diese Einschränkungen des Systems abbilden. Passt die Reihenfolge der Interaktion nicht zum Arbeitsablauf des Benutzers, so muss dieser Lösungswege suchen, wie er trotzdem zum gewünschten Ergebnis kommt. Die Entwicklung des Lösungsweges ist mit Aufwand verbunden. Zusätzlich kann die Ausführung der Aufgabe aufwändig sein, wenn der gefundene Lösungsweg zusätzliche Arbeitsschritte erfordert, die bei einer vom Benutzer wählbaren Abfolge der Arbeitsschritte nicht notwendig wären.

Ein weiterer Aspekt, der die *Steuerbarkeit* auch aus der Perspektive mentaler Modelle unverzichtbar macht, ist die Möglichkeit eines Benutzers, seine auf dem Modell basierenden Hypothesen zu überprüfen. Ein System mit einem vorgegebenen Interaktionsablauf schränkt den Benutzer hierbei stark ein. Zwänge beispielsweise ein fiktives Textverarbeitungsprogramm seine Benutzer zur Angabe eines Dateinamens und Speicherorts für ein neues Dokument, bevor Text eingegeben werden kann, so könnten Benutzer die Vermutung entwickeln, das Programm würde jeglichen eingegebenen Text unmittelbar in der spezifizierten Datei auf der Festplatte speichern. Werden die Eingaben hingegen – wie bei gängigen Textverarbeitungsprogrammen üblich – zunächst nur im flüchtigen Arbeitsspeicher des Computers abgelegt, führt diese Vorstellung zu unerwarteten Ergebnissen, wenn der Rechner ohne vorherige Sicherung des Arbeitsspeicherinhalts auf die Festplatte ausgeschaltet wird. Die Eingaben des Benutzers wären dann verloren, die spezifizierte Datei leer. Die in diesem Beispiel ohne sachlichen Zwang vorgegebene Reihenfolge wirkt sich also potenziell auf die Hypothesenbildung der Benutzer aus und schränkt ihre Möglichkeiten zur Überprüfung dieser Hypothesen ein. Eine Betrachtung von Usability auf der Grundlage von Hypothesenbildung und -überprüfung, wie sie in einer Theorie mentaler Modelle vorkommt, führt zu den gleichen Schlussfolgerungen bezüglich *Steuerbarkeit* wie sie in der EN ISO 9241-10 und bei Ulich zu finden sind.

Das Auftreten von Fehlern deutet Nielsen (1993) pauschal als Zeichen

mangelnder Usability. Die EN ISO 9241-10 ist diesbezüglich zurückhaltender und fordert *Fehlertoleranz* eines Systems. Diese Forderung nach *Fehlertoleranz* erscheint praxisgerechter, da sie der Rolle von Fehlern im Lernprozess Rechnung trägt. In keiner Konzeption wird die Rolle von Fehlern für den Lernprozess deutlich. Fehler werden in der Lernphase immer auftreten und gehören zum Prozess der Bildung und Überprüfung eines mentalen Modells. So betrachtet muss unterschieden werden zwischen Fehlern, die in der Lernphase auftreten, und solchen, die dauerhaft zu beobachten sind. Treten dauerhaft Fehler auf, können zugrundeliegende Konzepte für den Benutzer zu komplex sein. Der Benutzer kann kein geeignetes Modell des Systemverhaltens entwickeln und wird immer wieder von unerwartetem Systemverhalten überrascht. Eine zweite Möglichkeit kann ein Widerspruch zwischen bewährtem schematischem Wissen der Benutzer und dem zugrundeliegenden Muster des Systems sein. Die Vorstellung eines mentalen Modells des Benutzers ermöglicht eine detaillierte Betrachtung von Fehlern, die auf ein unvollständiges oder unpassendes mentales Modell zurückgeführt werden können.

Die EN ISO 9241-10 und Ulichs Konzept betrachten die Individualisierbarkeit als Aspekt der Usability. Allerdings schränkt die Norm ein:

Obwohl es in vielen Fällen sehr wünschenswert ist, dem Benutzer anpassbare Dialogfunktionen zur Verfügung zu stellen, ist dies kein Ersatz für ergonomisch gestaltete Dialoge.

Die Vorschläge für die *Individualisierbarkeit* reichen von einer einstellbaren Vereinfachung der Benutzerschnittstelle für unerfahrene Benutzer bis zur Programmierbarkeit von Makros für wiederkehrende Aufgaben erfahrener Benutzer. Sowohl die Norm wie auch Ulich lassen offen, was genau für welche Benutzer individualisierbar sein soll.

Vor dem theoretischen Hintergrund mentaler Modelle wäre es vorteilhaft, unerfahrenen Benutzern nur eine vereinfachte Bedienoberfläche mit wenigen Funktionen zu präsentieren. Dies könnte die Modellbildung erleichtern, weil weniger modelliert werden kann und muss (Carroll & Mack, 1984). Allerdings weisen schon Carroll und Mack auf die Schwierigkeiten hin, die Komplexität der Bedienoberfläche automatisch an den Kenntnisstand der Benutzer anzupassen. Außerdem würde hiermit die Steuerbarkeit durch den Benutzer eingeschränkt. Für eine Regulierung durch den Benutzer fehlen diesem jedoch gerade in der Lernphase die Voraussetzungen. Das Problem des Benutzers besteht genau darin, nicht zwischen für ihn Wichtigem und Unwichtigem unterscheiden zu können. Entsprechend ist er überfordert, die Auswahl selbst zu treffen.

Die Möglichkeit, Makros aufzuzeichnen, geht über die normale Benutzung eines Systems hinaus. Obwohl Makros eine sinnvolle Erweiterung der Funktionalität einer Anwendung ermöglichen können, ist eine Betrachtung im Rahmen dieser Arbeit nicht möglich.

Im Hinblick auf *Individualisierbarkeit* bietet eine theoretische Fundierung durch

mentale Modelle interessante Betrachtungsperspektiven: Individualisierung kann als Anpassung der Bedienoberfläche an das mentale Modell des Benutzers verstanden werden. *Individualisierbarkeit* wirft im Hinblick auf Usability Schwierigkeiten auf, die sich durch die wechselseitige Abhängigkeit von mentalem Modell und den Möglichkeiten zur Selbststeuerung des Lernprozesses ergeben. Erst ein passendes mentales Modell ermöglicht eine zielgerichtete Steuerung des Lernprozesses, der dem Aufbau bzw. der Erweiterung und Verfeinerung des mentalen Modells dienen soll. Der Aspekt der *Individualisierbarkeit* bedarf einer weiteren theoretischen Klärung, die nicht im Rahmen dieser Arbeit erfolgen kann.

Ein theoretisches Grundgerüst mentaler Modelle als Grundlage für die Interaktion zwischen Mensch und Informationsverarbeitungssystem kann die Aspekte verschiedener Usability-Konzepte erklären und begründen. Die meisten Einzelaspekte lassen sich mit Hilfe dieses Grundgerüsts einer differenzierteren Betrachtung unterziehen und werden somit klarer. Das schematische Wissen der Benutzer und daraus gebildete mentale Modelle scheinen als theoretischer Rahmen für Betrachtungen der Usability gut geeignet.

Präzisierung der Fragestellung

Seit der ersten Publikation des TAM durch Davis (1989) hat sich dieses theoretische Modell in zahlreichen Studien prinzipiell bewährt. Die Konzeption mit den beiden Konstrukten PU und PEOU ist zur Vorhersage der *Nutzungsintention* und der tatsächlichen Nutzung moderner Informationsverarbeitungssysteme geeignet. Das Konstrukt PU hat hierbei einen konstant hohen Einfluss auf die *Nutzungsintention* (siehe Kapitel Experimentelle Überprüfungen zum TAM). Der Zusammenhang des Konstrukts PEOU mit der *Nutzungsintention* ist komplexer. In einigen Untersuchungen hatte PEOU einen stärkeren Einfluss auf die *Nutzungsintention* als PU (z. B. Shih, 2004), während in anderen Studien ein Einfluss von PEOU statistisch nicht nachgewiesen werden konnte (z. B. Calisir und Calisir, 2004). In zwei Studien zeigte sich eine zeitliche Veränderung die darin bestand, dass ein zunächst vorhandener Einfluss von PEOU auf die *Nutzungsintention* über die Zeit verschwand.

Davis (1989) hat im Konstrukt PEOU Aspekte der Erlernbarkeit nicht von Aspekten der Benutzbarkeit getrennt. Zudem verschob sich im Prozess der Skalenentwicklung der Schwerpunkt der Skala zunehmend zur Erlernbarkeit, wobei die Benutzbarkeit für das TAM an Bedeutung verlor. Ziel dieser Arbeit ist es, eine Aufspaltung des Konstrukts PEOU in einen Bestandteil Erlernbarkeit und einen weiteren Bestandteil, der die Leichtigkeit der Benutzung unabhängig von Lernprozessen erfasst.

Die Erlernbarkeit der Benutzung eines komplexen technischen Systems verliert für den Benutzer mit zunehmendem Lernfortschritt an Bedeutung. Der Einfluss des Aspekts Erlernbarkeit innerhalb des Konstrukts PEOU ist entsprechend vom Lernfortschritt des Benutzers abhängig und zeitlich nicht konstant. Andere Aspekte

der Bedienung können unabhängig vom Kenntnisstand des Benutzers die Bedienung erschweren. Schlechte Navigationsmöglichkeiten zwischen den Eingabefeldern einer Eingabemaske oder eine nicht zur Aufgabe passende, erzwungene Eingabereihenfolge erschweren die Interaktion mit dem System für alle Benutzer. Wahrscheinlich werden solche Aspekte mit zunehmender Erfahrung mit einem System deutlicher wahrgenommen. Ein unerfahrener Benutzer ist vermutlich vorsichtiger mit negativen Beurteilungen eines Systems, da er nicht ausschließen kann, dass er eine bessere Interaktionsmöglichkeit schlicht noch nicht entdeckt hat. Er würde also im Vergleich zu einem erfahreneren Benutzer eine unbefriedigende Interaktion eher seinen eigenen, mangelhaften Kenntnissen zuschreiben als sie dem Informationsverarbeitungssystem anzulasten. Mit dem Fortschreiten des Lernprozesses wird sich dieses Attributionsmuster verändern, und es wird zunehmend die Bedienoberfläche als verantwortlich für Probleme bei der Interaktion betrachtet werden. Die Bewertung verschiedener Aspekte des Konstrukts PEOU ist somit vermutlich vom Kenntnisstand des Systembenutzers abhängig. Für den Aspekt der Bedienbarkeit wird eine entgegengesetzte Beziehung zwischen der Erfahrung des Benutzers und dem Einfluss auf die Systembewertung erwartet. Während die Bedienbarkeit an sich von unerfahrenen Benutzern noch nicht getrennt vom Lernprozess wahrgenommen werden kann, spielt sie für unerfahrene Benutzer keine bedeutsame Rolle bei der Bewertung eines Informationsverarbeitungssystems. Die *Nutzungsintention* erfahrener Benutzer hingegen wird wesentlich von deren Wahrnehmung der Bedienbarkeit des Systems abhängen.

Entsprechend wird für unerfahrene Benutzer ein starker Einfluss der wahrgenommenen Erlernbarkeit auf die *Nutzungsintention* des TAM erwartet, der mit zunehmender Systemerfahrung in dem Maße schwindet, wie Lernvorgänge bei der Systemnutzung in den Hintergrund treten. Für die wahrgenommene Benutzbarkeit existiert ein umgekehrter Zusammenhang. Am Anfang des Lernprozesses sind Einschätzungen der wahrgenommenen Benutzbarkeit noch unzuverlässig, der Lernvorgang dominiert die Wahrnehmung des Systems. Entsprechend ist für unerfahrene Benutzer noch kein starker Einfluss auf die *Nutzungsintention* zu erwarten. Es soll gezeigt werden, dass mit zunehmender Erfahrung der Wahrnehmungseindruck, der die Bedienbarkeit aus Perspektive eines erfahrenen Benutzers widerspiegelt, wichtiger für die Systemwahrnehmung wird und entsprechend in zunehmendem Maße die *Nutzungsintention* bestimmt. Der Einfluss der beiden neuen, PEOU ablösenden Konstrukte, existiert parallel zum etablierten Einfluss von PU auf die *Nutzungsintention*. Für das Konstrukt PU wird keine Abhängigkeit vom Lernfortschritt des Benutzers erwartet.

Die Aufnahme von Konstrukten, welche von der Lernerfahrung bzw. dem Kenntnisstand der Benutzer abhängen, verändern das Modell zu einem Prozessmodell. Die grundlegende Struktur des neuen Modells ist in Abb. 15 dargestellt. Die enger gefassten Konstrukte Perceived Learnability und Perceived

Ease of Use for Experienced Users ersetzen das TAM-Konstrukt Perceived Ease of Use. Eine Abhängigkeit des Konstrukts Perceived Usefulness von der Erfahrung des Benutzers wird nicht erwartet.

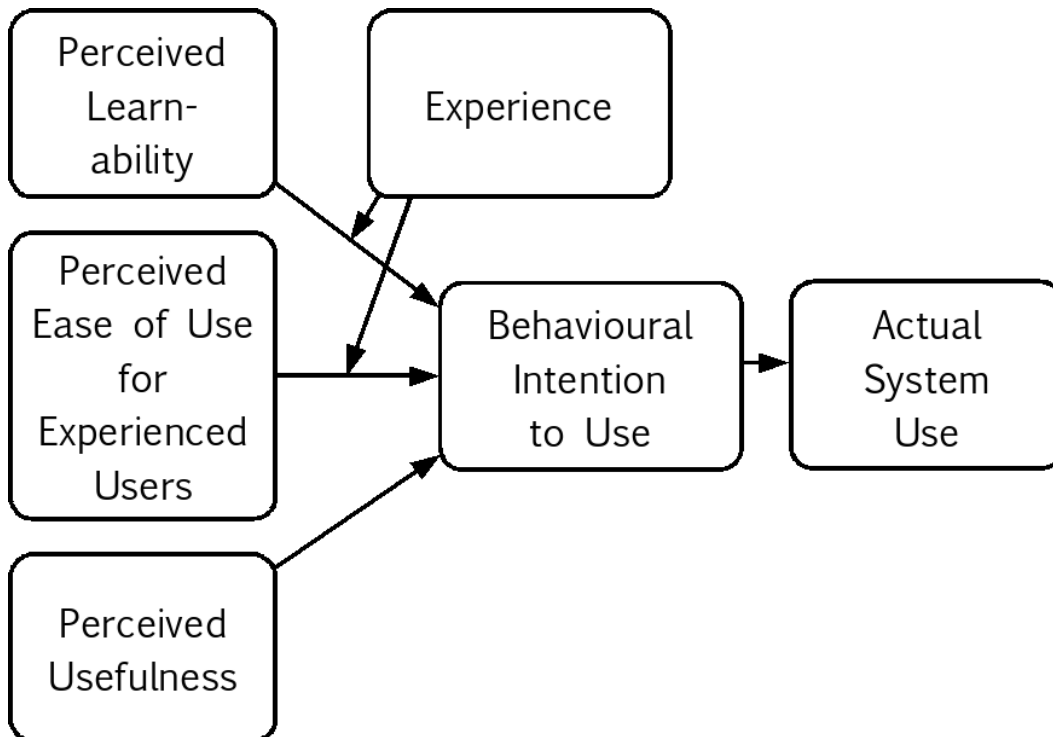


Abb. 15: Prozessmodell der Technologieakzeptanz

In Abhängigkeit von der Erfahrung eines Benutzers mit dem System ändert sich der Einfluss der neuen Konstrukte. Bei unerfahrenen Benutzern spielt die Perceived Learnability eine wichtige Rolle, während Perceived Ease of Use for Experienced Users keinen bedeutsamen Einfluss auf die *Nutzungsintention* hat (Abb. 16). Die graue Darstellung des Konstrukts Perceived Ease of Use for Experienced Users in Abb. 16 stellt den für unerfahrene Benutzer nicht erwarteten Einfluss dieses Konstrukts dar.

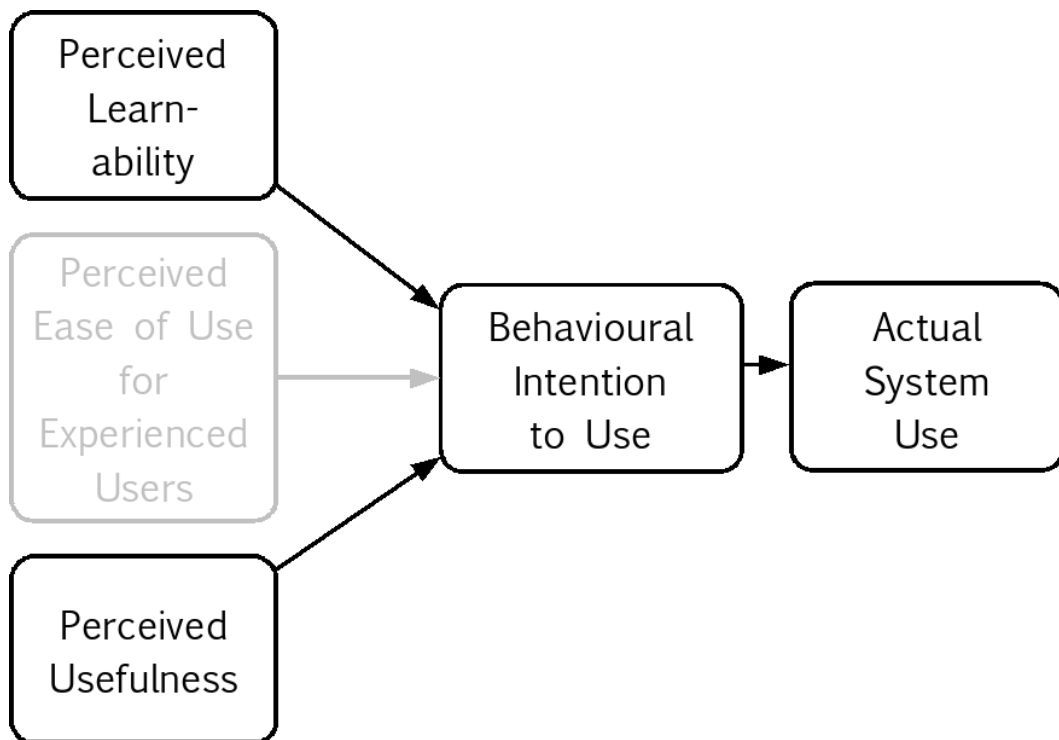


Abb. 16: Modell der Technologieakzeptanz unerfahrener Benutzer (nicht bedeutsame Konstrukte sind grau dargestellt)

Mit zunehmender Erfahrung bezüglich der Bedienung des Systems nimmt der Einfluss der Perceived Learnability ab, während gleichzeitig die Bedeutung von Perceived Ease of Use for Experienced Users zunimmt. Das in Abb. 17 grau dargestellte Konstrukt Perceived Learnability wirkt sich bei erfahrenen Benutzern nicht mehr auf die *Nutzungsintention* aus.

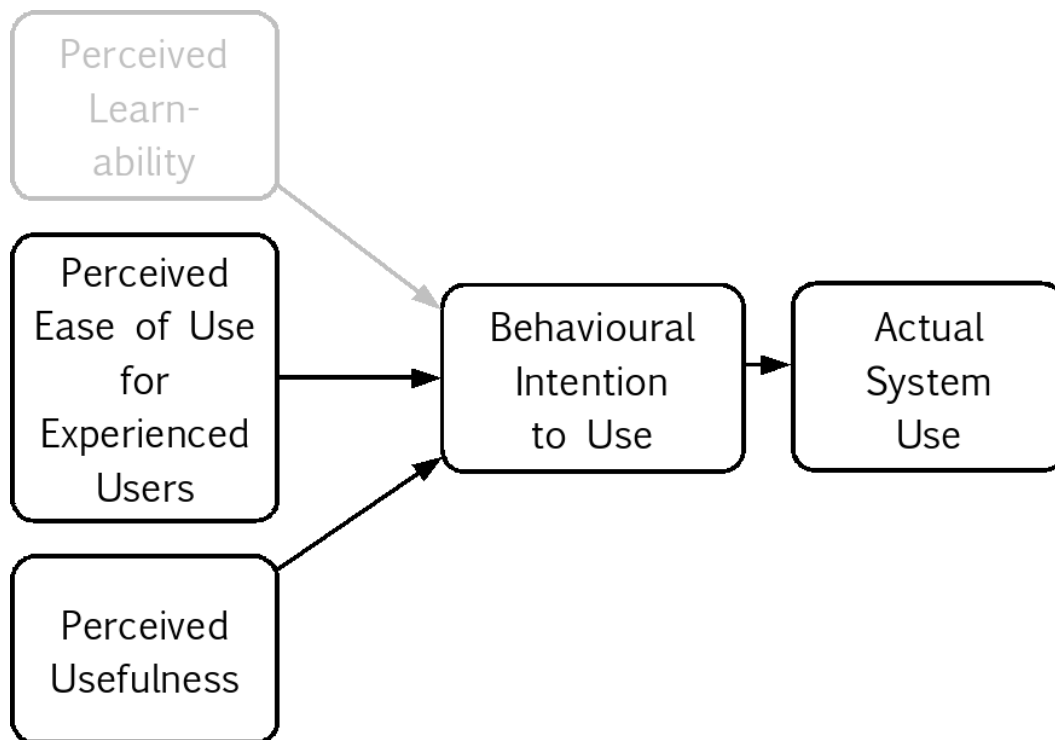


Abb. 17: Modell der Technologieakzeptanz erfahrener Benutzer (nicht bedeutsame Konstrukte sind grau dargestellt)

Das Ziel dieser Arbeit ist es, auf der Grundlage empirischer Daten zu überprüfen, ob ein Prozessmodell mit den beiden Konstrukten Perceived Learnability und Perceived Ease of Use for Experienced Users, deren Einfluss von der Erfahrung des Benutzers abhängt, eine bessere Vorhersage der Nutzungsintention als das TAM ermöglicht. Es wird erwartet, dass die Anpassungsgüte des Prozessmodells im Vergleich zum TAM besser ausfällt. In diesem Fall würde das Prozessmodell ein erweitertes Verständnis der der Beschreibung der Systemakzeptanz zugrundeliegenden kognitiven Prozesse ermöglichen.

Definition der neuen Konstrukte

Die Formulierung der Definition der neuen Konstrukte soll sich eng an die Definitionen des TAM anlehnen. Wie PU und PEOU sollen auch die neuen Konstrukte nicht auf objektive Eigenschaften eines Informationsverarbeitungssystems, sondern deren subjektive Wahrnehmung durch einen Benutzer zielen. Dies ergibt sich aus der alleinigen Bedeutung der Wahrnehmung für die Handlung einer Person. Obwohl die Wahrnehmungseindrücke natürlich nicht abgekoppelt von den Eigenschaften eines Systems betrachtet werden dürfen, sind die tatsächlichen Eigenschaften letztlich irrelevant. Eine – objektiv vorhandene – nicht wahrgenommene Eigenschaft eines Systems hat keinen Einfluss auf die *Nutzungsintention* oder die Handlungen einer Person bezüglich des Systems. Ob ein System beispielsweise die Ausgabe von Daten in einem bestimmten und erwünschten Format beherrscht oder nicht, ist für einen Benutzer nicht primär entscheidend. Kennt er die vorhandene Ausgabemöglichkeit

nicht oder weiß diese nicht zu nutzen, so ist dies für den Benutzer äquivalent mit dem Nichtvorhandensein der entsprechenden Ausgabemöglichkeit. Insofern ist eine Orientierung auf objektive Systemeigenschaften nicht zielführend, da sie nur vermittelt über die Wahrnehmung des Benutzers auf Einstellung und Verhalten wirken. Wie Davis seine Konstrukte als „wahrgenommene“ Systemeigenschaften definiert hat, sollen auch die beiden neuen Konstrukte als Wahrnehmungskorrelate von Systemeigenschaften definiert werden. Für die Bezeichnung der Konstrukte wurde als Sprache Englisch gewählt, damit die Konstrukte besser in den Kontext des TAM passen. Die Definitionen der beiden Konstrukte, die PEOU ersetzen sollen, lauten:

Perceived Learnability (PL):

Der Grad zu dem eine Person glaubt, dass die Benutzung einer bestimmten Software leicht zu erlernen ist.

Perceived Ease of use for experienced users (PEUEU):

Der Grad zu dem eine Person glaubt, dass die Nutzung einer bestimmten Software für den darin geübten Benutzer nicht aufwendig ist.

Empirischer Teil

Der empirische Teil der Arbeit gliedert sich in drei Abschnitte. Der erste Abschnitt beschreibt die grundlegende Entwicklung der beiden neuen Skalen für die Konstrukte Perceived Learnability und Perceived Ease of Use for Experienced Users. Für die neuen Skalen wurden zunächst potentielle Items formuliert. Als Grundlage für die Itemformulierung dienten die aufbereiteten Antworten einer Befragung zu den Konstrukt-Definitionen. Um eine möglichst vollständige Abdeckung des Konstrukts durch die Items der Skala sicherzustellen, wurden die Befragungsergebnisse geclustert und nach ihrer Passung zur Konstrukt-Definition in eine Rangreihenfolge sortiert. Dieses Vorgehen lieferte die Basis für eine erste Itemauswahl.

Im zweiten Abschnitt werden die empirische Testung der Skalen und die daraus resultierenden Verbesserungen der Skalen beschrieben. Die Skalen wurden eingesetzt, um Online-Routenplaner zu bewerten. Anhand der Daten wurden verschiedene statistische Analysen durchgeführt, um zu Erkenntnissen sowohl über einzelne Items wie auch die gesamte Skala zu gelangen. Die Betrachtung der einzelnen Items konzentrierte sich auf die Verteilungsform der Antworten und die Korrelationen des Items mit anderen Items. Für die Gesamtskala wurde Cronbachs Alpha als Kennwert berechnet.

Im letzten Abschnitt werden die neu entwickelten Skalen eingesetzt, um die neuen Konstrukte im Rahmen eines modifizierten TAM zu überprüfen. Hierzu wurden zwei Prototypen von Online-Routenplanern entwickelt, die sich bei gleicher Funktionalität durch ihre Bedienkonzepte unterschieden. Am Anfang und nach einer längeren Benutzungsphase wurden die Routenplaner mit den neuen Skalen bewertet. Die erste Messung erfasste die Beurteilung durch unerfahrene Benutzer, die zweite die Beurteilung durch erfahrene Benutzer. Statistisch wurde die Güte der Modellierung durch das TAM mit den Konstrukten PU und PEOU mit der Güte der Modellierung mit den neuen Konstrukten, also PU, PL und PEUEU verglichen.

Skalenentwicklung

Die empirische Psychologie steht regelmäßig vor dem Problem, Konstrukte erfassen zu müssen, die der direkten Beobachtung nicht zugänglich sind. Um dieses Problem zu lösen, werden Messinstrumente entwickelt, die einer Messung zugängliche Größen erheben, welche in einem engen theoretischen Zusammenhang mit dem eigentlichen Konstrukt stehen. So kann beispielsweise der Erfolg beim Lösen einer mathematischen Aufgabe gemessen werden, um daraus Rückschlüsse auf die Intelligenz zu ziehen. Prinzipiell können bei dieser Art der Messung zwei verschiedene Fehler auftreten:

- die gewählten Kenngrößen stehen nicht im vermuteten Zusammenhang mit dem Konstrukt

•die Messung ist nicht zuverlässig

Bei der Bewertung der Güte von psychologischen Messinstrumenten haben sich für diese beiden Probleme, bzw. deren Gegenteil zwei Begriffe etabliert:

Validität bezeichnet im weitesten Sinne, dass die Skala tatsächlich das misst, was sie ursprünglich messen soll (Anastasi, 1986). Im Falle einer perfekten Validität stehen also sinnvolle Kenngrößen für das Konstrukt zur Verfügung und die Zusammenhänge zwischen diesen Kenngrößen und dem Konstrukt existieren genau so wie vermutet. Außerdem sollten die Kenngrößen nicht von anderen Einflüssen abhängen, da sich sonst eine unerwünschte Abhängigkeit der Messwerte von anderen Konstrukten ergeben würde. Beim Beispiel des Lösungserfolgs einer mathematischen Aufgabe als Messung für Intelligenz wäre eine solche unerwünschte Abhängigkeit des Messwerts zur Vertrautheit mit dem Aufgabentyp gegeben. Eine hohe Vertrautheit mit einem entsprechenden Aufgabentyp erhöht die Lösungswahrscheinlichkeit, hat allerdings keinen Bezug zur Intelligenz der betreffenden Person. Eine hohe Validität einer Skala ist eine grundlegende Voraussetzung für einen sinnvollen Einsatz der Skala. Ein valider Test liefert Messwerte, die in einem engen Zusammenhang mit dem zu messenden Konstrukt stehen.

Reliabilität ist ein Maß für die Zuverlässigkeit eines Messinstruments (Steyer und Eid, 2001, Kapitel 9: Grundbegriffe der klassischen Testtheorie). Ein gutes Messinstrument sollte beim wiederholten Messen derselben Kenngröße möglichst gering voneinander abweichende Messwerte liefern. Außerdem sollten die Messwerte möglichst genau die gemessene Kenngröße abbilden. Eine – theoretisch – einfache Möglichkeit, die Reliabilität eines Messinstruments zu überprüfen, wäre eine häufige Wiederholung der Messung eines Wertes. Ein hoch reliabler Intelligenztest sollte bei wiederholter Messung derselben Person stets das gleiche Ergebnis liefern. Problematisch ist hierbei in der Praxis meist, dass bei psychologischen Messverfahren während der Anwendung Lernprozesse stattfinden, die sich auf das nachfolgende Messergebnis auswirken. Für die Beurteilung der Reliabilität wurden verschiedene statistische Verfahren entwickelt, die das Problem einer Messwiederholung umgehen. Eine hohe Reliabilität ist ebenfalls eine wichtige Voraussetzung für einen sinnvollen Einsatz einer Skala. Ein reliabler Test liefert zuverlässige Messwerte, die frei von zufälligen Einflüssen sind.

Ein häufig eingesetztes Verfahren, um nicht direkt beobachtbare Konstrukte zu messen, sind Fragebogenskalen. Diese setzen sich aus mehreren Aussagen oder Fragen zusammen, die bewertet oder beantwortet werden müssen. Die Aussagen oder Fragen werden als Items (Element, Übers. d. Verf.) bezeichnet. Die Formulierung und Auswahl geeigneter Items bestimmt maßgeblich die Güte einer Fragebogenskala, also deren Validität und Reliabilität. Mit Skalenentwicklung bezeichnet man den Prozess der Formulierung und Auswahl von Items, bis diese die Endfassung einer Skala mit gewünschten Eigenschaften für ein entsprechendes Konstrukt bilden.

Für die Beurteilung der Validität und Reliabilität existieren verschiedene Verfahren und statistische Kenngrößen. Die meisten erlauben eine Beurteilung der Güte einer Skala erst auf der Grundlage konkreter Messwerte der Skala. Entsprechend spielen die beiden Konzepte am Anfang einer Skalenkonstruktion oft eine untergeordnete Rolle, was Anastasi (1986) kritisierte. Sie fordert, die Skalenkonstruktion von klar definierten Konstrukten aus zu beginnen und die Sicherung von Validität und Reliabilität als fortlaufenden Prozess zu betrachten. Diese Ansicht teilen Clark und Watson (1995), die empfehlen, die klare Konstruktdefinition als Ausgangsbasis für die Itemsammlung zu nutzen. Die ursprüngliche Itemsammlung hat nach Clark und Watson eine nicht zu überschätzende Bedeutung für die resultierende Skala. Da der Prozess der Itemauswahl Defizite des Itempools nicht mehr kompensieren kann, würden vom Itempool nicht abgedeckte Aspekte eines Konstrukts von der Skala nicht erfasst. Entsprechend sollte der ursprüngliche Itempool explizit breit angelegt werden, um sicherzustellen, dass alle denkbaren Aspekte des Konstrukts in der Itemsammlung enthalten sind. Clark und Watson empfehlen hierfür, die Perspektive auf das Konstrukt über den eigenen Blickwinkel hinaus zu erweitern. Ein breites Verständnis soll durch Analyse existierender Skalen ähnlicher oder vergleichbarer Konstrukte und offene Befragungen zu den Konstruktdefinitionen erreicht werden. In dieser Arbeit wurden beide Möglichkeiten genutzt, nachfolgend werden die Ergebnisse der Befragung dargestellt.

Befragung zu den Konstruktdefinitionen

Um zu einem entsprechend breiten Blickwinkel auf die Konstrukte PL und PEUEU zu gelangen, wurden 12 Personen frei befragt, welche Eigenschaften eine Software haben muss, um die Definition für PL bzw. PEUEU zu erfüllen. Das Alter der Personen lag zwischen 28 und 59 Jahren, die Geschlechterverteilung war genau ausgewogen. Die genannten Stichpunkte sind in Anhang A vollständig aufgelistet. Es wurden 39 Stichworte zu PL und 27 Stichworte zu PEUEU genannt. Die genannten Stichpunkte waren sich inhaltlich teilweise sehr ähnlich und unterschieden sich im Abstraktionsgrad (z. B. als „Bubble Help“ bezeichnete MouseOver-Effekte und „gutes Hilfsprogramm“). Um zu einer stringenten inhaltlichen Gliederung der beiden Konstrukt-Bereiche zu gelangen, wurden die Stichworte verschiedenen Aspekten zugeordnet. Die Aspekte orientierten sich an den Kriterien der EN-ISO 9241-10, waren aber nicht darauf beschränkt. Manchen Stichworten wurden mehrere Aspekte zugeordnet, die Zuordnungen im Detail sind ebenfalls Anhang A zu entnehmen. Folgende Aspekte konnten bis auf zwei unklare Nennungen („Eingabespeicher“ und „Produktivität“) die genannten Stichworte gut abbilden:

- Übersichtlichkeit
- Selbsterklärend
- Verständlichkeit

- Erwartungskonformität
- Fehlerkorrektur
- Keine überflüssigen Arbeitsschritte notwendig
- Effizienz
- Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- die Interaktion wird vom Benutzer kontrolliert

Der Aspekt *Übersichtlichkeit* wurde Stichworten zu beiden Konstruktdefinitionen zugeordnet. Auch zum Aspekt *Fehlerkorrektur* wurde je ein Stichwort zu den beiden Konstruktdefinitionen genannt. Für die anderen Aspekte ergab sich eine eindeutige Zuordnung zu den beiden Konstruktdefinitionen. Bezüglich *Übersichtlichkeit* war die Häufigkeit der Zuordnung zwischen den Konstrukten stark unterschiedlich: 22 der 39 Stichworte zum Konstrukt PL hatten einen inhaltlichen Bezug zu *Übersichtlichkeit*. Einen entsprechender Bezug zeigten nur drei der 27 Stichworte zum Konstrukt PEUEU. Hier wurde *Übersichtlichkeit* folgenden Nennungen zugeordnet:

- „erweiterte Funktionen leicht zu finden“
- „häufig benötigte Items sind an prominenten Stellen“
- „Bedienelemente, die zur Durchführung einer Aufgabe dienen, gruppieren“

Der erste Punkt zielt bei genauer Betrachtung auf *Erlernbarkeit*, allerdings auf einem gehobenen Niveau. Es geht um die *Erlernbarkeit* einer fortgeschrittenen Funktionalität. Der Nutzer ist zwar prinzipiell erfahren mit dem Programm, aber nicht im Hinblick auf diese Funktionalität. Da es in der Praxis unwahrscheinlich scheint, dass ein Benutzer einen Kenntnisstand erreicht, bei dem weiteres Lernen unmöglich ist, ist Lernen auch für einen erfahrenen Nutzer weiterhin möglich.

In den beiden anderen Stichpunkten finden sich die Aspekte *Übersichtlichkeit* und *Effizienz* gemischt. Eine Gruppierung und Platzierung, die sich an der Struktur der Aufgaben orientiert, erleichtert einerseits die Bildung eines passenden mentalen Modells, erhöht aber auch direkt die Effizienz der Bedienung. Für erfahrene Benutzer wird der Fokus auf der Effizienzsteigerung liegen.

Analyse ähnlicher Skalen

Bei der Konstruktion neuer Skalen empfiehlt es sich, existierende Skalen inhaltlich ähnlicher Konstrukte zu berücksichtigen (Clark und Watson, 1995). Dieses Vorgehen soll dazu beitragen, ein breites Verständnis für das neue Konstrukt zu entwickeln. Im folgenden werden ein Usability-Fragebogen und verschiedene Varianten der PEOU-Skala im Hinblick auf die neuen Konstrukt-Definitionen analysiert.

Der ErgoNorm-Fragebogen ermöglicht, die Erfüllung der EN-ISO 9241-10 subjektiv zu bewerten. Er orientiert sich inhaltlich eng an der in Kapitel Usability in der EN ISO 9241-10 beschriebenen Norm. Der vollständige Fragebogen ist in

Anhang B zu finden. Das ErgoNorm-Testverfahren wurde als Konformitätsprüfung für Software im Hinblick auf die EN ISO 9241-10 entwickelt. Es ist zweistufig aufgebaut und besteht zunächst aus einem Fragebogen zur subjektiven Abfrage der Konformität einer Software mit der Norm. Falls im Fragebogen problematische Aspekte der Software festgestellt werden, erfolgt in einem weiteren Schritt eine sogenannte Erhärtungsprüfung. Dabei wird die gefundene Normabweichung auf ihre Auswirkungen auf *Effektivität* und *Effizienz* untersucht. Diese Analyse wird von Experten durchgeführt.

Der Fragebogen des Prüfverfahrens ist für normale Anwender in einem typischen Nutzungskontext konzipiert. Normale Benutzer tendieren in ihrem gewohnten Nutzungsumfeld dazu, sich bezüglich der Gebrauchstauglichkeit einer eingesetzten Software unkritisch zu äußern (Asher, 1963). Um dieser Tendenz entgegen zu wirken ist der Fragebogen zur Konformitätsprüfung so entwickelt, dass er den Benutzer zu einer kritischen Auseinandersetzung mit der benutzten Software anregt.

Der Fragebogen ist sehr umfangreich, um alle normrelevanten Aspekte einer Software zu erfassen. So beinhaltet alleine der Teil zur Aufgabenangemessenheit 9 Items. Insgesamt besteht der ErgoNorm-Fragebogen aus 28 Fragen, von denen lediglich eine direkt die Lernförderlichkeit adressiert. Dabei orientiert sich der Fragebogen hinsichtlich der verschiedenen Aspekte von Gebrauchstauglichkeit eng an der Norm. Allerdings werden die Definitionen über die Benutzeroberfläche hinaus ausgedehnt und auch auf funktionale Aspekte der Software angewandt. In der Definition für Aufgabenangemessenheit heißt es unter anderem:

„Brauchbar“ bedeutet, dass alle Tätigkeiten, die Sie erledigen müssen, vom Programm unterstützt werden“ (EN ISO 9241-10, 1996)

Diese Definition ist sehr allgemein gehalten und betont die enge Verzahnung von Software, Benutzer und Nutzungsumfeld. Sie geht damit über die Definitionen für die Konstrukte PL bzw. PEUEU deutlich hinaus. Wie hierdurch zu erwarten ist, gehen drei Fragen der Skala über Aspekte der Bedienbarkeit hinaus auf die Funktionalität einer Software ein. Im Bereich *Aufgabenangemessenheit* geht es im ersten Item um die Frage, ob die Funktionalität der Software zur Bearbeitung der Aufgaben ausreicht. Eine Frage des Bereichs *Steuerbarkeit* fragt nach der Reaktionsgeschwindigkeit des Gesamtsystems. Die Reaktionsgeschwindigkeit hängt stark von der Qualität des Programmcodes und der Leistungsfähigkeit der genutzten Hardware ab. Ein Bezug zur Bedienoberfläche ist nicht gegeben. Zur *Fehlertoleranz* wird gefragt, ob das Programm „immer zuverlässig und stabil“ arbeitet. Stabilität und Zuverlässigkeit einer Software sind stark vom Betriebssystem und der Hardware abhängig. Fehler in der Software selbst fallen in den Bereich der Funktionalität und sind unabhängig von der Nutzerschnittstelle.

Die restlichen Fragen des Fragebogens lassen sich folgenden Aspekten zuordnen:

- keine überflüssigen Arbeitsschritte notwendig
- Effizienz
- Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- Selbsterklärend
- Übersichtlichkeit
- Verständlichkeit
- Fehlerkorrektur
- die Interaktion wird vom Benutzer kontrolliert
- Erwartungskonformität

Wegen der großen Anzahl an Fragen, der thematischen Ausdehnung und der sehr spezifischen Formulierungen konnten Fragen des ErgoNorm-Fragebogens nicht direkt übernommen werden. Die inhaltliche Analyse der Items ergab allerdings, dass im Ergo-Norm-Fragebogen keine zusätzlichen Aspekte der Bedienbarkeit einer Software enthalten waren, die bisher keine Berücksichtigung fanden.

Aus der inhaltlichen Nähe der neuen Konstrukte PL und PEUEU zu Davis (1989) Konstrukt PEOU kann auf eine inhaltliche Ähnlichkeit zwischen den neu konstruierenden Skalen und der existierenden PEOU-Skala geschlossen werden. Die neuen Skalen sollen Davis Konstrukt PEOU inhaltlich differenzierter erfassen. Davis Skala für PEOU wurde von verschiedenen Autoren gegenüber der 1989 publizierten Version abgewandelt. Im Rahmen dieser Arbeit wurden 25 Untersuchungen zum TAM analysiert. In 14 davon war die eingesetzte Variante der PEOU-Skala detailliert dokumentiert. Die vollständige Auflistung dieser Skalen ist in Anhang C zu finden. Insgesamt wurden in den ausgewerteten 14 Skalen 16 verschiedene Items zur Erfassung des Konstrukts PEOU eingesetzt. Sie sind im folgenden aufgelistet, wobei alle Items im Präsens formuliert und – soweit möglich – die verschiedenen Produktnamen der in den Studien eingesetzten Software durch SOFTWARE als Platzhalter ersetzt wurden:

- Learning to operate SOFTWARE would be easy for me.
- I would find it easy to get SOFTWARE to do what I want it to do.
- My interaction with SOFTWARE would be clear and understandable.
- I would find SOFTWARE to be flexible to interact with.
- It would be easy for me to become skillful at using SOFTWARE.
- I would find SOFTWARE easy to use.
- I believe that a SOFTWARE is cumbersome to use.
- My using a SOFTWARE requires a lot of mental effort.
- Using a SOFTWARE is often frustrating.
- comfortable using SOFTWARE

- It will be impossible to use SOFTWARE without expert help.
- It takes too long time to learn to use SOFTWARE.
- It is easy to remember how to use SOFTWARE.
- I often become confused when I use the Internet for my shopping activities.
- I can use the Internet in a manner that allows me to obtain the information I want.
- Navigating through the Internet-based learning system was easy for me.
- I think finding what I want via mobile commerce is easy.

Die ersten sechs Items stammen von Davis (1989). Davis postuliert einen engen Zusammenhang zwischen PEOU und der Erlernbarkeit. Davis ursprüngliche Items konnten zu 3 Clustern zusammengefasst werden. Laut Davis entsprechen die Cluster dem physischen Aufwand der Nutzung, dem mentalen Aufwand sowie der Erlernbarkeit (Davis Skalenentwicklung ist im Kapitel Entwicklung der PEOU-Skala im Detail dargestellt).

Die nächsten drei Items wurden von Moore und Benbasat (1991) eingesetzt. Sie waren ursprünglich von Davis (1989) formuliert worden und waren in dessen Itempool enthalten. Im Rahmen von Davis Skalenentwicklung wurden diese Items aussortiert. Moore und Benbasat haben die Skalen ausgehend von Davis ursprünglichen Items neu entwickelt. Dabei kamen sie zu einer gegenüber Davis Version veränderten Itemauswahl. Allerdings bleiben alle von Davis gefundenen Cluster auch in der Auswahl von Moore & Benbasat berücksichtigt.

Das nächste Item („comfortable using SOFTWARE“) wurde in der Untersuchung von Karahanna & Straub (1999) eingesetzt. Die Autoren klären nicht, wie sich für sie der Aspekt „Komfort“ in das Konstrukt PEOU eingliedert. Aus einer zum Vergleich herangezogenen wissenschaftlichen Definition (z. B. Zhang, Helander & Drury, 1996) ist dies nicht ersichtlich.

Moon & Kim (2001) haben für ihre Untersuchung der Technologieakzeptanz des Internets drei neue Items formuliert. Das erste Item fragt ab, ob die Nutzung eines Systems ohne Expertenhilfe möglich scheint. Das zweite bezieht sich auf die zum Erlernen der Nutzung erforderliche Zeit und damit auf die generelle Erlernbarkeit. Inhaltlich zielt das dritte der neuen Items auf die *Einprägsamkeit* (*Memorability*, Nielsen, 1993).

Die Studie von Klopping & McKinney (2004) untersucht die Akzeptanz des Internet als kommerzielle Plattform. Sie verwenden in ihrer Skala ein Item aus Davis ursprünglichen Itempool. Davis (1989) hatte dieses Item nicht in die finale Skala für PEOU übernommen.

In der Studie von Shih (2004) stellt ebenfalls das Internet das untersuchte Software-System dar. Da den Autoren eine wortgetreue Übernahme von Davis (1989) Item „I would find it easy to get CHART-MASTER to do what I want it to do“

für diesen Kontext nicht sinnvoll erschien, formulierten sie das Item um zu „*I can use the Internet in a manner that allows me to obtain the information I want.*“. Obwohl das Item in der Untersuchung von Shih hoch auf dem Faktor PEOU lädt und niedrig auf dem Faktor PU, scheint die Formulierung Aspekte der PU zu beinhalten. Das umformulierte Item wird hier als neues Item betrachtet.

Für die Untersuchung eines internet-basierten Lernsystems haben Saade & Bahli (2005) ein neues Item formuliert, welches die Leichtigkeit der Navigation innerhalb des Systems abfragt.

Das letzte Item stammt von Wu & Wang (2005). Obwohl es im Wortlaut der Formulierung abweicht, scheint es inhaltlich stark dem Item von Shih zu ähneln.

Die Items der PEOU-Skala des TAMs wurden nicht im Hinblick auf eine Differenzierung zwischen Erlernbarkeit und Nutzbarkeit durch erfahrene Benutzer formuliert. Entsprechend fällt es für die meisten Items schwer, eine solche Zuordnung vorzunehmen. Ausnahmen sind zwei Items von Davis Originalskala („*Learning to operate CHART-MASTER would be easy for me*“ und „*It would be easy for me to become skillful at using CHART-MASTER*“), die sich eindeutig auf *Erlernbarkeit* fokussieren. Diese beiden Items wurden in den Itempool für die neuen Skalen übernommen. Für die anderen Items war keine klare Zuordnung im Hinblick auf die neuen, differenzierteren Konstrukte möglich. Inhaltlich sind diese Items durch die formulierten Aspekte abgedeckt.

Itempool

Im folgenden sind die gefunden Aspekte zusammenfassend aufgelistet. Dabei wurde zwischen einer Zuordnung zum Konstrukt PL bzw. PEUEU unterschieden:

Konstrukt PL

- Übersichtlichkeit
- Selbsterklärend
- Verständlichkeit
- Erwartungskonformität
- Fehlerkorrektur

Konstrukt PEUEU

- Übersichtlichkeit
- Keine überflüssigen Arbeitsschritte notwendig
- Effizienz
- Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- die Interaktion wird vom Benutzer kontrolliert
- Fehlerkorrektur

Übersichtlichkeit ist für beide Konstrukte relevant, wird allerdings deutlich häufiger im Zusammenhang mit dem Konstrukt PL genannt (20 Nennungen vs. 3 Nennungen). Clark und Watson (1995) fordern die Formulierung eines möglichst breiten Itempools, um eine vollständige Abdeckung der Konstruktdefinition sicherzustellen. Um diesen Anforderungen zu genügen, wurden für die beiden Skalen jeweils 17 Items formuliert. Zwei der Items für die PL-Skala sind, wie bereits erwähnt, Übersetzungen von Items, die Davis (1989) formuliert hat. Die gesamten Roh-Items sind zusammen mit den zugehörigen Konstruktdefinitionen nachfolgend aufgelistet:

Skala für Perceived Learnability

Definition: Der Grad zu dem eine Person glaubt, dass die Benutzung einer bestimmten Software leicht zu erlernen ist.

- Die Bedienung von SOFTWARE ist für mich leicht zu lernen.
- Es ist leicht für mich, ein geschickter Nutzer von SOFTWARE zu werden.
- SOFTWARE verhält sich so, wie ich es von anderen Programmen kenne.
- Es ist leicht für mich zu erkennen, welche Funktionen SOFTWARE bereitstellt.
- SOFTWARE wirkt für mich sehr übersichtlich.
- Wie SOFTWARE funktioniert, ist für mich leicht zu erkennen.
- Die Funktion von allen Bedienelementen ist offensichtlich.
- Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen.
- Ich brauche keine besonderen Kenntnisse für die Nutzung von SOFTWARE.
- Ich kann mich leicht in die Bedienung von SOFTWARE einarbeiten.
- Die Oberfläche von SOFTWARE wirkt auf mich gut strukturiert.
- Die Struktur von SOFTWARE ist für mich leicht zu erkennen.
- Ich komme mit SOFTWARE auf Anhieb gut zurecht.
- Die Funktionen von SOFTWARE sind für mich leicht zu erkennen.
- Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE.
- Die Bedienung von SOFTWARE ist für mich leicht verständlich.
- Ich kann die Bedienung von SOFTWARE gut durch Ausprobieren erlernen.

Skala für Perceived Ease of Use for Experienced Users

Definition: Der Grad zu dem eine Person glaubt, dass die Nutzung einer bestimmten Software für den darin geübten Benutzer nicht aufwändig ist.

- Mit SOFTWARE kann ich effizient arbeiten.
- Wichtige Funktionen sind für mich in SOFTWARE schnell erreichbar.
- SOFTWARE passt zu meinem Arbeitsstil.

- Die Bedienoberfläche von SOFTWARE ist für mich optimal gestaltet.
- Die Ausgaben von SOFTWARE sind genau so, wie ich es mir wünsche.
- Die Eingabe von Informationen in SOFTWARE ist nicht aufwendig.
- Ich fühle mich durch SOFTWARE nicht bevormundet.
- Die Bedienung von SOFTWARE entspricht genau meinen Anforderungen.
- Ich muss keine unnötigen Arbeitsschritte machen, wenn ich SOFTWARE bediene.
- Mit Übung kann ich SOFTWARE fehlerfrei nutzen.
- Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen.
- Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar.
- Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben.
- Ich werde durch die Oberfläche von SOFTWARE nicht eingeschränkt.
- Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen.
- Ich bekomme von SOFTWARE keine für mich überflüssigen Hilfestellungen.
- Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren.

Die Itemselektion erfolgte in zwei Stufen. In der ersten wurden die Items durch eine kleinere Gruppe von Personen bezüglich ihrer inhaltlichen Ähnlichkeit und ihrer Übereinstimmung mit der Konstruktdefinition sortiert. In einem zweiten Schritt wurde die hieraus gebildete Itemauswahl in einer größeren Fragebogenstudie zu Internet-Routenplanern eingesetzt. Die in dieser Studie gewonnenen Daten ermöglichten eine Beurteilung der Items auf der Grundlage verschiedener statistischer Parameter. Zunächst wird die Sortieraufgabe und der hieraus folgende Selektionsprozess beschrieben.

Erste Itemselektion (Sortieraufgabe)

Die erste Itemauswahl wurde auf der Grundlage verschiedener Sortieraufgaben durchgeführt. Die gesamten Items für die beiden neuen Skalen wurden von 16 Probanden insgesamt drei mal nach unterschiedlichen Vorgaben sortiert. Dieses Vorgehen führte zu einem tieferen Verständnis der inneren Struktur der Konstrukte und erlaubte die Bewertung der Güte der Items im Hinblick auf die Skalendefinition. Das Verfahren und die Ergebnisse werden nachfolgend beschrieben.

Methode

Sherif und Sherif (1967) beschrieben ein als „own category procedure“ (Methode [zum Auffinden] eigener Kategorien [einer Versuchsperson], Übers. d. Verf.) bezeichnetes Verfahren zur inhaltlichen Gliederung mehrerer Items. Dieses Verfahren wurde in abgewandelter Form von Davis für die Gliederung seiner Items eingesetzt. Auch in dieser Arbeit wurde ein Verfahren zur Gliederung genutzt, welches auf der Methode von Sherif und Sherif aufbaut. Die Items werden hierfür einzeln auf Karten

gedruckt. Die Karten werden von den Probanden nach ihrer Ähnlichkeit auf Stapel sortiert. Die Anzahl zu bildender Stapel wurde von Sherif und Sherif nicht vorgegeben. Anschließend sollen die Probanden die Stapel im Hinblick auf eine vorgegebene Aussage bewerten. Übertragen auf die Einstellungen zu einem Informationsverarbeitungssystem können beispielsweise verschiedene Aussagen über eine Software als Items dienen. Die Probanden würden dann zunächst Stapel bilden, auf den ähnliche Items zusammengefasst würden. In einem zweiten Schritt würde eine Sortierung der Stapel nach – je nach Fragestellung der Untersuchung – unterschiedlichen Vorgaben erfolgen. Die Vorgabe könnte beispielsweise sein, wie wichtig das Zutreffen der Aussagen der jeweiligen Stapel für die Eignung einer Software für den Büroalltag der Probanden sind. Oder die Stapel könnten nach der Wichtigkeit für die Benutzbarkeit der Software sortiert werden. Die Gruppierung in Stapel und deren abschließende Sortierung in eine Rangreihe ermöglicht einen Einblick in die Gliederung und Bewertung der entsprechenden Eigenschaften durch die Probanden.

Davis (1989) modifizierte das von Sherif und Sherif beschriebene Verfahren und setzte es ein, um die Items für die PU- bzw. PEOU-Skala zu gliedern (siehe Kapitel Entwicklung der PEOU-Skala). Abweichend von Sherif und Sherifs Methode gab Davis eine Ober- und Untergrenze für die Anzahl der zu bildenden Stapel vor. Eine Begründung für diese Einschränkung gab Davis nicht an, vermutlich waren praktische Überlegungen hinsichtlich der Auswertbarkeit der Ergebnisse hier ausschlaggebend. Als zweite Abweichung ließ Davis nicht die gebildeten Stapel in eine Rangreihenfolge sortieren, sondern die einzelnen Items und zwar unabhängig von ihrer Einsortierung in die Stapel.

Das von Sherif und Sherif (1967) entwickelte Verfahren zur inhaltlichen Gliederung einer Itemsammlung erscheint für einen ersten Schritt zur Selektion geeigneter Items gut geeignet. Im Rahmen dieser Arbeit wurde versucht, die Vorteile des ursprünglichen Verfahrens von Sherif und Sherif mit sinnvollen Modifikationen durch Davis (1989) zum Einsatz zu bringen. Bestehend an Sherif und Sherifs Ansatz ist, die Anzahl der Kategorien offen zu lassen. Dieses Verfahren ordnet sich vollständig der Kategorisierung durch die Probanden unter und macht hier ursprünglich keinerlei Vorgabe. Die Einführung von Ober- und Untergrenzen für die zu bildende Kategorienganzheit durch Davis stellt entsprechend keine Verbesserung des Verfahrens dar und wurde für diese Arbeit nicht übernommen. Anders verhält es sich mit der zweiten Veränderung der Methode durch Davis, die Rangreihenfolge nicht auf der Ebene der gebildeten Kategorien, sondern der einzelnen Items bilden zu lassen. Dieses Vorgehen ist für das angestrebte Ziel – der Auswahl einzelner Items – besser geeignet als eine entsprechende Sortierung der Stapel. Zumal es sich für die Konstruktion einer Skala ohnehin verbietet, eine zum Konstrukt gehörige Kategorie (also einen Stapel) komplett zu entfernen. Im Rahmen der Itemselektion interessiert es nicht, welche Kategorie mit den zugehörigen Items am unwichtigsten

ist, um anschließend die entsprechenden Items komplett zu entfernen. Vielmehr sollen innerhalb jeder Kategorie jene Items identifiziert werden, die am besten zur Konstruktdefinition der Skala passen. Es ist also hinsichtlich der Itemselektion vorteilhaft, die einzelnen Items in eine Rangreihenfolge sortieren zu lassen. Zusätzlich wurde eine weitere Veränderung vorgenommen. Sherif und Sherif äußern sich nicht hinsichtlich der Anwendung der *own category procedure* für verschiedene Konstrukte. Davis legte den Probanden die Items für die PU- und die PEOU-Skala getrennt vor. Damit wurde eine erste Möglichkeit zur inhaltlichen Überprüfung, wie stark die beiden Konstrukte sich unerwünscht überlappen und ob die angenommenen Zuordnungen der Items zu den Konstrukten bestätigt werden, nicht genutzt. In dieser Arbeit wurden die Items für die Sortieraufgaben vermischt. Dieses Vorgehen ermöglichte eine erste Analyse, ob die vorgenommene Aufteilung des Konstrukts PEOU in die beiden neuen Konstrukte PL und PEUEU sinnvoll war.

Erhebung

Die drei verschiedenen Sortierungen der Items wurden von 16 Personen vorgenommen. Das Alter der Personen lag zwischen 28 und 39 Jahren, neun Personen waren weiblich und sieben männlich. Jede Person sortierte die insgesamt 34 Items (je 17 Items für die PL- bzw. PEUEU-Skala) drei mal nach verschiedenen Vorgaben. Hierfür waren die Items einzeln auf laminierte Karten gedruckt. Die für die Auswertung der Items notwendige Kodierung der Karten, die eine Zuordnung der Items zu den Konstrukten ermöglichte, war auf der Kartenrückseite aufgedruckt und konnte die Sortierung entsprechend nicht beeinflussen. Die Probanden konnten also nicht erkennen, welcher Konstrukt-Definition ein Item vorab zugeordnet war. Im ersten Durchlauf sollte eine beliebige Anzahl Stapel so gebildet werden, das alle inhaltlich ähnlichen Items auf einem Stapel zusammengefasst wurden. Danach sollten je zwei Rangreihenfolgen nach den folgenden Vorgaben gebildet werden:

- paßt die Itemaussage zu „dem Grad, zu dem Sie glauben, dass die Benutzung einer bestimmten Software leicht zu erlernen ist“
- paßt Itemaussage zu „dem Grad, zu dem Sie glauben, dass die Nutzung einer bestimmten Software für den darin geübten Benutzer nicht aufwändig ist“

Die Instruktionen wurden schriftlich gegeben und – falls nötig – mündlich erläutert. Instruktionen sind in Anhang D enthalten.

Auswertung der Daten

Die Auswertung der Daten der verschiedenen Sortieraufgaben erfolgte in mehreren Schritten. Zunächst wurden die hinsichtlich der zwei Konstruktdefinitionen gebildeten Rangreihenfolgen analysiert. Die von den Probanden zugewiesenen Positionen in der Rangreihenfolge wurden linear mit Werten von 1 bis 34 bewertet. Das am wenigsten zur jeweiligen Konstruktdefinition passende Item erhielt für den letzten Platz den Wert 34 zugewiesen, das am besten passende den Wert 1. Durch

diese Umsetzung der Rangposition in numerische Werte hat quasi jeder Proband jedem Item einen Wert zwischen 1 und 34 zugeordnet, wobei niedrige Werte eine gute Passung zur Konstruktdefinition ausdrücken und hohe Werte eine schlechte Passung. Diese Werte wurden aufsummiert, um zu einer Gesamtreihenfolge über alle Probanden zu gelangen. Für die zwei Konstruktdefinitionen ergaben sich entsprechend zwei Rangreihenfolgen.

Die Auswertung der Stapelbildung erfolgte über eine Matrix, in der alle möglichen Itempaarungen abgebildet werden konnten. Die Matrix hatte hierfür genauso viele Zeilen und Spalten, wie Items vorhanden waren. Zeilen und Spalten wurden mit den Itembezeichnungen beschriftet. Diese Matrix war symmetrisch entlang der Diagonalen, weshalb nur eine Hälfte mit Werten gefüllt wurde. Pro Versuchsperson wurde in diese Matrix der Wert „1“ eingetragen, wenn zwei Items von dieser Person auf denselben Stapel sortiert worden waren, oder der Wert „0“, falls die Items auf verschiedene Stapel gruppiert worden waren. Die Aufsummierung dieser Werte über alle Versuchspersonen führte zu einer Ähnlichkeitsmatrix, in der jede mögliche Kombination aus zwei Items einen Wert zwischen null und sechzehn zugewiesen hatte. Hohe Werte zeigen eine starke inhaltliche Ähnlichkeit der entsprechenden Items an, während niedrige Werte auf inhaltliche Unterschiede schließen lassen. Die Ähnlichkeitsmatrix für die Items ist in Anhang E dargestellt.

Die Ähnlichkeitsmatrix konnte durch eine multidimensionale Skalierung (MDS) in verschiedener Hinsicht analysiert werden. Die MDS ist ein statistisches Verfahren, bei dem Objekte hinsichtlich ihrer Ähnlichkeit zueinander in einem möglichst niedrigdimensionalen Raum dargestellt werden (Borg & Groenen, 2005). Es ist ein Verfahren – ähnlich wie die Faktorenanalyse – zur Datenreduktion. Im Unterschied zur Faktorenanalyse begnügt sich die MDS mit einem ordinalen Messniveau der Skalen. Eine MDS versucht Objekte in einer vorgegebenen Anzahl von Dimensionen so anzuordnen, dass die Abstände zwischen den angeordneten Objekten möglichst genau den vorgegebenen (meist gemessenen) Abständen bzw. Ähnlichkeitsmaßen entsprechen. Zunächst wurde eine MDS mit nur einer Dimension gerechnet. Die Items wurden dadurch entlang einer Achse entsprechend ihrer Ähnlichkeitswerte verteilt. Ähnliche Items wurden nahe zueinander angeordnet und unähnliche Items weit voneinander entfernt. Durch die eindimensionale MDS wurde also eine dritte Rangreihenfolge der Items erzeugt. Diese ermöglichte eine zusätzliche Betrachtung, ob die Items entsprechend der beiden unterschiedlichen Konstrukte unterschieden wurden.

Um zu einer Clusterung der Items innerhalb eines Konstrukts zu gelangen, wurden für die konstruktspezifischen Items mehrere MDS mit einer variablen Anzahl von Dimensionen gerechnet. Ein wichtiger Schritt bei der MDS ist die Festlegung der notwendigen Anzahl von Dimensionen für die Beschreibung der Objekte. Das Ziel der Datenreduktion gibt vor, die Anzahl an Dimensionen möglichst gering zu wählen. Dabei werden die Abweichungen zwischen den abgebildeten Distanzen und den

ursprünglichen Ähnlichkeitsmaßen mit einer abnehmenden Anzahl an Abbildungsdimensionen zunehmen. Ein Maß für die Güte der Abbildung ist der sog. Stress (Borg & Groenen, 2005). Der Stress errechnet sich als Summe der quadrierten Abweichungen der Distanzen der Modellierung von den vorgegebenen Distanzwerten geteilt durch die Summe der vorgegebenen Distanzwerte (Venables und Ripley, 2002). Zur Beurteilung der Güte der Modellierungen und zur Auswahl einer geeigneten Anzahl an Dimensionen wurden sog. Scree-Plots angefertigt. Dabei werden die Stress-Maße für Modellierungen mit zunehmender Anzahl an Dimensionen über die Anzahl der Dimensionen graphisch dargestellt. Dieses Vorgehen wurde getrennt für die Items der beiden Konstrukte durchgeführt.

Die statistischen Berechnung wurden mit der freien Software R in der Version 2.2.1 durchgeführt, die über cran.r-project.org verteilt wird. R ist eine freie Implementierung der Statistik-Umgebung S bzw. S-Plus der Insightful Corporation und weitgehend zu dieser kompatibel (Dalgaard, 2002). Zusätzlich installierte Pakete, die in der Grundversion nicht vorhandene Funktionen verfügbar machen, waren MASS (Venables und Ripley, 2002) und NLME (Pinheiro und Bates, 2004).

Ergebnisse

Je nach Sortierung hinsichtlich der Passung zur Konstruktdefinition für PL oder PEUEU ergaben sich folgende Rangreihenfolgen (die Werte in Klammern geben die aufsummierten Rangplatzbewertungen aller Probanden an):

Passung zur PL-Konstruktdefinition	Passung zur PEUEU-Konstruktdefinition
L01 (112)	E01 (96)
L06 (139)	E02 (111)
L16 (144)	E15 (112)
L10 (151)	E11 (112)
L17 (161)	E09 (121)
L03 (162)	E13 (172)
L08 (178)	E08 (172)
L13 (181)	E17 (174)
L09 (201)	E03 (184)
L07 (204)	E06 (195)
L02 (207)	E04 (199)
L05 (211)	E05 (218)
L14 (213)	E14 (238)
L04 (225)	E16 (239)
L12 (255)	E12 (241)
L11 (257)	E07 (296)
E11 (273)	L14 (304)
E02 (276)	L15 (323)
E17 (312)	L05 (328)

E04	(315)	L11	(331)
L15	(318)	L07	(334)
E10	(322)	L06	(341)
E03	(349)	E10	(348)
E16	(356)	L16	(351)
E09	(370)	L04	(361)
E12	(370)	L12	(366)
E08	(386)	L03	(371)
E06	(388)	L08	(379)
E13	(394)	L13	(385)
E15	(396)	L02	(396)
E01	(412)	L10	(424)
E14	(417)	L17	(425)
E07	(421)	L09	(427)
E05	(444)	L01	(446)

Die Items für die PL-Skala sind mit dem Buchstaben *L* und einer zweistelligen Ziffer kodiert, die Items für die PEUEU-Skala beginnen mit dem Buchstaben *E*. Die vollständige Auflistung der Rohitems mit Kodierung befindet sich in Anhang F.

Ein geringer Wert mit einem entsprechenden oberen Platz in der Rangreihe zeigt an, dass viele Probanden das Item als für das Zutreffen der Konstruktdefinition wichtig eingestuft haben. Umgekehrt zeigt ein unterer Platz in der Auflistung (mit einem hohen Wert) an, dass viele Probanden die Aussage des Items als für das Zutreffen der Konstruktdefinition unwichtig betrachtet haben. Für die Rangreihenfolge im Hinblick auf die Konstruktdefinition von PL wurde eine Einstufung der PL-Items als relevant und der PEUEU-Items als wenig relevant erwartet.

Das Ergebnis der Sortieraufgabe bestätigte diese Erwartung im Wesentlichen. Die PL-Items (mit dem Buchstaben *L* kodiert) befinden sich in der Rangreihe oben, die PEUEU-Items (mit dem Buchstaben *E* kodiert) unten. Lediglich das Item L15 (Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE) weicht von diesem Muster ab. Es wurde mit einem Rangplatz 21 von 34 als für das Zutreffen der Konstruktdefinition als wenig wichtig eingestuft.

Für die zweite Sortieraufgabe, in der die Rangreihenfolge bezüglich der Konstruktdefinition von PEUEU gebildet wurde, wurde eine umgekehrte Anordnung der Items erwartet. Die PL-Items sollten hier als unwichtig am unteren Ende der Rangreihe stehen, während die PEUEU-Items oben zu finden sein sollten. Auch für diese Sortierung wurden die Erwartungen grundlegend bestätigt. Bis auf Item E10 (Mit Übung kann ich SOFTWARE fehlerfrei nutzen) waren alle PEUEU-Items (mit dem Buchstaben *E* kodiert) oben in der Rangreihenfolge zu finden, die PL-Items im unteren Teil. Das Item E10 war auf Rang 23 von 34 möglichen Rängen sortiert worden.

Eine weitere Überprüfung der Aufteilung der Items in zwei Blöcke, und damit auch eine entsprechende Bestätigung der inhaltlichen Trennung des Konstrukts PEOU in die Konstrukte PL und PEUEU, konnte auf der Grundlage der von den Probanden gebildeten Stapel durchgeführt werden. Hierzu wurde eine eindimensionale MDS der Ähnlichkeitsmatrix durchgeführt. Eine eindimensionale MDS ordnet alle Objekte entlang einer Achse an, im Ergebnis wurden die Items in einer Abfolge angeordnet. Die Reihenfolge der Items, die sich aus der MDS ergab, ist in Abb. 18 dargestellt.

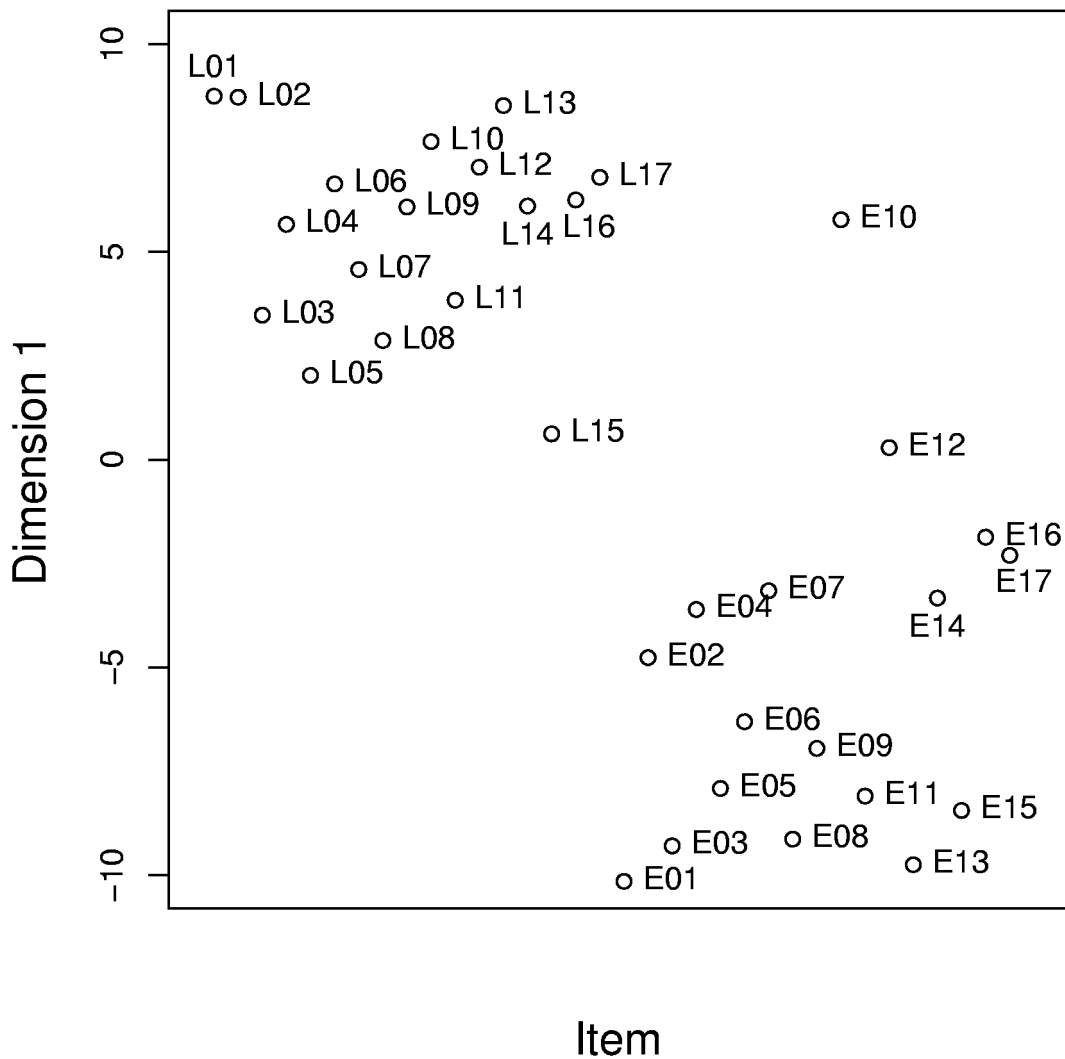


Abb. 18: Eindimensionale MDS der Ähnlichkeit der Items

In der Abbildung sind die Items entlang der Abszisse der Reihen nach angeordnet (zuerst die L-Items, dann die E-Items in aufsteigender Sortierung). Die vertikale Symmetrieachse unterteilt die Abbildung entsprechend in PL- bzw. PEUEU-Items. Die Anordnung entlang der Ordinate stellt das Ergebnis der MDS dar. Items, die sich ähnlich sind – also von vielen Probanden gemeinsam auf einen Stapel sortiert worden waren – sind entlang der Ordinate nahe beieinander angeordnet. Items, die selten gemeinsam auf einen Stapel sortiert worden waren, sind entlang

der Ordinate weit voneinander positioniert.

Die Einheit der Ordinate kann als Ähnlichkeit bzw. Unähnlichkeit der Items interpretiert werden. Die Distanz zweier Items entlang der Ordinate bildet ab, wie unähnlich sich diese Items sind bzw. von wie vielen Probanden diese Items nicht auf einen gemeinsamen Stapel sortiert wurden. Da die Abbildung nicht einzelne Itempaarungen darstellt, sondern eine integrierte Anordnung aller denkbaren Paarungen, entsprechen die Abstände zwischen zwei beliebigen Items nicht genau der Anzahl von Probanden, die dieses Paar nicht zusammen auf einen Stapel sortiert hat. Beispielsweise beträgt der Abstand zwischen den Items L01 und E10 in der Abbildung ungefähr 20, obwohl maximal 16 Probanden die Items nicht auf denselben Stapel sortieren konnten. Die MDS ordnet alle Items genau so an, dass die Distanzen zwischen zwei Items möglichst genau abgebildet wird.

Die Anordnung der Items in Abb. 18 entspricht im Wesentlichen den Erwartungen. Sowohl die PL-Items (mit dem Buchstaben L kodiert) wie auch die PEUEU-Items (mit dem Buchstaben E kodiert) bilden eine zusammenhängende Punktwolke. Lediglich das Item E10 weicht von diesem Muster deutlich ab. Dieses Item wurde also bei der Sortieraufgabe von vielen Probanden zusammen mit PL-Items gruppiert, obwohl es als PEUEU-Item konzeptioniert war. Die Items L15 (Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE) und E12 (Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar) sind nahe der horizontalen Symmetrieachse der Abbildung angeordnet. Das bedeutet, diese Items wurden inhaltlich ähnlich oft mit PL-Items wie auch mit PEUEU-Items zusammen sortiert. Diese Items lassen sich entsprechend schlecht eindeutig zu einem der beiden Konstrukte zuordnen.

Insgesamt lässt sich für die eindimensionale MDS zusammenfassen, dass sich in der Anordnung der Items die Trennung der beiden Konstrukte PL und PEUEU wiederfindet. Die meisten Items sind sich innerhalb eines Konstrukts ähnlich und zu den Items des anderen Konstrukts unähnlich. Das Item E10 (Mit Übung kann ich SOFTWARE fehlerfrei nutzen) verletzt diese Regel eindeutig, während die Items L15 und E12 eine klare Zuordnung vermissen lassen.

Zusammenfassend zeigt sich die angestrebte Differenzierung in die beiden unterschiedlichen Konstrukte in jeder der drei Sortierungen. Die Ergebnisse der Sortieraufgaben sind somit eine erste Bestätigung des Versuchs, das Konstrukt PEOU aufzuspalten. Die inhaltliche Trennung in das Konstrukt PL und PEUEU wurde von den 16 Probanden bei den Sortierungen der zusammengefassten Items wieder hergestellt. Hierbei fielen zwei Items unerwünscht auf: Item L15 wurde in einer der drei Reihenfolgen nicht entsprechend den Erwartungen einsortiert und Item E10 in zwei der drei Reihenfolgen. Die offensichtlich falsche Eingruppierung des Items E10 als PL-Item in der MDS gab den Ausschlag für dessen Ausschluss aus dem Itempool. In allen weiteren Auswertungen war dieses Item nicht mehr enthalten.

Ein weiteres Item musste aufgrund aufgetretener inhaltlicher Probleme aus

dem Itempool entfernt werden. Während der Sortieraufgabe wurde wiederholt nach der Bedeutung des Begriffs „Ausgaben“ in der Formulierung des Items E05 (Die Ausgaben von SOFTWARE sind genau so wie ich es mir wünsche) gefragt. Der Begriff kann im Kontext des Items unterschiedlich interpretiert werden. Zusätzlich zur angestrebten Bedeutung im Sinne von Arbeitsergebnis bzw. der Formatierung der Arbeitsergebnisse, kann der Begriff auf die Version einer Software (beispielsweise OpenOffice 1.1.3 oder OpenOffice 2.2.1) bezogen werden. Da diese unerwünschte Interpretation häufig auftrat (6 Probanden fragten während der Sortieraufgaben bezüglich der Bedeutung nach), wurde das Item als für die Skala unbrauchbar aussortiert. Auf eine Umformulierung des Items wurde verzichtet, da eine erneute Überprüfung anhand einer Replikation der Sortieraufgaben nicht geleistet werden konnte. Eine umformulierte Version des Items wäre entsprechend ohne grundlegende Überprüfung im Itempool verblieben, was nicht optimal erschien.

Um Aufschluss über eine inhaltliche Gliederung der Items innerhalb des Itempools pro Konstrukt zu erhalten, wurden getrennte MDS sowohl für die PL-Items wie auch für die PEUEU-Items durchgeführt. Hierbei wurde der gesamte Pool an PL-Items ausgewertet, während aus den PEUEU-Items die Items E05 (wegen Verständnisproblemen der Itemformulierung) und E10 (wegen zu großer Ähnlichkeit zu PL-Items) ausgeschlossen worden waren.

Ein generelles Problem datenreduzierender Verfahren wie der MDS oder der Faktorenanalyse ist die Auswahl einer Abbildung, die zwar zu einer erheblichen Reduktion der Datenmenge führt, dabei aber noch eine ausreichend genaue Abbildung der Daten leistet. Je größer die Anzahl an Dimensionen, die eine MDS zur Abbildung der Daten nutzt, desto genauer kann sie die Daten abbilden. Allerdings fällt die Reduktion der Daten umso geringer aus, je höher die Anzahl an Dimensionen gewählt wird. Es gilt also abzuwägen zwischen einer möglichst genauen Abbildung der Daten einerseits und einer möglichst einfachen Abbildung andererseits (Backhaus, Erichson, Plinke und Weiber, 2006, Kapitel 10). Ein etabliertes Vorgehen ist hierbei anhand eines Gütekriteriums die einfachste Abbildung, bei der eine gewünschte Abbildungsgüte noch erreicht wird, auszuwählen. Für die MDS ist ein geeignetes Kriterium zur Beurteilung der Abbildungsgüte der sogenannte Stress (Venables und Ripley, 2002, S. 306-308).

Graphisch kann der Zusammenhang zwischen der Komplexität einer MDS und der Abbildungsgüte der Daten in einem sogenannten Scree-Plot dargestellt werden. In einem Scree-Plot wird auf der Abszisse ein Index für die Komplexität einer Abbildung abgetragen und auf der Ordinate wird das zugehörige Maß für die Abbildungsgüte dargestellt. Abb. 19 stellt einen solchen Scree-Plot für die PL-Items dar.

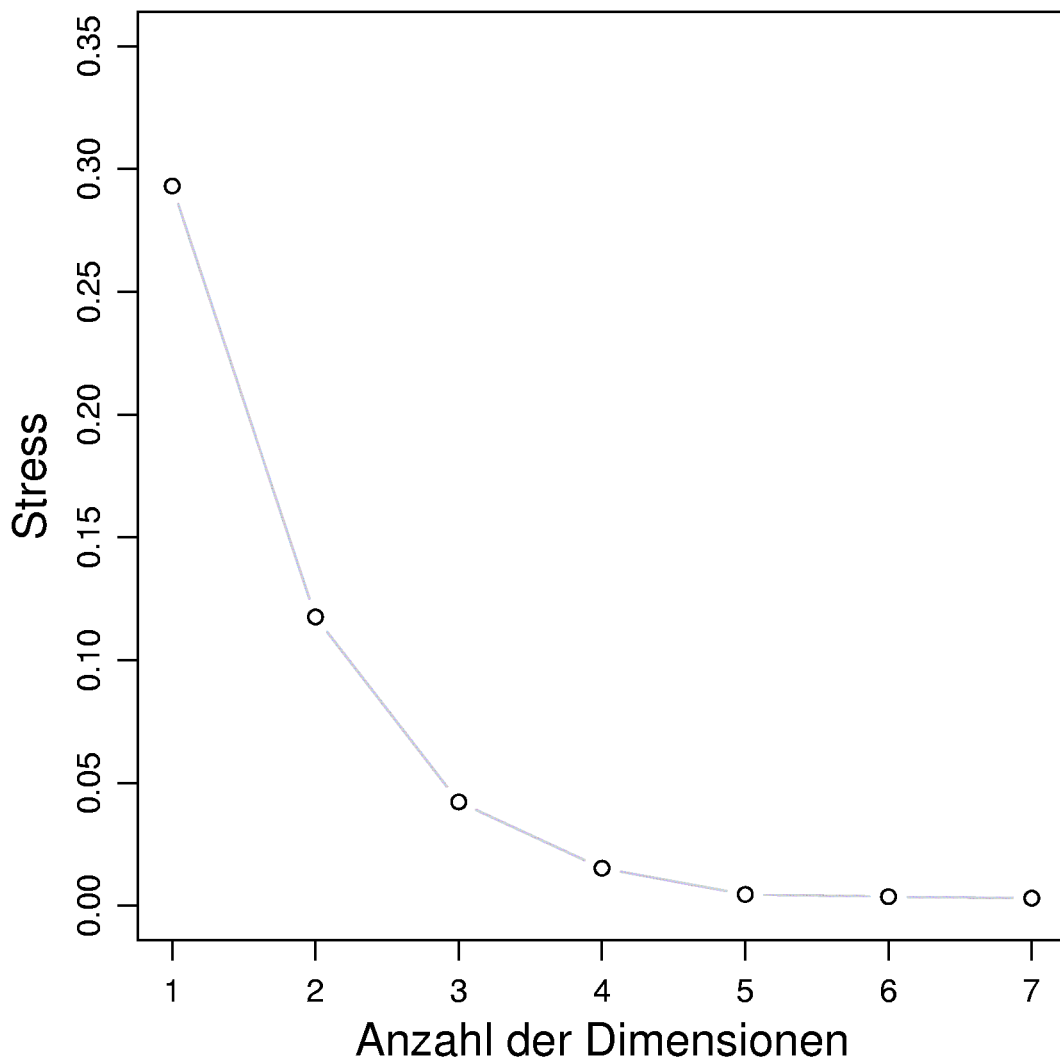


Abb. 19: Scree-Plot der MDS der PL-Items

Die Abszisse stellt die Anzahl der MDS-Dimensionen dar. Mit zunehmendem Wert auf der Abszisse nimmt entsprechend die Komplexität der Abbildung der Daten zu. Die Werte auf der Ordinate stellen das zugehörige Stress-Maß der Abbildung dar. Je geringer der Wert, desto genauer gibt die Abbildung die originalen Daten wieder. Die Auswahl einer geeigneten Anzahl von Dimensionen kann nach dem Elbow-Kriterium (Borg & Groenen, 2005, S. 47-53) erfolgen³. Hierbei werden die Punkte im Scree-Plot durch eine gedachte Linie verbunden. Falls diese Linie in einem Punkt erkennbar abknickt, sich also die Steigung vor und hinter diesem Punkt deutlich unterscheiden, so entspricht die durch diesen Punkt repräsentierte Anzahl

³ Die Scree-Methode zur Auswahl einer geeigneten Anzahl von Dimensionen bei einer MDS unterscheidet sich von der – im Kapitel „Korrelationsmatrix der Items“ erwähnten – Scree-Methode zur Auswahl einer geeigneten Anzahl an Faktoren einer Faktorenanalyse. Bei der MDS schließt die Anzahl der Dimensionen den Knick-Punkt mit ein, bei der Faktorenanalyse hingegen werden nur die durch Punkte vor dem Knick repräsentierte Faktoren als bedeutsam betrachtet; der durch den Knick-Punkt dargestellte Faktor ist bereits unbedeutend.

an Dimensionen der Anzahl bedeutsamer Dimensionen. Die Dimensionalität der MDS sollte entsprechend festgesetzt werden. In Abb. 19 ist kein ausgeprägter Knick zu erkennen, denkbar wären nach dem Elbow-Kriterium zwei oder drei Dimensionen als optimale Lösung.

Eine zusätzliche Entscheidungshilfe bietet die erreichte Güte der Abbildung. Hierfür hat Kruskal (zitiert nach Backhaus et al., 2006, S. 643) folgende Werte vorgeschlagen (Tab 4):

Anpassungsgüte	Stress
gering	0,200
ausreichend	0,100
gut	0,050
ausgezeichnet	0,025
perfekt	0,000

Tab 4: Anhaltswerte zur Beurteilung des Stress

Der Stress für eine MDS der PL-Items mit zwei Dimensionen beträgt 0.118, mit drei Dimensionen sinkt der Wert auf 0.042 ab. Da ein Wert größer 0.100 entsprechend Tab 4 im besten Fall nur noch als ausreichend betrachtet werden kann, wurde die Anzahl der Dimensionen auf drei festgelegt. Der zugehörige Stress entspricht einer guten Abbildung der Originalwerte durch die MDS.

Die gleiche Prozess für die Auswahl einer geeigneten Anzahl an Dimensionen wurde auch für die PEUEU-Items durchgeführt. Die Stress-Werte der MDS der PEUEU-Items sind in Abb. 20 als Scree-Plot dargestellt.

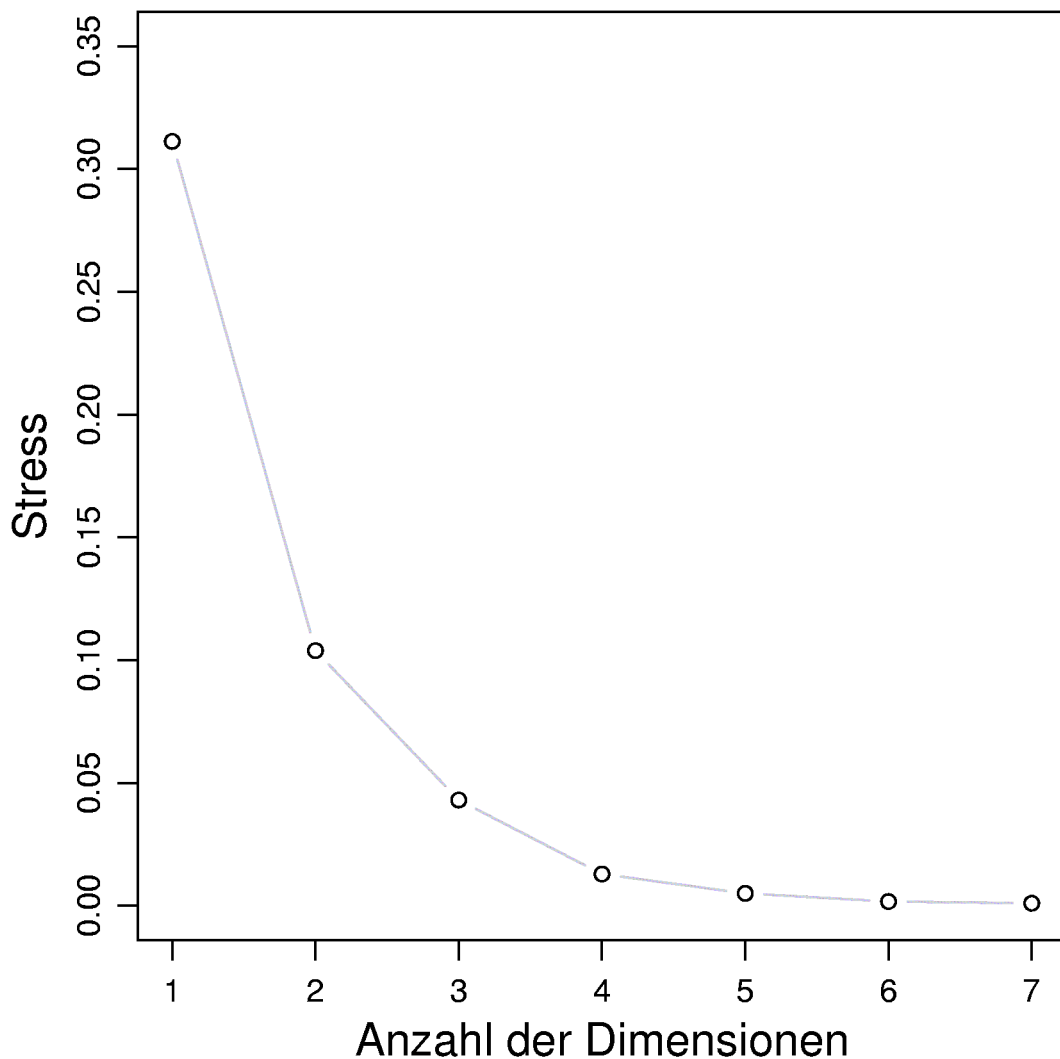


Abb. 20: Scree-Plot der MDS der PEUEU-Items

Auch im Scree-Plot zu den MDS der PEUEU-Items ist kein eindeutiger Knick zu erkennen. Mögliche Werte wären ebenfalls zwei oder drei Dimensionen. Der zugehörige Stress beträgt 0.104 für zwei Dimensionen und 0.043 für drei Dimensionen. Da ein Wert größer 0.100 bestenfalls noch als ausreichend betrachtet werden kann, während ein Wert kleiner 0.050 als gut zu betrachten ist, wurde – analog zum Vorgehen für die PL-Items – die geeignete Anzahl von Dimensionen auf drei festgesetzt.

Über die Ähnlichkeiten der PL-Items wurde eine dreidimensionale MDS gerechnet, um zu einer inhaltlichen Gliederung der Items zu gelangen. Auf dieser Grundlage sollten stark ähnliche Items identifiziert werden. Zu ähnliche Items sind in einer Skala nicht erwünscht, da das mehrfache Abfragen desselben Aspekts eines Konstrukts keine zusätzlichen Informationen liefert (Yousfi, 2005). Entsprechend sollten sich Items inhaltlich nicht zu sehr ähneln. Andererseits sollten keine Items aus der Skala entfernt werden, die wenig Ähnlichkeit zu allen anderen Items aufwiesen. Diese Items könnten einen sonst nicht mehr durch die Skala berücksichtigten Aspekt des Konstrukts erfassen, der in der Skala nicht verloren

gehen sollte.

Eine dreidimensionale MDS ordnet die Items auf der Grundlage ihrer Ähnlichkeitswerte räumlich so an, dass die Abstände zwischen den Items bestmöglich der Ähnlichkeit entsprechen. Die Ergebnisse für die PL-Items sind in Abb. 21 dargestellt.

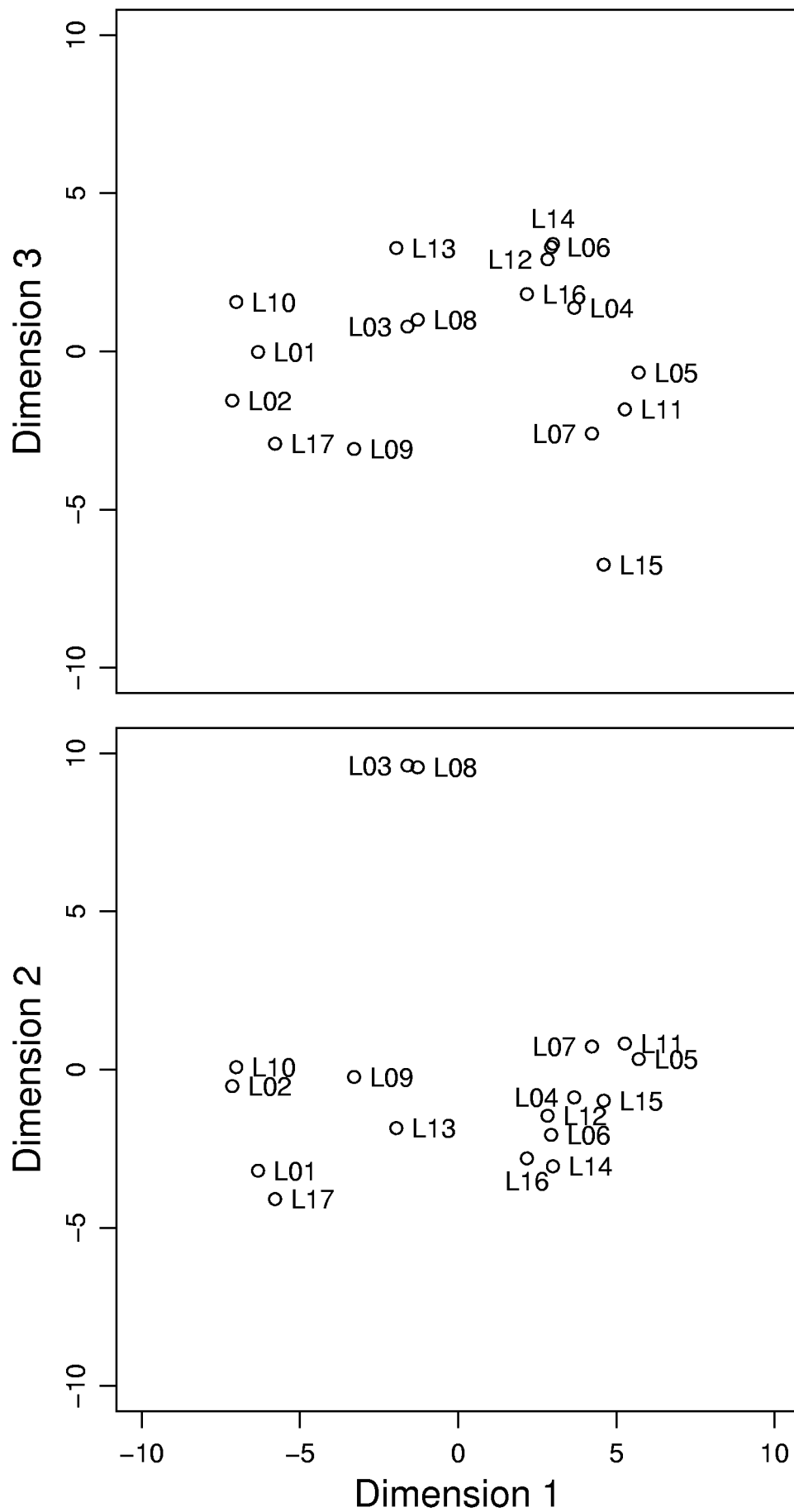


Abb. 21: 3-dimensionale MDS der PL-Items

Für die zweidimensionale Darstellung einer dreidimensionalen Anordnung sind zwei Abbildungen notwendig. Auf der Abszisse ist in beiden Abbildungen einheitlich die Position der Items in der ersten Dimension dargestellt. Die Ordinaten zeigen einmal die Position in der zweiten und in der anderen Abbildung die Position in der dritten Dimension an. Die Skalierung wurde für alle Achsen gleich gewählt, so dass eine leichte Vergleichbarkeit der Abstände in den unterschiedlichen Dimensionen gegeben ist.

Die Einheit der Achsen entspricht der Häufigkeit, mit der Items nicht auf denselben Stapel sortiert wurden, also einem Komplementärwert zur Ähnlichkeit, nämlich der Differenz zwischen dem Ähnlichkeitswert und der Gesamtzahl der Versuchspersonen. Der absolute Wert der Skalen lässt sich nicht sinnvoll interpretieren, sondern lediglich die Differenzen zwischen Items als Maß für deren Unterschiedlichkeit. Punkte, die in allen Abbildungen nahe beisammenliegen (z. B. L03 und L08) sind inhaltlich stark ähnlich. Punkte, die in einer (z. B. L04 und L15) oder mehreren (z. B. L03 und L15) Dimensionen weit auseinander liegen, unterscheiden sich inhaltlich stark voneinander. Auf eine Benennung der Dimensionen der MDS wurde verzichtet. Eine entsprechende inhaltliche Bezeichnungen der Dimensionen erschien schwierig und somit problembehaftet und ist für den Selektionsprozess der Items an sich nicht notwendig.

Auf der Grundlage von Abb. 21 wurden folgende Items zu Clustern zusammengefasst:

Cluster A:

- L03: SOFTWARE verhält sich so, wie ich es von anderen Programmen kenne.
- L08: Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen.

Cluster B:

- L05: SOFTWARE wirkt für mich sehr übersichtlich.
- L07: Die Funktion von allen Bedienelementen ist offensichtlich.
- L11: Die Oberfläche von Software wirkt auf mich gut strukturiert.

Cluster C:

- L01: Die Bedienung von SOFTWARE ist für mich leicht zu lernen.
- L02: Es ist leicht für mich ein geschickter Nutzer von SOFTWARE zu werden.
- L10: Ich kann mich leicht in die Bedienung von SOFTWARE einarbeiten.
- L17: Ich kann die Bedienung von SOFTWARE gut durch Ausprobieren erlernen.

Cluster D:

- L04: Es ist leicht für mich zu erkennen, welche Funktionen SOFTWARE bereitstellt.
- L06: Wie SOFTWARE funktioniert ist für mich leicht zu erkennen.
- L12: Die Struktur von SOFTWARE ist für mich leicht zu erkennen.
- L14: Die Funktionen von SOFTWARE sind für mich leicht zu erkennen.

- L16: Die Bedienung von SOFTWARE ist für mich leicht verständlich.

Die übrigen Items wurden keinen Clustern zugeordnet, da keine zu ihnen inhaltlich ähnlichen Items vorhanden waren. Diese Items sind im Folgenden aufgelistet:

- L09: Ich brauche keine besonderen Kenntnisse für die Nutzung von SOFTWARE.
- L13: Ich komme mit SOFTWARE auf Anhieb gut zurecht.
- L15: Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE.

Abb. 22 entspricht der Abb. 21, wobei die zu Clustern zusammengefassten Items in der Darstellung durch gleiche Symbole abgebildet sind. Alle Items, deren Punkte in der Abbildung durch dasselbe Symbol dargestellt sind, wurden also zu einem Cluster zusammengefasst.

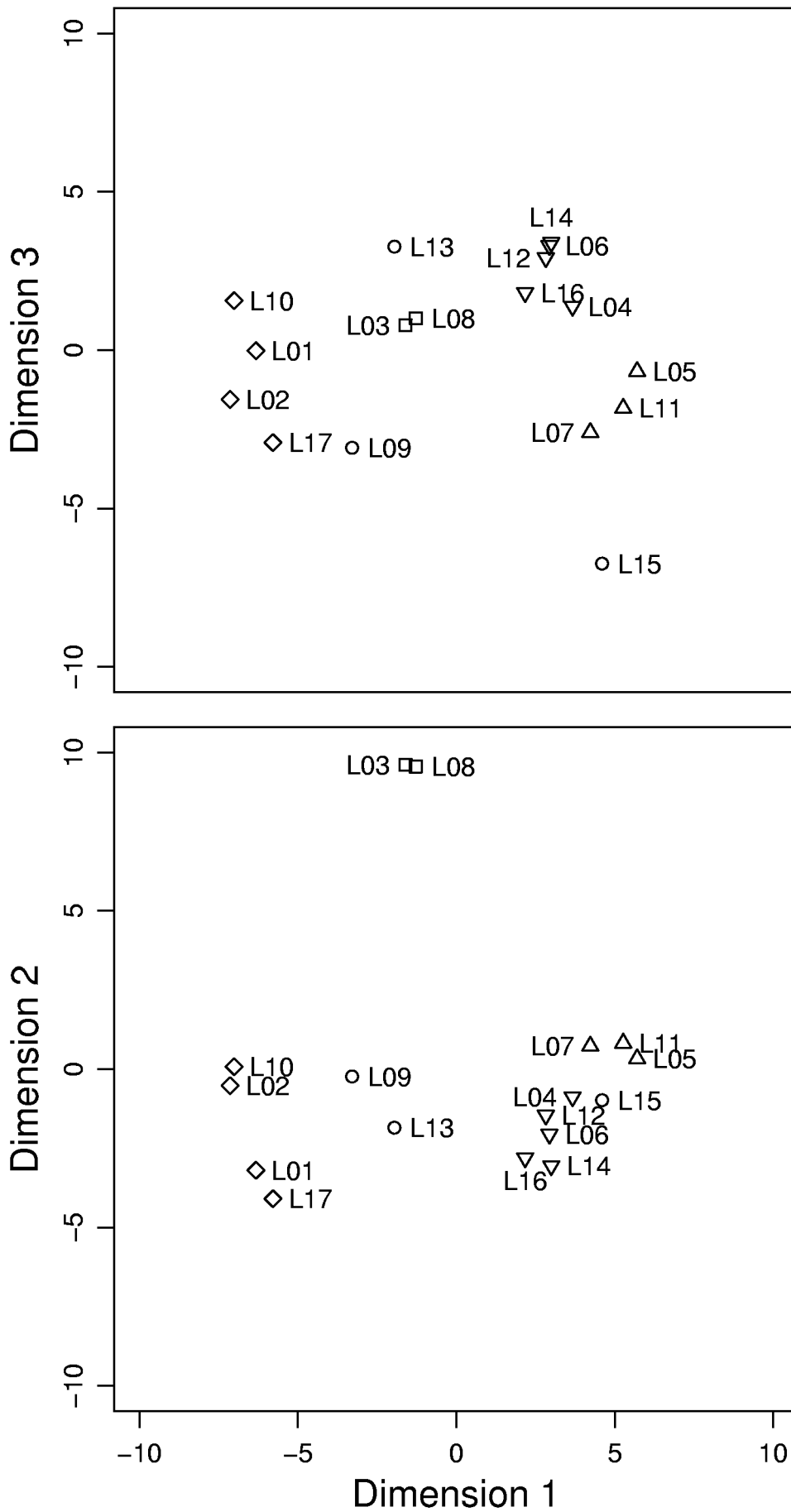


Abb. 22: 3-dimensionale MDS der PL-Items (Cluster hervorgehoben)

Für die Clusterung wurde kein absolutes Kriterium wie der maximale Abstand zwischen zwei Items bezüglich einer Dimension oder im Raum festgelegt. Die Items innerhalb des Cluster C haben teilweise einen größeren Abstand zueinander (L10 und L17) als eine Itempaarung mit je einem Item aus Cluster B und Cluster D (L04 und L05). Dies ist unproblematisch, da der Zweck der Clusterung nicht in einer absoluten Bewertung eines Items in Bezug auf eine äußeres Kriterium bestand, sondern die Items nur im Verhältnis zueinander bewertet wurden. Innerhalb der größeren Ansammlung der Items aus Cluster B und D zusammen hat die Aufteilung eine zusätzliche Strukturierung geleistet, die für die Itemauswahl hilfreich war.

Aus jedem Cluster sollten zwei Items in der Skala enthalten bleiben. Somit konnte in einem weiteren Schritt der Skalenentwicklung noch ein Item pro Cluster entfernt werden, ohne den inhaltlichen Aspekt des Clusters aus der Skala zu verlieren. Mehr als zwei Items aus einem Cluster zu behalten, erschien nicht optimal, da hierdurch die inhaltliche Redundanz der Items unnötig groß geworden wäre. Die Auswahl der auszuschließenden Items wurde über die Platzierung des Items in der eindimensionalen MDS (Abb. 18) getroffen. Die Wahl fiel hier auf die MDS, da die Abfolgen der Items in den unterschiedlichen Reihen teilweise voneinander abwichen, und die MDS durch die Verdichtung der Daten das höchste Skalenniveau (intervallskaliert, Backhaus et al., 2006, S. 645-647) erreichte.

Aus Cluster A wurden keine Items ausgeschlossen, da nur zwei Items in diesem Cluster enthalten waren. Aus Cluster B wurde das Item L05 (SOFTWARE wirkt für mich sehr übersichtlich) ausgeschlossen, da es sich in der MDS am wenigsten von den PEUEU-Items unterschied. Aus Cluster C wurden die Items L10 (Ich kann mich leicht in die Bedienung von SOFTWARE einarbeiten) und L17 (ich kann die Bedienung von SOFTWARE gut durch Ausprobieren erlernen) entfernt. Aus dem größten Cluster D mussten drei Items eliminiert werden. Dies waren die Items L04 (Es ist leicht für mich zu erkennen, welche Funktionen SOFTWARE bereitstellt), L14 (Die Funktionen von SOFTWARE sind für mich leicht zu erkennen) und L16 (Die Bedienung von SOFTWARE ist für mich leicht verständlich). Die Items, die keinem Cluster zugeordnet werden konnten, blieben alle in der Skala enthalten. Da für sie jeweils kein ähnliches Item existierte, wäre inhaltlich in jedem Fall ein Aspekt des Konstrukts aus der Skala verloren gegangen. Im folgenden ist die Anordnung der Items abgebildet, die sich in der dreidimensionalen MDS ergeben hat. Allerdings sind in Abb. 23 nur noch die Items dargestellt, die in der Skala verblieben sind. So ist eine visuelle Kontrolle möglich, dass kein vorher durch ein Item abgedeckter Bereich des Raumes durch den Ausschluss eines hierfür wichtigen Items unberücksichtigt ist.

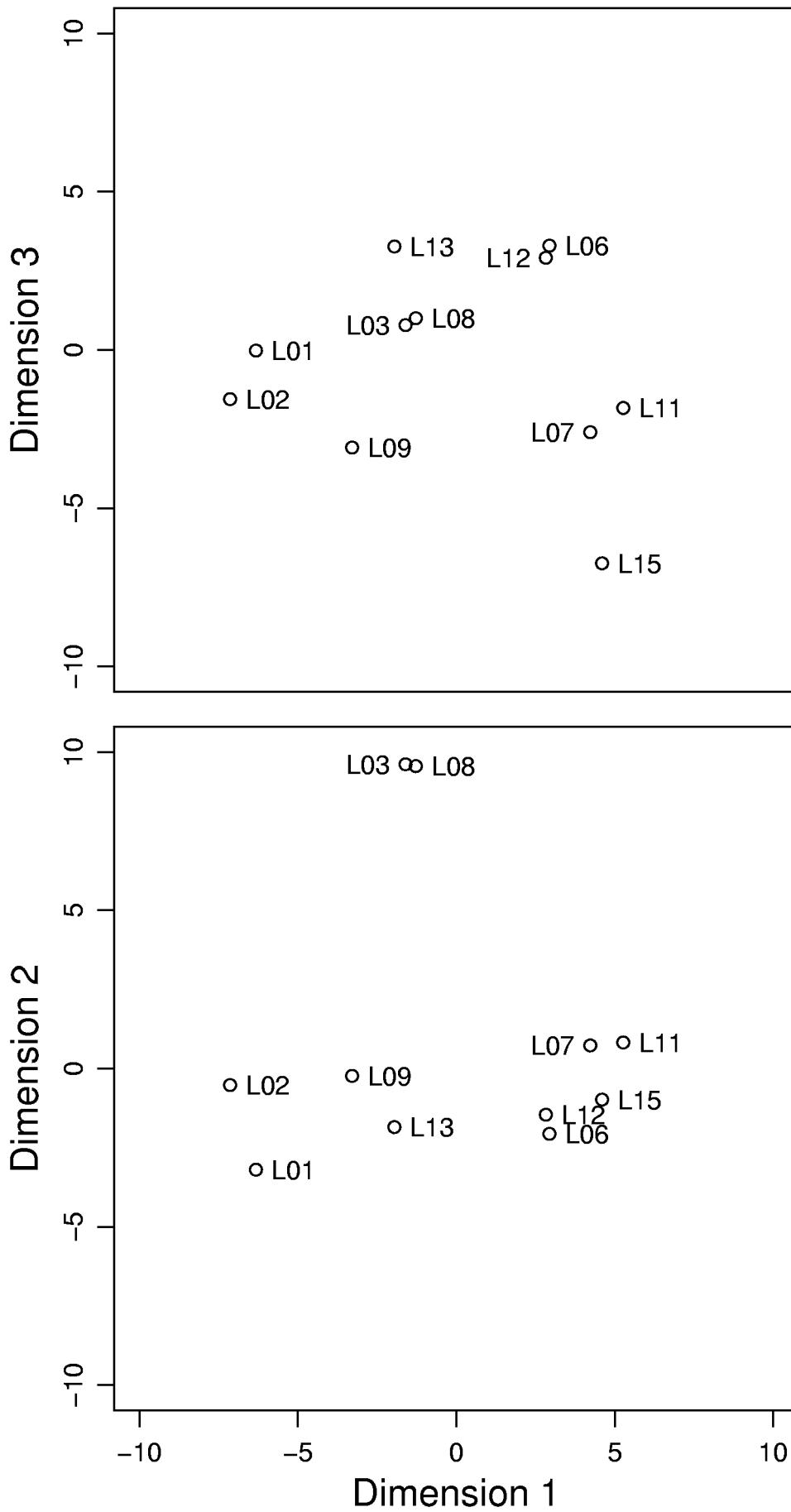


Abb. 23: 3-dimensionale MDS der PL-Items (eliminierte Items nicht dargestellt)

Die verbliebenen Items für die PL-Skala sind im Folgenden zusammenfassend aufgelistet:

- L01: Die Bedienung von SOFTWARE ist für mich leicht zu lernen.
- L02: Es ist leicht für mich ein geschickter Nutzer von SOFTWARE zu werden.
- L03: SOFTWARE verhält sich so, wie ich es von anderen Programmen kenne.
- L06: Wie SOFTWARE funktioniert ist für mich leicht zu erkennen.
- L07: Die Funktion von allen Bedienelementen ist offensichtlich.
- L08: Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen.
- L09: Ich brauche keine besonderen Kenntnisse für die Nutzung von SOFTWARE.
- L11: Die Oberfläche von Software wirkt auf mich gut strukturiert.
- L12: Die Struktur von SOFTWARE ist für mich leicht zu erkennen.
- L13: Ich komme mit SOFTWARE auf Anhieb gut zurecht.
- L15: Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE.

Auch für die MDS der PEUEU-Items waren drei Dimensionen als optimale Anzahl festgelegt worden. Auch hier sollte die Anordnung der Items durch die MDS Aufschluss darüber geben, wie ähnlich sich Items sind. Abb. 24 stellt die Ergebnisse der MDS der PEUEU-Items dar. In dieser Analyse sind die Items E05 und E10 nicht mehr enthalten, da sie bereits wegen Verständnisproblemen (E05) bzw. wegen starker Assoziation zum Konstrukt PL (E10) ausgeschlossen worden waren..

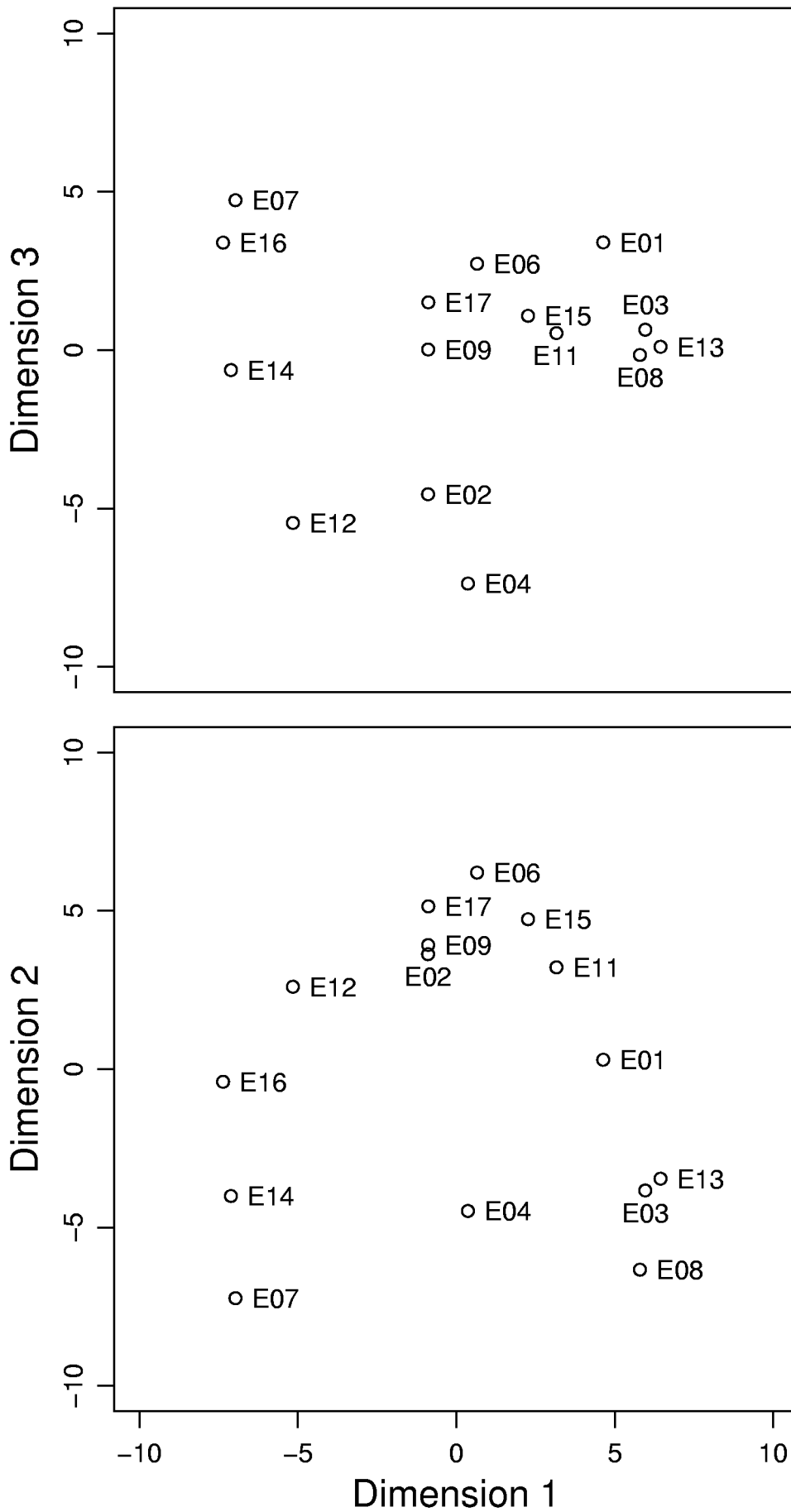


Abb. 24: 3-dimensionale MDS der PEUEU-Items

Die Abszisse gibt in beiden Diagrammen die Position der Items in der ersten Abbildungsdimension der MDS wieder. Die Ordinaten der Diagramme zeigen die Position des Items in der zweiten bzw. in der dritten Abbildungsdimension an. In den Diagrammen sind alle Achsen einheitlich skaliert, damit die entsprechenden Werte leicht verglichen werden können. Die Einheit der Achsen entspricht dem Komplementärwert der Häufigkeit, mit der zwei Items auf denselben Stapel sortiert wurden, ist also ein Maß dafür, wie unähnlich sich Items sind. Absolute Werte lassen sich nicht interpretieren, sondern lediglich die Distanzen zwischen den Items haben Aussagekraft. Auch für die PEUEU-Items wurde auf eine inhaltliche Benennung der Dimensionen verzichtet.

Auf der Grundlage von Abb. 24 wurden folgende Items zu Clustern zusammengefasst:

Cluster A:

- E06: Die Eingabe von Informationen in SOFTWARE ist nicht aufwendig.
- E09: Ich muss keine unnötigen Arbeitsschritte machen, wenn ich SOFTWARE bediene.
- E11: Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen.
- E15: Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen.
- E17: Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren.

Cluster B:

- E03: SOFTWARE passt zu meinem Arbeitsstil.
- E05: SOFTWARE wirkt für mich sehr übersichtlich.
- E08: Die Bedienung von SOFTWARE entspricht genau meinen Anforderungen.
- E13: Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben.

Die übrigen Items wurden keinen Clustern zugeordnet, da keine zu ihnen inhaltlich ähnlichen Items vorhanden waren. Diese Items sind im Folgenden aufgelistet:

- E01: Mit SOFTWARE kann ich effizient arbeiten.
- E02: Wichtige Funktionen sind für mich in SOFTWARE schnell erreichbar.
- E04: Die Bedienoberfläche von SOFTWARE ist für mich optimal gestaltet.
- E07: Ich fühle mich durch SOFTWARE nicht bevormundet.
- E12: Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar.
- E14: Ich werde durch die Oberfläche von SOFTWARE nicht eingeschränkt.
- E16: Ich bekomme von SOFTWARE keine für mich überflüssigen Hilfestellungen.

In Abb. 25 sind die zu Clustern zusammengefassten Items durch ein einheitliches Symbol dargestellt. Die Items des Clusters A werden durch dasselbe Symbol repräsentiert, ebenso die Items, die Cluster B zugewiesen wurden.

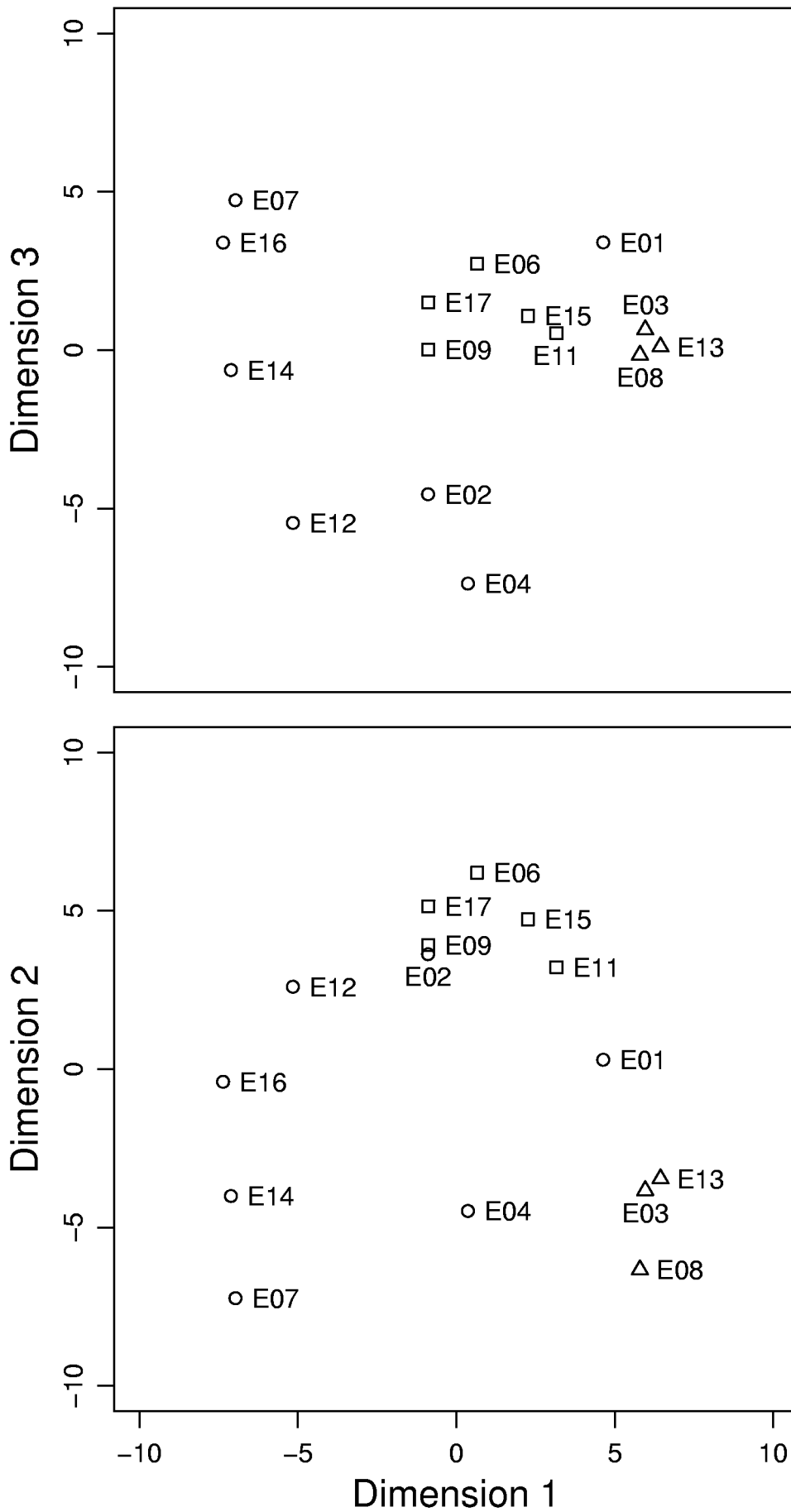


Abb. 25: 3-dimensionale MDS der PEUEU-Items (Cluster hervorgehoben)

Auch für die Clusterung der PEUEU-Items wurde kein absolutes Kriterium angesetzt. Wie schon bei den PL-Items bestand auch bei den PEUEU-Items der Sinn der Clusterung darin, stark ähnliche Items zu identifizieren. Um eine Aufnahme zu ähnlicher Items, die denselben Aspekt eines Konstrukts mehrfach abfragen, in die Skala zu vermeiden, wurden auch von den PEUEU-Items jeweils nur zwei pro Clustern in der Skala belassen. Die Auswahl, welche Items pro Cluster in der Skala blieben, erfolgte – wie auch bei den PL-Items – nach der Platzierung in der eindimensionalen MDS.

Cluster A enthielt fünf Items, entsprechend mussten drei Items ausgeschlossen werden, damit zwei Items dieses Clusters in der Skala blieben. Die drei Items mit einer Platzierung am weitesten in Richtung der PL-Items in der eindimensionalen MDS (Abb. 18) waren Item E11 (Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen), E15 (Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen) und E17 (Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren).

Item E17 war das einzige PEUEU-Item des Itempools, welches den Aspekt der *Fehlerkorrektur* im Rahmen von PEUEU adressierte. Da *Fehlerkorrektur* als Aspekt in den Stichwortsammlungen zu beiden Konstrukten genannt worden war und im Rahmen der Bildung und Benutzung mentaler Modelle eine zentrale Bedeutung hat, wurde das Item entgegen der Auswahlkriterien nicht aus der Skala entfernt. Das Item wurde als zusätzliches Item im Itempool belassen, so dass aus Cluster A drei Items in der Skala blieben.

Von den drei Items des Clusters B war das Item E08 (Die Bedienung von SOFTWARE entspricht genau meinen Anforderungen) in der eindimensionalen MDS am weitesten zu den PL-Items platziert worden. Entsprechend wurde dieses Item aus dem Itempool entfernt. Alle übrigen Items, die keinem Cluster zugeordnet worden waren, blieben im Itempool enthalten, um weiterhin eine möglichst breite Abdeckung des Konstrukts zu gewährleisten. Abb. 26 zeigt die dreidimensionale Abbildung der PEUEU-Items durch die MDS an, wobei die aus dem Itempool entfernten Items nicht mehr dargestellt sind. Die Abbildung ermöglicht eine visuelle Kontrolle, ob alle vor der Itemauswahl abgedeckten Bereiche der drei Dimensionen mit Items besetzt sind.

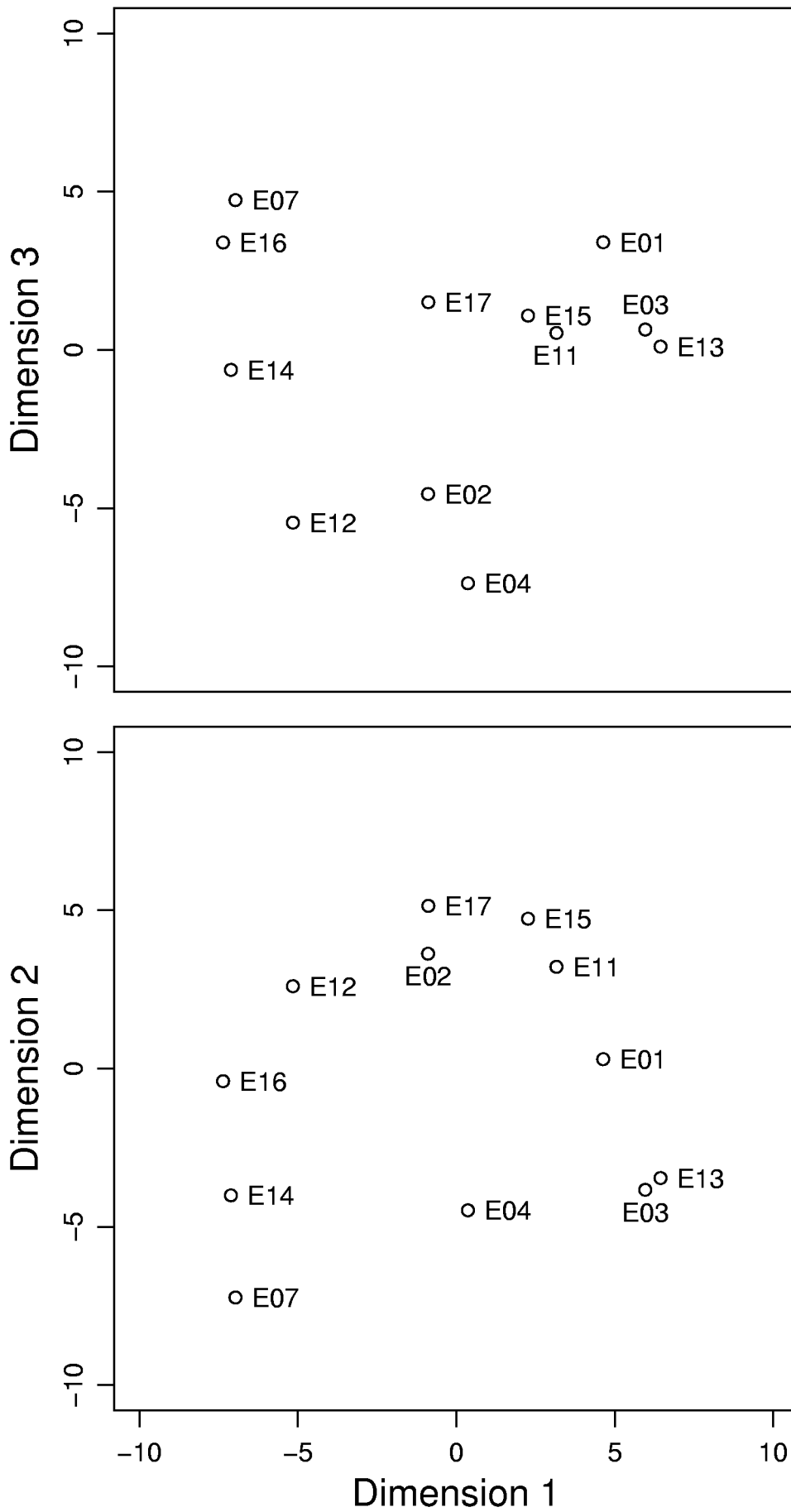


Abb. 26: 3-dimensionale MDS der PEUEU-Items (eliminierte Items nicht dargestellt)

Zusammengefasst sind folgende Items im Itempool für die PEUEU-Skala verblieben:

- E01: Mit SOFTWARE kann ich effizient arbeiten.
- E02: Wichtige Funktionen sind für mich in SOFTWARE schnell erreichbar.
- E03: SOFTWARE passt zu meinem Arbeitsstil.
- E04: Die Bedienoberfläche von SOFTWARE ist für mich optimal gestaltet.
- E07: Ich fühle mich durch SOFTWARE nicht bevormundet.
- E11: Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen.
- E12: Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar.
- E13: Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben.
- E14: Ich werde durch die Oberfläche von SOFTWARE nicht eingeschränkt.
- E15: Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen.
- E16: Ich bekomme von SOFTWARE keine für mich überflüssigen Hilfestellungen.
- E17: Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren.

Zusammenfassung der ersten Itemauswahl

Die erste Schritt zur Auswahl geeigneter Items basierte auf je drei verschiedenen Sortierungen der Items, die von 16 Personen gebildet worden waren. Es wurden sowohl zwei verschiedene Rangreihen der Items gebildet, als auch alle ähnlichen Items zu Stapeln zusammengefasst. Die beiden Rangreihen bestätigten die Trennung in die Konstrukte PL und PEUEU. Im wesentlichen wurden alle PL-Items in die eine Hälfte der Rangreihenfolgen und alle PEUEU-Items in die andere Hälfte sortiert.

Eine deutliche Ausnahme bildet das Item E10 (Mit Übung kann ich SOFTWARE fehlerfrei nutzen), welches als Konsequenz aus dem Itempool entfernt wurde. Bei Item E05 (Die Ausgaben von SOFTWARE sind genau so wie ich es mir wünsche) gab es mehrfach Verständnisschwierigkeiten, die eine unklare Formulierung des Items aufzeigten. Auch dieses Item wurde aus dem Itempool ausgeschlossen.

Die Daten der Stapelbildung wurden so aggregiert, dass für jedes Itempaar ein Wert dafür vorlag, von wievielen Probanden das Itempaar gemeinsam auf einen Stapel sortiert worden war. Die entsprechenden Werte bilden ein Maß für die Ähnlichkeit der Items. Die Ähnlichkeitsmatrix wurde durch eine Multidimensionale Skalierung vereinfacht.

Die einfachste Abbildung der Itemähnlichkeiten ordnete diese eindimensional entlang einer Achse an. Auch durch die MDS wurden die Items so angeordnet, dass sich im Wesentlichen eine Hälfte mit PL-Items und die andere Hälfte mit PEUEU-Items bildete.

Für die abschließende Beurteilung der inhaltlichen Ähnlichkeit war die eindimensionale MDS nicht geeignet, da sie die Ähnlichkeitsstruktur der Items zu stark vereinfachte. Die Analyse mehrerer MDS, die jeweils getrennt für die PL- und die PEUEU-Items berechnet wurden, zeigte, dass die Ähnlichkeitsstruktur der Items für beide Konstrukte durch dreidimensionale MDS optimal abgebildet werden

konnten.

In der dreidimensionalen MDS der PL-Items waren vier Cluster aus zwei bis fünf Items erkennbar. Drei Items konnten keinem Cluster zugeordnet werden, da ihre Distanz zu den umliegenden Items zu groß war. Aus den vier Clustern wurden jeweils zwei Items beibehalten, die übrigen wurden aus dem Itempool entfernt. Die Auswahl erfolgte so, dass die beiden Items beibehalten wurden, die sich in der eindimensionalen MDS näher am Rand der Anordnung befanden, also weiter von den Items des anderen Konstrukts am anderen Ende der Anordnung entfernt waren. Die drei Items, die keinem Cluster zugeordnet werden konnten, wurden alle im Itempool belassen. Der Itempool für die PL-Skala bestand nach der Selektion noch aus 11 Items.

Die inhaltliche Struktur der PEUEU-Items, die durch die dreidimensionale MDS aufgezeigt wurde, unterschied sich deutlich von der der PL-Items. Für die PEUEU-Items waren nur zwei Cluster erkennbar, von denen das erste fünf Items und das zweite drei Items umfasste. Die übrigen sieben Items verteilten sich so innerhalb der drei Dimensionen, dass ein Zusammenfassen weiterer Items nicht sinnvoll möglich war. Auch für die PEUEU-Skala wurden aus jedem Cluster die beiden Items ausgewählt, die in der eindimensionalen MDS aller Items näher zum Rand der Anordnung platziert worden waren. Die anderen Items wurden mit einer Ausnahme aus der Skala ausgeschlossen.

Aus dem ersten gebildeten Cluster hätte Item E17 (Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren) ausgeschlossen werden müssen. Jedoch nur dieses Item adressierte den Aspekt der Fehlerkorrekturmöglichkeiten für erfahrene Benutzer. Die Möglichkeit zur Fehlerkorrektur ist wichtig für die Überprüfung gebildeter mentaler Modelle, spielt also in der Lernphase eine wichtige Rolle. Trotzdem war die Möglichkeit zur einfachen Korrektur gemachter Fehler auch als Aspekt für PEUEU genannt worden. Da Unklarheit darüber bestand, ob die Möglichkeit zur Fehlerkorrektur eindeutig einem der beiden Konstrukte zugeordnet werden könnte, wurde E17 als zusätzliches Item im Itempool belassen.

Auch bei den PEUEU-Items wurden alle Items, die keinem Cluster zugewiesen werden konnten, im Itempool belassen. Hiermit bestand der Itempool für die PEUEU-Skala nach der Selektion aus 12 Items. Für beide Skalen zusammen ergaben sich 23 Items.

Zweite Itemselektion – Beurteilung von Routenplanern

Die 23 Items der beiden Skalen wurden in einer Erhebung zur Akzeptanz von Online-Routenplanern (z. B. www.reiseplanung.de) eingesetzt. Hierbei wurde eine deutlich größere Stichprobe (n=45) als für die erste Itemselektion erhoben, so dass die Eigenschaften der Items auf der Grundlage verschiedener statistischer Parameter beurteilt werden konnten. Die Probanden bewerteten die Itemaussagen auf fünfstufigen Antwortskalen. Bei der Itemauswahl wurden sowohl Charakteristika der

einzelnen Items (z. B. die Verteilungsform der Antworten) wie auch das Korrelationsmuster zwischen den Items berücksichtigt. Die angewandten Verfahren und die Ergebnisse werden nachfolgend beschrieben. Auf der Grundlage der durchgeführten Auswertungen konnten die beiden Skalen für PL und PEUEU auf jeweils acht Items verkürzt werden.

Untersuchung

Für eine statistische Beurteilung der Items wurden diese in einer Untersuchung zur Bewertung von Online-Routenplanern eingesetzt. Die Probanden mussten 30 Minuten lang vorgegebene Aufgaben mit den Routenplanern bearbeiten, um einen Eindruck der Routenplaner zu bekommen. Jeder Proband arbeitete ausschließlich an einem von drei verschiedenen Routenplanern. Als Routenplaner wurden www.mappy.de, www.reiseplanung.de und www.viamichelin.de eingesetzt. Durch die verschiedenen Routenplaner sollte die Variation der erhobenen Daten vergrößert werden. Die Untersuchung wurde im Rechnerraum der Universität Lüneburg durchgeführt, in dem 12 Probanden parallel an den Routenplanern arbeiten konnten. Um Reihenfolge-Effekte zu vermeiden, wurde in allen vier Sitzungen mit allen Routenplanern gearbeitet. Benachbarte Probanden arbeiteten dabei stets an unterschiedlichen Routenplanern, um eine gegenseitige Beeinflussung der Probanden möglichst auszuschließen.

An der Untersuchung nahmen 45 Psychologiestudenten/innen teil. Das Alter bewegte sich zwischen 19 und 37 Jahren bei einem Median von 22 Jahren. 40 der 45 Versuchspersonen waren weiblich. Alle Versuchspersonen hatten Erfahrung im Umgang mit einem Computer und dem Internet seit mindestens 2 Jahren. Der Median der Computererfahrung beträgt 9 Jahre und der Median der Erfahrung mit dem Internet 6 Jahre. Alle 45 Versuchspersonen gaben an, über Zugang zum Internet zu verfügen.

Aufgabenstellung

Insgesamt waren 6 Planungsaufgaben formuliert, um die Probanden 30 Minuten lang mit dem Routenplaner zu beschäftigen. Die Ergebnisse der Planungsaufgaben waren für diese Arbeit irrelevant und wurden nicht ausgewertet. Die Probanden wurden hierüber nicht informiert. Die Aufgaben gaben Start- und Zielort einer fiktiven Reise, die geplant werden sollte vor. Die Aufgaben sind in Anhang G enthalten. Teilweise bauten die Aufgaben aufeinander auf. So sollte in Aufgabe 4 die Rückreise zum Startort der vorausgegangenen Aufgabe geplant und dabei modifiziert werden. Zu allen Aufgaben wurden den Probanden Fragen gestellt, die eine gründliche Beschäftigung mit den Aufgaben sicherstellen sollten. Während der Bearbeitung der Aufgaben sollten die Probanden die Antworten auf diese Kontrollfragen in einen Antwortbogen eintragen. Die Probanden sollten so viel Aufgaben wie möglich innerhalb der vorgegebenen Zeit bearbeiten. Im Anschluss an die Bearbeitung der Planungsaufgaben bewerteten die Probanden den benutzten

Routenplaner mit den Items der neuen Skalen. Je nach Routenplaner wurde hierbei der Platzhalter *SOFTWARE* in der Itemformulierung durch *mappy*, *reiseplanung.de* oder *ViaMichelin* ersetzt. Für die Abfolge der Items in den Fragebögen gab es drei Varianten. Dies sollte Reihenfolge-Effekt verringern. Insgesamt kamen somit neun verschiedene Varianten des Fragebogens zum Einsatz. Die Zustimmung zu den Items erfolgte auf einer fünfstufigen Antwortskala, deren Layout und Formulierungen von Rohrman (1978, S. 231, Skala B: Bewertung von Aussagen) übernommen wurden.

Auswertungen

Die Itemauswahl für die finale Skala wurde anhand verschiedener Kriterien getroffen. Die Verteilungsform der Antworten für jedes Item sollten annähernd einer Normalverteilung entsprechen. Dies sicherte eine mittlere Schwierigkeit (Erläuterung folgt im nächsten Kapitel) der Items und ist gleichzeitig Voraussetzung für weitere statistische Analysen. Neben der geeigneten Verteilungsform der Antworten mussten die Korrelationen eines Items mit den anderen Items der zugehörigen Skala innerhalb gewisser Grenzen liegen. Eine zu niedrige Korrelation zwischen zwei Items ist hierbei ebenso wenig wünschenswert wie eine zu hohe Korrelation. Bei einer zu niedrigen Korrelation stellt sich die inhaltliche Frage, ob die beiden Items dasselbe Konstrukt erfassen. Bei einer zu hohen Korrelation gewinnt die Skala zu wenig an zusätzlicher Information durch die Aufnahme beider Items (Cattell und Tsujikawa, 1964).

Ein verbreiteter Parameter zur Beurteilung der Reliabilität einer Skala ist Cronbachs α ; auch dieser Parameter wurde berücksichtigt. Für die ausgewählten Items wurde eine Faktorenanalyse durchgeführt. Diese sollte zeigen, ob die Zusammenhänge zwischen den Items den theoretischen Erwartungen aufgrund der zugrundeliegenden Konstruktstruktur entsprachen.

Verteilungsform der Itemantworten

Bei der statistischen Beurteilung eines Items spielt die Verteilungsform der Antworten eine wichtige Rolle. Die Antworten müssen sich aus inhaltlichen Gründen zwar nicht normal (im Sinne einer Gaußschen Glockenverteilung) verteilen, allerdings setzen statistische Verfahren, die auf einer Interpretation der Inter-Itemkorrelationen aufbauen (z. B. die Faktorenanalyse) eine Normalverteilung der Antworten voraus (Bortz, 1993, S. 197-198). Zusätzlich kann anhand der Verteilung der Antworten auf die Schwierigkeit eines Items geschlossen werden.

Unter Itemschwierigkeit versteht man, wie hoch die Lösungswahrscheinlichkeit für dieses Item ist. Der Begriff soll zunächst am Beispiel zweistufiger Items erläutert werden. Itemschwierigkeit ist ein statistisches Maß und beschreibt, wie sich die Antworten auf die beiden Möglichkeiten verteilen. Im Falle von Intelligenztests stimmt die tatsächliche Schwierigkeit mit der statistischen Schwierigkeit überein. Hier ist eine Aufgabe schwierig, die nur von wenigen Probanden gelöst werden kann. Geht

die Lösungsquote einer Aufgabe gegen null, d.h. fast kein Proband kann die Aufgabe lösen, so gewinnt man durch dieses Item keine Information, da es nicht zur Differenzierung der Probanden beiträgt. Gleiches gilt analog auch für zu leichte Items. Bei mehrstufigen Skalen beschreibt die Itemschwierigkeit, wie stark sich die Antworten am einen oder am anderen Ende der Skala ballen. Verteilen sich die gegebenen Antworten nicht ausreichend über den Skalenbereich, so ist ein Item zu einfach oder zu schwierig und liefert somit wenig zusätzliche Information. Über die Verteilungsform können solche Items leicht erkannt werden, da sie eine schiefe Verteilung aufweisen.

Die Betrachtung der Verteilungen der Antworten sicherte somit die statistischen Vorbedingungen für weitere Analysen ab und half gleichzeitig, zu einfache sowie zu schwierige Items zu identifizieren. Da durch die fünfstufige Antwortskala keine stetige Verteilung der Antworten gegeben war, wurde ein χ^2 -Test zur Überprüfung auf Normalverteilung eingesetzt (Bortz, Lienert & Boehnke, 2000). Der Mittelwert und die Standardabweichung der Sollverteilung wurden aus den Daten des jeweiligen Items geschätzt. Die erwarteten Häufigkeiten eines Skalenwerts wurden aus den entsprechenden Quantilen errechnet, wobei die arithmetischen Mittelwerte zwischen zwei Skalenstufen als Grenzen für die Quantilen dienten.

Da bei der Überprüfung auf Abweichung von der Normalverteilung eigentlich der β -Fehler interessiert, dieser jedoch nur gegen eine spezifische H_1 getestet werden kann, empfehlen Bortz et al. ein hohes α -Niveau anzusetzen. Die Überprüfung der Verteilungsform wurde mit einem α -Grundniveau von 5 % durchgeführt.

Da für jedes Item die Durchführung eines Tests auf Abweichung von der Normalverteilung erforderlich war, wurden insgesamt 23 Tests auf Abweichung von der Normalverteilung durchgeführt. Hierbei steigt die kumulierte Wahrscheinlichkeit zufällig signifikanter Ergebnisse deutlich über die zuvor festgelegte Irrtumswahrscheinlichkeit α . Um die Gesamtirrtumswahrscheinlichkeit auf dem angestrebten Sollwert zu halten, hat Bonferroni eine Korrekturmethode für das α -Niveau der Einzeltest vorgeschlagen. Diese wurde von Holm (1979) zu einem weniger konservativen Verfahren modifiziert. Die Methode vom Holm wurde eingesetzt, um das α -Niveau für die Einzeltests zu justieren. Da sie bezüglich des α -Fehlers weniger konservativ ist als die Bonferroni-Methode, entscheidet sie bezüglich des β -Fehler konservativer, was hier durchaus gewünscht ist. Die Kennwerte der χ^2 -Test sind in Tab 5 im Einzelnen aufgelistet.

Item	χ^2 -Wert	p-Wert	α -Wert (nach Holm)	Ergebnis
L03	12,29	0,0153	0,0022	n. sig.
L13	7,51	0,1110	0,0023	n. sig.
E07	6,66	0,1548	0,0024	n. sig.
L12	5,61	0,2304	0,0025	n. sig.
L07	5,39	0,2492	0,0026	n. sig.
E03	5,39	0,2497	0,0028	n. sig.
L09	5,38	0,2506	0,0029	n. sig.
E02	3,96	0,4116	0,0031	n. sig.
E15	3,72	0,4457	0,0033	n. sig.
E16	3,70	0,4479	0,0036	n. sig.
E17	2,97	0,5635	0,0038	n. sig.
E04	2,77	0,5968	0,0042	n. sig.
L15	2,77	0,5972	0,0045	n. sig.
E01	2,70	0,6099	0,0050	n. sig.
E14	2,54	0,6369	0,0056	n. sig.
L11	2,35	0,6723	0,0063	n. sig.
E11	2,28	0,6847	0,0071	n. sig.
L01	2,11	0,7159	0,0083	n. sig.
L06	1,92	0,7510	0,0100	n. sig.
E12	1,59	0,8108	0,0125	n. sig.
L08	0,63	0,9598	0,0167	n. sig.
E13	0,47	0,9762	0,0250	n. sig.
L02	0,06	0,9996	0,0500	n. sig.

Tab 5: Überprüfung der Verteilung der Itemantworten auf Abweichung von einer Normalverteilung durch χ^2 -Tests ($df=4$, α -Grundniveau=5 %)

Die Reihenfolge der Items in der Tabelle ist so gewählt, dass sich eine absteigende Abfolge der zugehörigen χ^2 -Werte bzw. eine aufsteigende Abfolge der p-Werte ergibt. Die Antworten des obersten Items L03 (SOFTWARE verhält sich so, wie ich es von anderen Programmen kenne.) weisen somit die deutlichsten Abweichungen von der Normalverteilung auf. Beim untersten Item L02 (Es ist leicht für mich ein geschickter Nutzer von SOFTWARE zu werden.) stimmt die Verteilung der Antworten am ehesten mit der Normalverteilung überein. Zur Veranschaulichung ist die Antwortverteilung für Item L03 in Abb. 27 gemeinsam mit der zugehörigen Normalverteilungsfunktion dargestellt. Alle Abweichungen von der Normalverteilung

sind statistisch nicht signifikant.

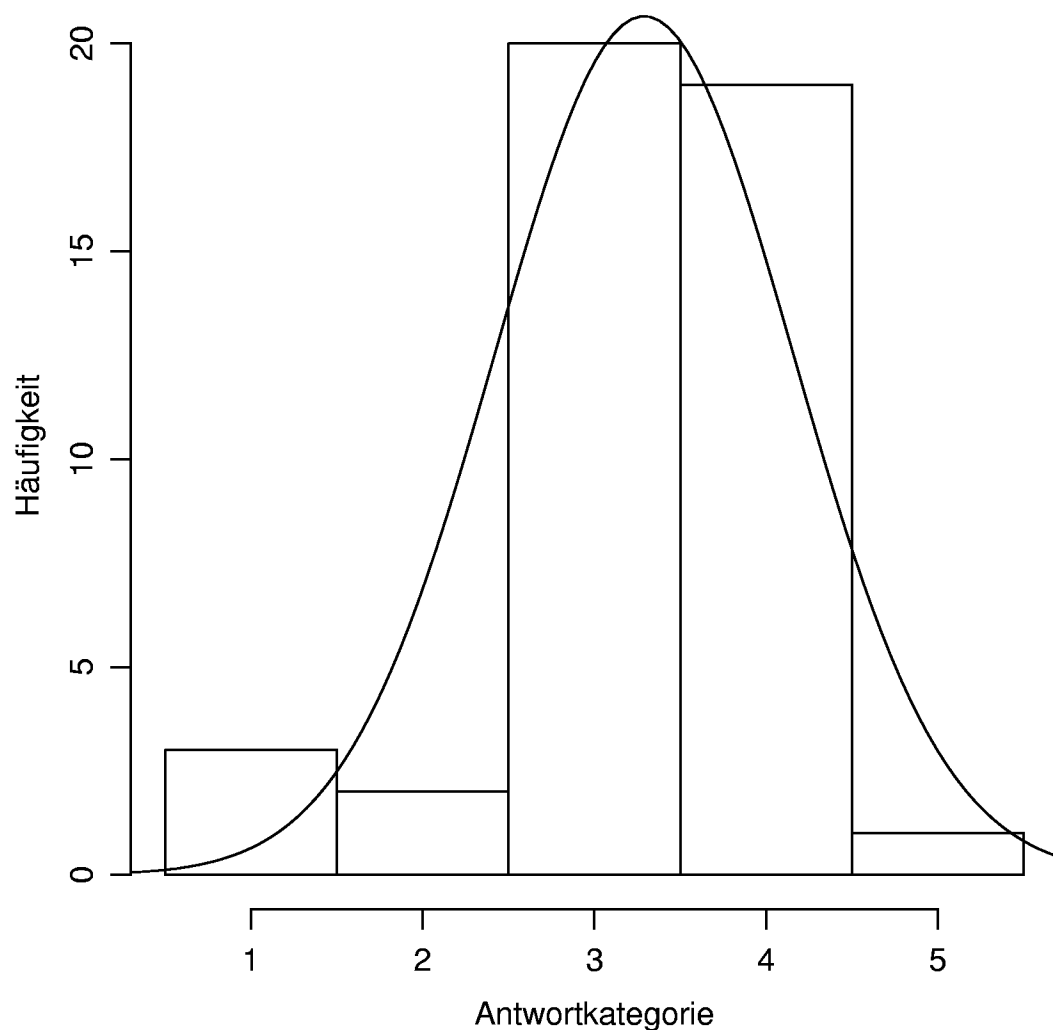


Abb. 27: Verteilung der Antworten für Item L03

Die erste Antwortkategorie ist deutlich überbesetzt, hier hat die Verteilung einen zweiten Gipfel. Die zweite Antwortkategorie ist unterbesetzt, ebenso wie die fünfte Antwortkategorie. Die Verteilung hat eine Schiefe von $-0,978$, ist also linksschief. Der Exzess der Verteilung beträgt $0,949$; die Verteilung ist also steilgipflig. Die Verteilungsformen der Antworthäufigkeiten der übrigen Items sind in Anhang H dargestellt.

Korrelationsmatrix der Items

Die Korrelationsmatrix gibt an, wie hoch die Produkt-Moment-Korrelation für jedes Itempaar ist. Da eine Korrelation einen ungerichteten Zusammenhang zwischen zwei Variablen beschreibt, ist die Korrelationsmatrix eine symmetrische Matrix. Die Korrelationsmatrix der Items ist in Tab 6 dargestellt.

	E01	E02	E03	E04	E07	E11	E12	E13	E14	E15	E16	E17	L01	L02	L03	L06	L07	L08	L09	L11	L12	L13	L15
E01		0,47	0,73	0,63	0,22	0,56	0,43	0,69	0,35	0,33	0,30	0,35	0,51	0,44	0,21	0,35	0,46	0,23	0,61	0,30	0,39	0,38	0,59
E02	0,47		0,46	0,40	0,43	0,53	0,39	0,67	0,41	0,31	0,31	0,33	0,22	0,07	0,06	0,08	0,37	0,07	0,27	-0,01	0,09	0,23	0,43
E03	0,73	0,46		0,74	0,31	0,44	0,54	0,57	0,39	0,44	0,20	0,32	0,64	0,52	0,24	0,51	0,51	0,24	0,47	0,40	0,54	0,57	0,71
E04	0,63	0,40	0,74		0,29	0,36	0,65	0,52	0,50	0,39	0,14	0,44	0,49	0,37	0,15	0,39	0,36	0,26	0,35	0,47	0,50	0,46	0,58
E07	0,22	0,43	0,31	0,29		0,47	0,27	0,39	0,68	0,44	0,12	0,51	0,39	0,14	0,05	0,26	0,25	-0,05	0,02	-0,01	0,04	0,08	0,30
E11	0,56	0,53	0,44	0,36	0,47		0,24	0,46	0,49	0,36	0,30	0,44	0,45	0,39	0,18	0,15	0,36	0,23	0,42	0,05	0,25	0,25	0,50
E12	0,43	0,39	0,54	0,65	0,27	0,24		0,54	0,51	0,35	0,32	0,47	0,21	0,20	-0,03	0,03	0,08	0,02	-0,04	0,11	0,06	0,05	0,30
E13	0,69	0,67	0,57	0,52	0,39	0,46	0,54		0,51	0,43	0,45	0,33	0,31	0,27	0,13	0,09	0,24	0,16	0,25	0,10	0,14	0,16	0,54
E14	0,35	0,41	0,39	0,50	0,68	0,49	0,51	0,51		0,48	0,23	0,53	0,35	0,30	0,26	0,27	0,25	0,17	0,00	0,15	0,03	0,08	0,31
E15	0,33	0,31	0,44	0,39	0,44	0,36	0,35	0,43	0,48		0,19	0,30	0,25	0,18	0,13	0,11	0,08	0,02	-0,07	0,00	0,03	0,14	0,37
E16	0,30	0,31	0,20	0,14	0,12	0,30	0,32	0,45	0,23	0,19		0,20	0,04	0,02	0,04	-0,02	-0,05	-0,06	0,19	-0,18	-0,08	-0,03	0,22
E17	0,35	0,33	0,32	0,44	0,51	0,44	0,47	0,33	0,53	0,30	0,20		0,26	0,23	-0,08	0,16	0,32	0,02	0,06	-0,05	0,05	0,15	0,16
L01	0,51	0,22	0,64	0,49	0,39	0,45	0,21	0,31	0,35	0,25	0,04	0,26		0,62	0,38	0,48	0,52	0,17	0,56	0,37	0,63	0,51	0,52
L02	0,44	0,07	0,52	0,37	0,14	0,39	0,20	0,27	0,30	0,18	0,02	0,23	0,62		0,51	0,35	0,30	0,35	0,44	0,43	0,47	0,56	0,42
L03	0,21	0,06	0,24	0,15	0,05	0,18	-0,03	0,13	0,26	0,13	0,04	-0,08	0,38	0,51		0,53	0,31	0,61	0,34	0,36	0,41	0,39	0,34
L06	0,35	0,08	0,51	0,39	0,26	0,15	0,03	0,09	0,27	0,11	-0,02	0,16	0,48	0,35	0,53		0,58	0,50	0,43	0,44	0,56	0,49	0,47
L07	0,46	0,37	0,51	0,36	0,25	0,36	0,08	0,24	0,25	0,08	-0,05	0,32	0,52	0,30	0,31	0,58		0,32	0,40	0,42	0,51	0,54	0,39
L08	0,23	0,07	0,24	0,26	-0,05	0,23	0,02	0,16	0,17	0,02	-0,06	0,02	0,17	0,35	0,61	0,50	0,32		0,26	0,35	0,38	0,38	0,34
L09	0,61	0,27	0,47	0,35	0,02	0,42	-0,04	0,25	0,00	-0,07	0,19	0,06	0,56	0,44	0,34	0,43	0,40	0,26		0,31	0,64	0,53	0,36
L11	0,30	-0,01	0,40	0,47	-0,01	0,05	0,11	0,10	0,15	0,00	-0,18	-0,05	0,37	0,43	0,36	0,44	0,42	0,35	0,31		0,57	0,47	0,40
L12	0,39	0,09	0,54	0,50	0,04	0,25	0,06	0,14	0,03	0,03	-0,08	0,05	0,63	0,47	0,41	0,56	0,51	0,38	0,64	0,57		0,68	0,51
L13	0,38	0,23	0,57	0,46	0,08	0,25	0,05	0,16	0,08	0,14	-0,03	0,15	0,51	0,56	0,39	0,49	0,54	0,38	0,53	0,47	0,68		0,52
L15	0,59	0,43	0,71	0,58	0,30	0,50	0,30	0,54	0,31	0,37	0,22	0,16	0,52	0,42	0,34	0,47	0,39	0,34	0,36	0,40	0,51	0,52	

Tab 6: Korrelationsmatrix der Einzelitems

Das Ziel bei der Betrachtung der Korrelationsmatrix bestand in der Identifizierung von für die Skala besonders geeigneten Items. Solche Items zeichnen sich nicht durch besonders hohe Korrelationen mit den anderen Items aus, da die Aufnahme eines Items mit hohen Korrelationen zu anderen Items nur wenig zusätzliche Information liefert (Yousfi, 2005). Nach Clark und Watson (1995) sollten die Korrelationen zwischen Items einer Skala im Bereich von 0,150 bis 0,500 liegen. Dabei sollten die einzelnen Inter-Itemkorrelationen nicht zu stark vom Mittelwert der Korrelationen abweichen. Entsprechend wurden für jedes Item der Mittelwert⁴ der Korrelationen mit den anderen Items und die Variationsbreite dieser Korrelationen berücksichtigt. Die Items sollen dabei so ausgewählt werden, dass ein „ruhiges, aber beharrliches Meer aus niedrigen, sehr ähnlichen Korrelationen“ (Green, 1978, S. 665-666, Übers. d. Autors) entsteht. In Tab 7 sind die niedrigste, die höchste und die mittlere Korrelation jedes Items mit Items desselben Konstrukts und mit Items des anderen Konstrukts dargestellt.

4 Da das Skalenniveau von Korrelationen eine direkte arithmetische Mittelung nicht sinnvoll zulässt, wurden die Korrelationen zunächst einer Fishers Z-Transformation unterzogen (Bortz, 1993, S. 201-202). Die arithmetisch gemittelten Werte wurden anschließend mittels der Umkehrfunktion zurück transformiert.

Item	eigenes Konstrukt			fremdes Konstrukt		
	Minimum	Maximum	Mittelwert	Minimum	Maximum	Mittelwert
L01	0,175	0,627	0,485	0,037	0,641	0,353
L02	0,299	0,619	0,450	0,018	0,517	0,266
L03	0,314	0,612	0,423	-0,081	0,264	0,113
L06	0,348	0,581	0,486	-0,021	0,515	0,203
L07	0,299	0,581	0,435	-0,050	0,506	0,275
L08	0,175	0,612	0,373	-0,059	0,257	0,109
L09	0,262	0,641	0,435	-0,068	0,607	0,224
L11	0,307	0,568	0,414	-0,182	0,474	0,117
L12	0,379	0,685	0,543	-0,082	0,544	0,180
L13	0,379	0,685	0,513	-0,031	0,570	0,220
L15	0,336	0,524	0,431	0,165	0,705	0,431
E01	0,222	0,725	0,478	0,209	0,607	0,414
E02	0,306	0,674	0,435	-0,014	0,426	0,173
E03	0,195	0,745	0,485	0,237	0,705	0,499
E04	0,138	0,745	0,478	0,152	0,577	0,403
E07	0,120	0,685	0,386	-0,052	0,392	0,136
E11	0,240	0,558	0,426	0,054	0,500	0,300
E12	0,240	0,652	0,436	-0,036	0,297	0,091
E13	0,332	0,688	0,515	0,090	0,537	0,222
E14	0,234	0,685	0,470	0,003	0,349	0,200
E15	0,191	0,483	0,368	-0,068	0,371	0,115
E16	0,120	0,450	0,252	-0,182	0,224	0,009
E17	0,200	0,533	0,388	-0,081	0,316	0,118

Tab 7: Variationsbreite und Mittelwert der Interitemkorrelationen für jedes Item

Für die Items des Konstrukts PL beträgt die niedrigste gemittelte Korrelation 0,373 bei Item L08. Die höchste gemittelte Korrelation gehört mit 0,543 zu Item L12. Die Items L01 und L08 korrelieren mit 0,175 am niedrigsten innerhalb dieses Konstrukts miteinander. Die höchste Korrelation tritt zwischen den Items L12 und L13 auf. Sie beträgt 0,685.

Von allen PEUEU-Items hat Item E16 die niedrigste gemittelte Korrelation zu den anderen Items dieses Konstrukts. Der gemittelte Korrelationskoeffizient beträgt für dieses Item 0,252. Der zweitniedrigste gemittelte Wert ist 0,368 für Item E15.

Der höchste Mittelwert ergibt sich für die Korrelationen von Item E13. Er beträgt 0,515. Die niedrigste Einzelkorrelation innerhalb des Konstrukts tritt mit 0,120 zwischen Item E07 und E16 auf. Die mit 0,745 höchste Einzelkorrelation gehört zu den Items E03 und E04.

Hinsichtlich der Skalenhomogenität sind Korrelationen innerhalb eines Konstrukts unter der Grenze von 0,15 problematisch (Clark und Watson, 1995). Die Items des Konstrukts PL korrelieren im Sinne der Skalenhomogenität alle ausreichend hoch miteinander. Von den Korrelationen der Items des Konstrukts PEUEU liegen zwei unterhalb dieser Schwelle. Neben der bereits erwähnten Korrelation der Items E07 und E16 korrelieren auch die Items E04 und E16 mit einem Wert von 0,138 schwächer als empfohlen.

Die empfohlene Obergrenze für Korrelationen zwischen einzelnen Items beträgt 0,500. Die Korrelationen mehrerer Itempaare überschreiten diesen Wert innerhalb beider Konstrukte. Aus den Items des Konstrukts PL korrelieren 20 Paare höher als empfohlen. Insgesamt werden innerhalb dieses Konstrukts 55 Korrelationswerte betrachtet. Von den Items des Konstrukts PEUEU korrelieren 17 Paare höher als 0,500. Hier werden insgesamt 66 Paare betrachtet.

Bestimmte Items sind an mehreren der zu hohen Korrelationen beteiligt. Von den PL-Items korreliert das Item L12 mit sieben anderen Items zu hoch, die beiden Items L01 und L13 mit jeweils sechs Items und Item L15 mit drei Items. Auf die übrigen Items entfallen maximal zwei zu hohe Korrelationen. Von den PEUEU-Items ist das Item E13 mit sechs Korrelationen und sind die Items E01, E03, E04, E12 und E14 mit vier Korrelationen an Korrelationen über 0,500 beteiligt. Die anderen Items haben maximal zwei zu hohe Korrelationen.

Zusätzlich sollten die Items der verschiedenen Konstrukte nicht hoch miteinander korrelieren. Campbell und Fiske (1959) fordern im Rahmen ihrer *Multitrait-Multimethod-Methode* eine höhere Korrelation zwischen den Messwerten unterschiedlicher Tests für dieselbe Eigenschaft als zwischen den Messwerten unterschiedlicher Tests für unterschiedliche Eigenschaften. Die Methode von Campbell und Fiske lässt sich nicht unverändert auf die Auswahl von Items übertragen, da sie für die Validitätsüberprüfung von Skalen entwickelt worden war. Die gemittelten Testwerte von Skalen, die Gegenstand des Artikels von Campbell und Fiske waren, sind robust gegenüber Ausreißern in einzelnen Items. Als Konsequenz können an Einzelitems nicht entsprechend rigorose Anforderungen wie an Skalenwerte gestellt werden. Korrelationen mit Items des anderen Konstrukts wurden im Folgenden ab einem Wert von 0,300 für Einzelkorrelationen bzw. für mittlere Korrelationen als problematisch betrachtet.

Insgesamt zeigten sich 42 Korrelationen zwischen je einem PL und einem PEUEU-Item, die über 0,300 lagen. Die Items E01, E03 und E04 waren jeweils an neun dieser Korrelationen beteiligt. Item L15 war an acht der Korrelationen über 0,300 beteiligt und Item L01 an sieben. Zusammen waren diese fünf Items an 36

der 42 zu hohen Korrelationen beteiligt. Auch bezüglich der gemittelten Korrelationen zu Items des jeweils anderen Konstrukts hatten diese fünf Items die fünf höchsten Werte (zwischen 0,353 und 0,499).

Von allen Items zeigte Item L03 die am deutlichsten von einer Normalverteilung abweichende Verteilung der Antworten. Das Item korrelierte mit 0,612 höher als empfohlen mit Item L08. In der vorherigen Itemselektion clusterte das Item auch entsprechend eng mit Item L08. Eine weitere Korrelation zu Item L06 lag mit 0,530 über 0,500. Aus der Kombination dieser Merkmale ergab sich, dass Item L08 besser für die Skala geeignet wäre. Da Item L03 keinen wesentlich über Item L08 hinaus gehenden Aspekt zur Skala beitragen würde, wurde das Item aus der Skala entfernt. Item L01 und L15 hatten innerhalb der Skala für das Konstrukt PL die meisten Korrelationen zu PEUEU-Items über 0,300. Für beide Items lag auch die gemittelte Korrelation zu den PEUEU-Items über dem Grenzwert von 0,300. Zusätzlich korrelierte Item L01 mit sechs und Item L15 mit drei der zehn übrigen PL-Items höher als 0,500. Als Konsequenz wurden die Items L01 und L15 aus der PL-Skala entfernt. L01 war in der vorherigen Itemselektion genau wie Item L02 Cluster B zugeordnet worden. Item L15 konnte keinem Cluster zugeordnet werden. Entsprechend ging durch den Ausschluss der beiden Items kein inhaltliches Cluster aus der Skala verloren.

Folgende Items zeigten für die PL-Skala geeignete Eigenschaften und verblieben in der Skala:

- L02: Es ist leicht für mich ein geschickter Nutzer von SOFTWARE zu werden.
- L06: Wie SOFTWARE funktioniert ist für mich leicht zu erkennen.
- L07: Die Funktion von allen Bedienelementen ist offensichtlich.
- L08: Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen.
- L09: Ich brauche keine besonderen Kenntnisse für die Nutzung von SOFTWARE.
- L11: Die Oberfläche von Software wirkt auf mich gut strukturiert.
- L12: Die Struktur von SOFTWARE ist für mich leicht zu erkennen.
- L13: Ich komme mit SOFTWARE auf Anhieb gut zurecht.

Zur Vervollständigung sind nachfolgend die aus der Skala entfernten Items aufgelistet:

- L01: Die Bedienung von SOFTWARE ist für mich leicht zu lernen.
- L03: SOFTWARE verhält sich so, wie ich es von anderen Programmen kenne.
- L15: Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE.

Von den PEUEU-Items waren die Items E01, E03 und E04 für den Verbleib in der Skala ungeeignet, da sie zu hohen Korrelationen innerhalb des eigenen Konstrukts und zu Items des Konstrukts PL aufwiesen. Item E16 korrelierte innerhalb des eigenen Konstrukts zu gering mit den anderen Items. Es hatte die niedrigste mittlere Korrelation und war an den beiden Korrelationen unter 0,150 beteiligt und wurde deshalb ebenfalls aus der Skala entfernt. Alle vier Items wurden aus der Skala entfernt. Da die vier Items in der ersten Itemselektion keinem inhaltlichen

Cluster zugeordnet werden konnten, ging durch ihren Ausschluss aus der Skala kein inhaltliches Cluster verloren.

Nach der zweiten Itemselektion sind folgende Items im Itempool für die PEUEU-Skala verblieben:

- E02: Wichtige Funktionen sind für mich in SOFTWARE schnell erreichbar.
- E07: Ich fühle mich durch SOFTWARE nicht bevormundet.
- E11: Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen.
- E12: Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar.
- E13: Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben.
- E14: Ich werde durch die Oberfläche von SOFTWARE nicht eingeschränkt.
- E15: Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen.
- E17: Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren.

Folgende Items wurden aus der Skala entfernt:

- E01: Mit SOFTWARE kann ich effizient arbeiten.
- E03: SOFTWARE passt zu meinem Arbeitsstil.
- E04: Die Bedienoberfläche von SOFTWARE ist für mich optimal gestaltet.
- E16: Ich bekomme von SOFTWARE keine für mich überflüssigen Hilfestellungen.

Im folgenden sind die Eckdaten der Korrelationsmatrix der beschriebenen Itemauswahl dargestellt. Die niedrigste gemittelte Korrelation innerhalb des Konstrukts PL beträgt 0,386 für das Item L08, die höchste gemittelte Korrelation 0,616 für das Item L12. Die niedrigste Einzelkorrelation tritt mit einem Wert von 0,254 zwischen den beiden Items L08 und L09 auf. Die höchste Einzelkorrelation beträgt 0,679 zwischen Item L12 und L13. Von den 28 Korrelationen liegen elf über dem empfohlenen Wert von 0,500. Keine der Korrelationen liegt unter der empfohlenen Grenze von 0,150.

Die niedrigste gemittelte Korrelation innerhalb der PEUEU-Items ist 0,415 für Item E15. Die höchste gemittelte Korrelation liegt bei Item E14 vor (0,594). Die niedrigste Einzelkorrelation beträgt 0,310 zwischen den Items E17 und E15. Die höchste Einzelkorrelation beträgt 0,696 zwischen Item E07 und Item E14. Keine der Einzelkorrelationen ist geringer als 0,150 und acht der Korrelationen liegen über 0,500.

Die Korrelationen zwischen den Items der verschiedenen Konstrukte liegen zwischen -0,074 und 0,413. Korrelationen zwischen Items unterschiedlicher Konstrukte sollten den Wert von 0,300 nicht überschreiten. Von den 64 Korrelationen zwischen Paaren aus Items unterschiedlicher Konstrukte überschreiten sieben diesen Wert. Die gemittelten Korrelationen, die robuster gegenüber einzelnen Ausreißern sind, bleiben alle unter 0,300.

Faktorenanalyse der Items

Um die inhaltliche Trennung der Skalen zusätzlich zu überprüfen, wurde eine Maximum-Likelihood-Faktorenanalyse über die Korrelationsmatrix der ausgewählten

Items gerechnet. Die Ladungsmatrix wurde nach dem Varimax-Kriterium rotiert, um die Interpretierbarkeit der Faktoren zu maximieren (Bortz, 1993, S. 507). Ein problematischer Vorgang bei der Faktorenanalyse ist die Festlegung der Anzahl bedeutsamer Faktoren. Es existieren unterschiedliche Methoden (Kaiser-Guttman-Kriterium, Scree-Test, Parallel-Analyse), die regelmäßig zu verschiedenen Ergebnissen führen. Alle Verfahren bewerten die Bedeutsamkeit eines Faktors anhand der Eigenwerte. Die ersten sieben Eigenwerte sind in Abb. 28 als Scree-Plot dargestellt.

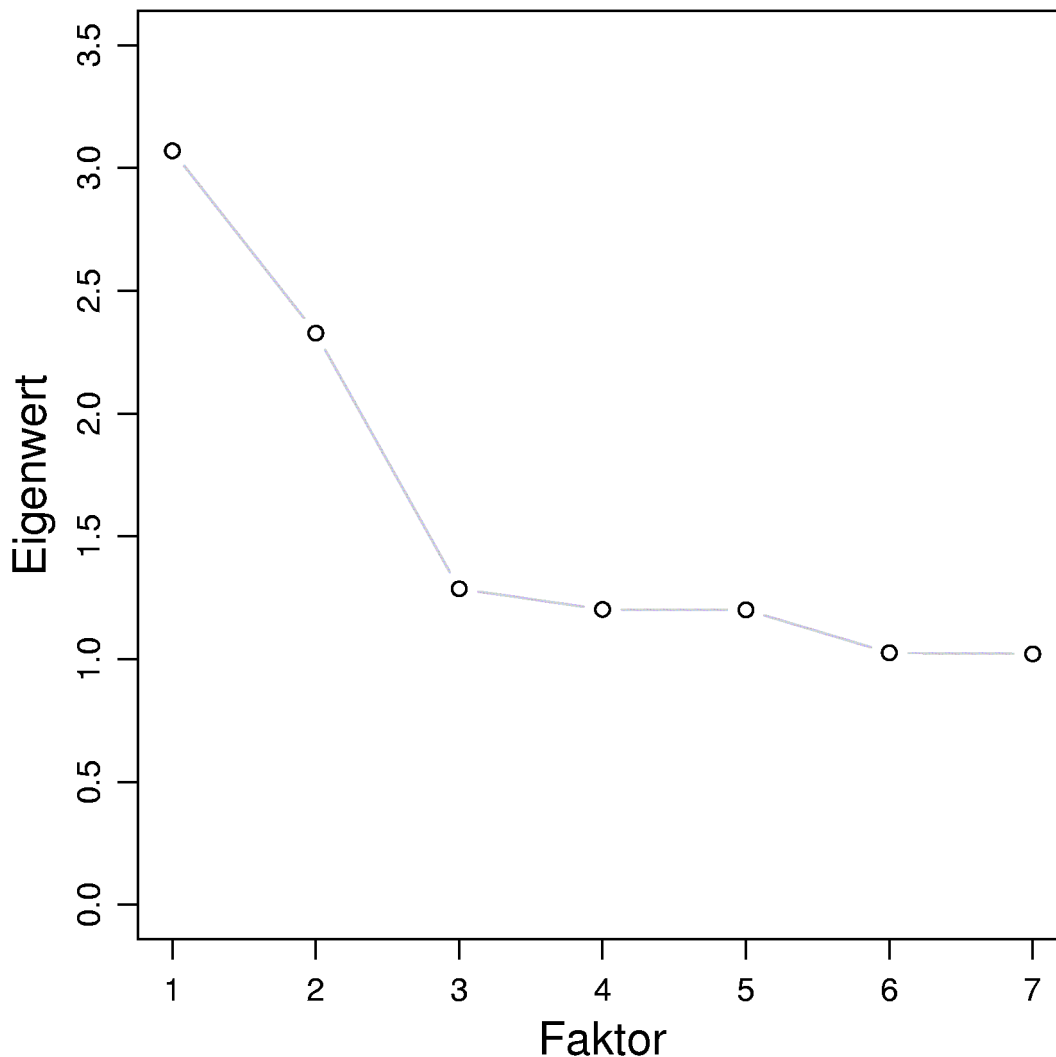


Abb. 28: Scree-Plot der Eigenwerte der Maximum-Likelihood-Faktorenanalyse

Nach dem Kaiser-Guttman-Kriterium werden alle Faktoren, deren Eigenwert größer als 1,00 ist, als bedeutsamer Faktor betrachtet. Allerdings weist Bortz (1993, S. 503) darauf hin, „dass die Anzahl der bedeutsamen Faktoren nach dieser Regel meistens überschätzt wird“. Beim sogenannten Scree-Test wird die Anzahl bedeutsamer Faktoren visuell anhand des Scree-Plots beurteilt. In einem Scree-Plot sind die Eigenwerte der Faktoren auf der Ordinate abgetragen, die Abszisse stellt die Faktoren in absteigender Reihenfolge ihrer Varianzaufklärung dar. Für die Festlegung der Anzahl bedeutsamer Faktoren wird in der Abbildung nach einem

Knick in der Kurve, die die Punkte verbindet, gesucht. In Abb. 28 ist ein solcher Knick beim dritten Faktor zu erkennen. Alle Faktoren vor dem Knick werden entsprechend der Scree-Methode als bedeutsam betrachtet. Die Antworten der Items können also durch zwei Faktoren sinnvoll beschrieben werden. Die beiden Faktoren erfassen zusammen 45,7 % der Varianz der Daten. Die Ladungen der Items auf die beiden Faktoren sind Tab 8 zu entnehmen.

Item	Faktor 1	Faktor 2
L02	0,236	0,566
L06	0,200	0,615
L07	0,241	0,580
L08	0,079	0,502
L09	0,029	0,676
L11	0,013	0,629
L12	0,032	0,826
L13	0,116	0,802
E02	0,615	0,156
E07	0,739	0,047
E11	0,610	0,311
E12	0,574	0,046
E13	0,684	0,178
E14	0,828	0,075
E15	0,580	0,063
E17	0,622	0,075

Tab 8: Ladungen der ausgewählten Items auf die beiden Faktoren

Die PL-Items laden auf dem ersten Faktor zwischen 0,013 und 0,241. Die Ladungen auf dem zweiten Faktor bewegen sich zwischen 0,502 und 0,826. Die PL-Items werden entsprechend hauptsächlich durch den zweiten Faktor repräsentiert, der entsprechend mit *Perceived Learnability* bezeichnet werden kann.

Die PEUEU-Items laden zwischen 0,574 und 0,828 auf dem ersten und zwischen 0,047 und 0,311 auf dem zweiten Faktor. Die PEUEU-Items werden hauptsächlich durch den ersten Faktor repräsentiert. Der erste Faktor entspricht deshalb *Perceived Ease of Use for Experienced Users*.

Die Faktorenanalyse bestätigt mit der Zuordnung der PL- bzw. PEUEU-Items zu zwei eigenen Faktoren die zugrundeliegende Annahme einer inhaltlichen Trennung der beiden Konstrukte. Die für die Skalen ausgewählten Items scheinen inhaltlich

homogen zu sein und sich ausreichend von der anderen Skala zu unterscheiden.

Ein gängiges Maß zur Beurteilung der Reliabilität einer Skala ist Cronbachs α (Steyer und Eid, 2001). Eine genaue Definition von Cronbachs α ist schwierig, der Begriff wird in der Literatur verschieden benutzt (Cortina, 1993). Unter der Annahme unkorrelierter Messfehler ist Cronbachs α eine untere Schranke für die Reliabilität einer Skala. Eine Schwäche von Cronbachs α besteht in der ausgeprägten Abhängigkeit von der Anzahl der Items einer Skala (Cortina, 1993). Generell steigt Cronbachs α mit der Anzahl der Items an. Nach Cortina wird die Reliabilität einer Skala im Allgemeinen ab einem Wert von 0,700 als brauchbar bewertet. Cronbachs α ist nicht für eine Bewertung der Homogenität einer Skala geeignet. Es kann allerdings bei homogenen Skalen genutzt werden, um Items mit einer – unerwünscht – hohen item-spezifischen Varianz zu entdecken.

Beim vorausgegangenen Prozess der Itemauswahl sollten Items mit einer ausgeprägten item-spezifischen Varianz schon erkannt und aussortiert worden sein. Die Werte für Cronbachs α bestätigen diese Annahme, sie liegen mit 0,861 für die PL-Skala und 0,864 deutlich über der Schwelle von 0,700. Die Reliabilität der Skalen ist damit für eine weitere Verwendung der Skalen ausreichend hoch.

Zusammenfassung der zweiten Itemauswahl

Bei der Itemselektion kamen statistische Verfahren zum Einsatz, die eine Normalverteilung der Antworten voraussetzen. Deshalb wurden die Antworten der Items zunächst auf durch χ^2 -Tests statistische signifikante Abweichungen von der Normalverteilung untersucht. Bei einer Korrektur des α -Niveaus nach Holm (1979) für die einzelnen Tests blieben alle Ergebnisse nicht signifikant.

Wichtigstes Kriterium für die Auswahl geeigneter Items waren die Korrelationen der Items untereinander. Insgesamt waren 253 Korrelationen zu berücksichtigen. Items, die für eine gemeinsame Skala geeignet sind, sollten zwischen 0,150 und 0,500 miteinander korrelieren. Innerhalb eines Konstrukts wurden alle Korrelationen, die außerhalb dieses Bereichs lagen, als problematisch betrachtet.

Nach Clark und Watson (1995) sollten die Korrelationen zwischen Items einer Skala im Bereich von 0.150 bis 0.500 liegen. 20 der 55 Korrelationen zwischen Items des Konstrukts PL lagen oberhalb der empfohlenen Schwelle von 0,500, keine Korrelation war niedriger als empfohlen. Von den Items des Konstrukts PEUEU korrelierten zwei zu niedrig und 17 höher als 0,500.

Im Hinblick auf die Skalenhomogenität sind hohe Korrelationen mit Items eines anderen theoretischen Konstrukts besonders problematisch. Alle Korrelationen zwischen je einem Item der beiden Konstrukte wurden ab einer Höhe von 0,300 als problematisch betrachtet. 42 von 132 solcher Korrelationen waren höher als 0,300. Es konnten insgesamt fünf Items (L01, L15, E01, E03, E04) identifiziert werden, die an 35 von diesen 42 unerwünschten Korrelationen beteiligt waren.

Diese fünf Items wurden aus den Skalen ausgeschlossen, um eine hohe Homogenität der Skalen zu gewährleisten. Zusätzlich wurde das Item L03 ausgeschlossen, da es eine ungünstige Verteilung der Antworten zeigte und zudem inhaltlich weitgehend mit Item L08 identisch war. Item L08 wies eine besser Verteilung der Antworten auf. Als letztes Item wurde E16 aus der Skala entfernt, da es zu geringe Korrelationen zu den anderen PEUEU-Items aufwies. Im Mittel korrelierte es nur 0,252 mit den anderen PEUEU-Items.

Über die Korrelationen der verbliebenen Items wurde eine Maximum-Likelihood-Faktorenanalyse durchgeführt. Anhand eines Scree-Plots wurden zwei bedeutsame Faktoren zur Beschreibung der Items identifiziert. Die Items der PL-Skala laden hauptsächlich auf dem zweiten Faktor, während die PEUEU-Items vorwiegend auf dem ersten Faktor laden. Somit bestätigt die Faktorenanalyse die erwünschte Trennung der beiden Skalen.

Abschließend wurde für beide Skalen Cronbachs α berechnet. Cronbachs α bildet bei homogenen Skalen mit unkorrelierten Messfehlern eine untere Schranke für die Reliabilität der Skala. Die Werte der beiden Skalen liegen mit 0,861 für die PL-Skala und 0,864 über der für Cronbachs α empfohlenen Schwelle von 0,700.

Die PL-Skala besteht nach der Itemselektion aus folgenden Items:

- L02: Es ist leicht für mich ein geschickter Nutzer von SOFTWARE zu werden.
- L06: Wie SOFTWARE funktioniert ist für mich leicht zu erkennen.
- L07: Die Funktion von allen Bedienelementen ist offensichtlich.
- L08: Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen.
- L09: Ich brauche keine besonderen Kenntnisse für die Nutzung von SOFTWARE.
- L11: Die Oberfläche von Software wirkt auf mich gut strukturiert.
- L12: Die Struktur von SOFTWARE ist für mich leicht zu erkennen.
- L13: Ich komme mit SOFTWARE auf Anhieb gut zurecht.

Die PEUEU-Skala setzt sich nach der Itemselektion aus folgenden Items zusammen:

- E02: Wichtige Funktionen sind für mich in SOFTWARE schnell erreichbar.
- E07: Ich fühle mich durch SOFTWARE nicht bevormundet.
- E11: Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen.
- E12: Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar.
- E13: Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben.
- E14: Ich werde durch die Oberfläche von SOFTWARE nicht eingeschränkt.
- E15: Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen.
- E17: Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren.

Empirische Überprüfung des Prozessmodells

Das Ziel der dritten empirischen Studie bestand darin, zu überprüfen, ob die Modellierung der Nutzungsintention durch die beiden neuen Konstrukte PL und

PEUEU signifikant gegenüber der Modellierung mit dem Konstrukt PEOU verbessert wurde. Die TAM-Konstrukte PU und PEOU wurden hierzu ebenso wie die neuen Konstrukte PL und PEUEU an einer Stichprobe von 84 Probanden gemessen. Zusätzlich wurde die Nutzungsintention erfasst. Der erste Schritt der Auswertung bestand in der statistischen Aufbereitung der Daten. Hierbei wurden die Eigenschaften der neuen Skalen nochmals überprüft. Zusätzlich zur Kontrolle der neuentwickelten Skalen diente dieser Schritt der Überprüfung der statistischen Voraussetzungen für eine parametrische Modellierung. Die Modellierung nach dem postulierten Prozessmodell wurde bezüglich der erreichten Anpassungsgüte mit einer Modellierung nach dem TAM verglichen. Als Kriterium für die Anpassungsgüte diente das *Bayesian Information Criterion* (BIC, Schwarz, 1978). Im Folgenden wird zunächst ein detaillierterer Überblick über die einzelnen Schritte gegeben.

Design der Studie

Die postulierte konzeptionelle Trennung von PL und PEUEU ergibt sich durch die unterschiedlichen Schwerpunkte bei der Wahrnehmung von Software durch unerfahrene bzw. erfahrene Benutzer. Eine Überprüfung dieser Annahme erforderte eine gezielte Manipulation der Erfahrung der Probanden hinsichtlich der verwendeten Software. Für die Studie wurde eine neue entwickelte Software eingesetzt, welche den Probanden unbekannt war. Es kam der Prototyp eines Online-Routenplaners mit integriertem Alertness-Management zum Einsatz. Mit Alertness-Management wird ein Ansatz zur Verringerung müdigkeitsbedingter Unfälle bezeichnet, bei dem Übermüdung vorgebeugt werden soll. Im Rahmen einer Routenplanung erstellt das Alertness-Management einen zeitlichen Verlauf für die Reise, der das Auftreten kritischer Übermüdung verhindern kann. Da bislang nur kurzfristig ein solcher Routenplaner mit Alertness-Management bei der Volkswagen AG online war, konnte vorausgesetzt werden, dass alle Probanden zu Beginn der Studie keine Erfahrung mit dieser Art von Software hatten. Durch das Bearbeiten von gestellten Aufgaben mit der Software sammelten die Probanden während der Untersuchung Erfahrung mit der Software, wurden also zu erfahrenen Benutzern. Die Erhebung der Einschätzung von unerfahrenen bzw. erfahrenen Benutzern erfolgte durch eine wiederholte Messung. Die erste Messung am Anfang der Untersuchung erhob die Eindrücke unerfahrener Benutzer. Die zweite Messung erhob am Ende der Untersuchung die Wahrnehmung der Software durch die nun erfahrenen Benutzer.

Für den Online-Routenplaner wurden zwei unterschiedliche Bedienoberflächen realisiert. Bei gleicher Funktionalität des Routenplaners unterschieden sich die beiden Bedienoberflächen im Grad der Benutzerführung. Es wurde erwartet, dass eine stärker ausgeprägte Benutzerführung durch unerfahrene Benutzer bevorzugt würde. Die Bedienoberfläche mit schwächerer Benutzerführung sollte entsprechend durch die erfahrenen Benutzer positiver bewertet werden. Da die Funktionalität des Online-Routenplaners ungeachtet der Bedienoberfläche gleich blieb, musste von einer ausgeprägten Übertragung von Lerneffekten an dem einen Bedienkonzept auf

das andere Bedienkonzept ausgegangen werden. Entsprechend konnte ein Proband nur als gleichermaßen unerfahren bzw. erfahren bezüglich beider Bedienkonzepte betrachtet werden. Als Konsequenz war es nicht sinnvoll möglich, Probanden mit beiden Bedienoberflächen zu konfrontieren. Die Unterschiede in der Wahrnehmung der Bedienkonzepte konnten also nicht abhängig gemessen werden. Statt dessen wurden die Probanden zufällig in zwei Gruppen eingeteilt, von denen die je eine ausschließlich mit einem der beiden Bedienkonzepte arbeitete.

Als Folge ergab sich ein Design, in dem teilweise wiederholt für denselben Probanden gemessene Werte, aber auch an unterschiedlichen Probanden gemessene Werte miteinander verglichen werden mussten. Die statistische Auswertung dieses Designs erfolgte im Rahmen einer sog. gemischten Modellierung.

Zunächst folgt eine Beschreibung des Software-Prototypen mit den beiden unterschiedlichen Bedienkonzepten. Anschließend wird das Design der Studie dargestellt und über die Durchführung berichtet.

Der Online-Routenplaner mit Alertness-Management

Das Konzept für die in dieser Arbeit eingesetzte Software wurde am Deutschen Zentrum für Luft- und Raumfahrt in der Arbeitsgruppe Verkehrssicherheit entwickelt. Das Ziel der Arbeitsgruppe Verkehrssicherheit besteht in einer Verringerung müdigkeitsbedingter Unfälle im Straßenverkehr durch ein sog. Alertness-Management. Alertness-Management ist hierbei die Bezeichnung für einen integrierten Ansatz, der versucht, Übermüdung im Straßenverkehr durch ein Bündel verschiedener Maßnahmen zu reduzieren. Online-Routenplaner eignen sich in diesem Kontext nach Einschätzung der Arbeitsgruppe, um die – im Gegensatz zu professionellen Fahrern – wenig organisierten Fahrer im privaten Personenverkehr zu erreichen. Die Integration eines Alertness-Management in einen Online-Routenplaner kann in dieser Zielgruppe Problembewusstsein hinsichtlich Übermüdung wecken und verstärken. Zusätzlich können auf diesem Wege, neben konkreten Vorschlägen für den zeitlichen Ablauf einer geplanten Fahrt, allgemeines Wissen über Ursachen von Müdigkeit und Strategien gegen Übermüdung vermittelt werden.

Müdigkeit hängt von verschiedenen Variablen ab, unter anderem der Tageszeit (zirkadiane Rhythmik) und der seit dem letzten Schlaf vergangenen Zeit. Zusätzlich haben verschiedene Tätigkeiten einen ermüdenden Effekt. Müdigkeit kann in engen Grenzen durch kurze Pausen reduziert werden, generell hilft allerdings nur ausreichender Schlaf gegen Müdigkeit.

In der Arbeitsgruppe Verkehrssicherheit wurde eine Software entwickelt, die die Vorhersage von Müdigkeit ermöglicht. Dabei hängt die Güte der Vorhersage direkt von den verfügbaren Informationen ab. Wegen der starken Abhängigkeit des Ermüdungsverlaufs von der Tageszeit ist eine Prognose der Ermüdung ohne Kenntnis der Tageszeit nicht sinnvoll. Die Güte der Vorhersage verbessert sich, wenn das konkrete Schlafverhalten bekannt ist, kann aber auch von der Software

geschätzt werden.

Für die Integration eines Alertness-Management in einen Online-Routenplaner muss zur Prognose der Ermüdung der absolute zeitliche Verlauf der Reise bekannt sein. Zusätzlich zum relativen Verlauf der Fahrzeit muss hierfür entweder die Abfahrts- oder die Ankunftszeit vorgegeben sein. Ist der zeitliche Verlauf entsprechend bekannt, kann die Software berechnen, ob die komplette Fahrt ohne ein Auftreten einer kritischen Übermüdung wie geplant durchgeführt werden kann. Ist dies nicht der Fall, werden Fahrtunterbrechungen wie beispielsweise in Abb. 29 vorgeschlagen, um eine Übermüdung des Fahrers zu verhindern. Eine detaillierte Erläuterung des Ausgabeformats des Routenplaners erfolgt anhand Abb. 33 auf Seite 144.




7km 06:07Uhr		Wechseln Sie am Autobahnkreuz Nordwestkreuz Frankfurt auf die A5 in Richtung Flughafen, Basel. Folgen Sie der A5 für 323km.
194km 08:00Uhr		Nutzen Sie die nächste Gelegenheit für eine kurze Pause (ca. 15 Min.)
331km 09:38Uhr		Sie passieren die Grenze zwischen Deutschland und der Schweiz.

Abb. 29: Pausenvorschlag in einem Routenplan des Prototypen

Für diese Arbeit wurde ein Prototyp eines Online-Routenplaners mit integriertem Alertness-Management erstellt. Eine vollständige Funktionalität war für die Untersuchung nicht notwendig. Die Programmierung einer Wegfindung basierend auf digitalem Kartenmaterial wäre zu aufwendig gewesen und war für die Untersuchung nicht notwendig. Da die zu planenden Routen in der Untersuchung vorgeben wurden, genügte ein Prototyp, der anhand der Benutzereingaben die entsprechende Route auswählen und einen zugehörigen Routenplan anzeigen konnte. Der Prototyp sollte – entsprechend einem vollständig implementierten Online-Routenplaner – in einem Browserfenster laufen. Die Bedienoberfläche wurde entsprechend in CSS-formatiertem HTML geschrieben und die Funktionalität der Routenauswahl und -darstellung in JavaScript realisiert. In der Untersuchung wurde der frei verfügbare Browser Firefox in der Version 2.0 zur Darstellung des Prototypen eingesetzt. Die Funktionalität des Prototypen wurde nur auf diesem Browser getestet.

Für den Prototypen wurden zwei verschiedene Bedienkonzepte entwickelt. Die Funktionalität des Prototypen blieb unabhängig vom gewählten Bedienkonzept gleich. Die Bedienkonzepte unterschieden sich hinsichtlich der Flexibilität der Interaktionsmöglichkeiten und der Benutzerführung. Während die eine Variante der Bedienoberfläche eine flexible Interaktion auf Kosten einer geringen Benutzerführung erlaubte, maximierte die andere Oberfläche die Benutzerführung und schränkte

hierfür die Interaktionsmöglichkeiten ein. Abb. 30 und Abb. 34 zeigen Screenshots der beiden Bedienoberflächen des Prototypen, wie sie im Browser dargestellt wurden. Zunächst wird das Bedienkonzept A näher beschrieben (Abb. 30).

Abb. 30: Bedienkonzept A mit schwacher Benutzerführung und hoher Flexibilität

Innerhalb des Bedienkonzeptes A konnten alle Eingaben in beliebiger Reihenfolge vorgenommen werden. Die Adresseingabe erfolgte in drei Formularfeldern je Adresse. Straßename und Hausnummer konnten in beliebiger Abfolge in die erste Zeile eingegeben werden, Postleitzahl und Ortsname ebenfalls in beliebiger Abfolge in die zweite Zeile. Hausnummern wurden in der Routenplanung generell nicht berücksichtigt. Für die Spezifikation des Ortes waren entweder die Postleitzahl oder der Ortsname ausreichend. Die Angabe des Landes erfolgte durch ein Listenauswahlfeld, in welchem Deutschland vorausgewählt war. Die Anordnung der Felder folgte hiermit dem gängigem Schema zur Angabe von Adressen, wie es z. B. auch auf Briefen oder in den meisten Formularen zur Anwendung kommt.

Unterhalb der beiden Eingabebereiche für die Start- bzw. die Zieladresse folgte der für das Alertness-Management notwendige Bereich zur Eingabe der Abfahrts- oder Ankunftszeit. Die Eingabe der Abfahrts- oder Ankunftszeit erfolgte in zwei Listenauswahlfeldern für Stunden und Minuten. Über sog.

Radiobuttons (Mehrfachoptionsfeld⁵) wird ausgewählt, ob die eingegebene Zeit als Abfahrts- oder Ankunftszeit interpretiert werden soll.

Unterhalb der Zeiteingabe können optional gewünschte Zwischenstationen eingegeben werden. Die Eingabe erfolgt genau wie für die Start- bzw. Zieladresse. Ganz unten stehen dem Benutzer zwei durch Icons gekennzeichnete Buttons zur Verfügung. Die Icons sollten gezielt eine gewisse Unklarheit bezüglich der Funktion der Schaltflächen erzeugen. Die linke, mit einem „X“ gekennzeichnete Schaltfläche diente zum Zurücksetzen der Eingabemaske, verwarf also gemachten Eingaben ohne weitere Bearbeitung. Die rechte, mit einem nach rechts weisenden Pfeil markierte Schaltfläche löste die Bearbeitung der getätigten Eingaben aus.

Falls unklare, widersprüchliche oder nicht bekannte Adressen eingegeben wurden, wurde dies durch einen roten Hinweistext unterhalb des entsprechenden Eingabebereichs angezeigt. Abb. 31 zeigt zwei mögliche Fehlermeldungen.

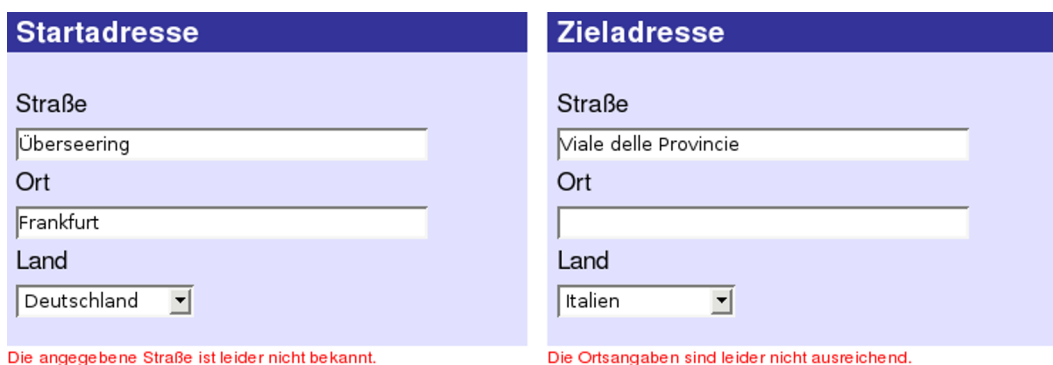


Abb. 31: Fehlermeldung in Bedienkonzept A

Bei der Startadresse wurde die Stadt korrekt erkannt, der Straßename *Überseering* ist allerdings in Frankfurt nicht bekannt. Die Zieladresse ist unvollständig eingegeben und wird entsprechend nicht erkannt. Die Fehlermeldungen unterbrechen den Bedienablauf nicht, sie erforderten keine gesonderte Reaktion des Benutzers, wie beispielsweise eine Bestätigung durch Klicken auf eine OK-Schaltfläche. Nach einer Korrektur der Fehler konnte eine erneute Routenberechnung durch Betätigen des entsprechenden Buttons ausgelöst werden.

Abb. 32 zeigt als Beispiel den Anfang der Ausgabe einer geplanten Route von Frankfurt nach Rom. Die Ausgabe der Ergebnisse begann mit einem Kopfteil, in dem Informationen zur geplanten Route zusammengefasst wurden. Die eingegebenen Daten wurden hier wiederholt und um errechnete Werte wie die voraussichtliche Fahrtzeit und die entsprechende Ankunftszeit ergänzt. Unterhalb des Kopfteils war eine Übersichtskarte der geplanten Route angeordnet.

5 In einem Mehrfachoptionsfeld kann immer genau eine Auswahl aus einer Anzahl von Möglichkeiten ausgewählt werden. Die englische Bezeichnung stammt von der Analogie zum Verhalten der mechanischen Tasten zur Auswahl des Empfangskanals alter Radiogeräte, mit denen ebenfalls immer nur ein Sender ausgewählt werden konnte.

Berechnete Route	
Start	Ziel
Glauburgstraße 60318 Frankfurt Deutschland	Viale delle Provincie 00162 Rom Italien
Abfahrtszeit 06:00 Uhr	Ankunftszeit (inklusive Pausen) 20:24 Uhr
Streckenlänge 1253 km	
Fahrzeit ohne Pausen 12 Std. 54 Min.	
Fahrzeit mit empfohlenen Pausen 14 Std. 24 Min.	

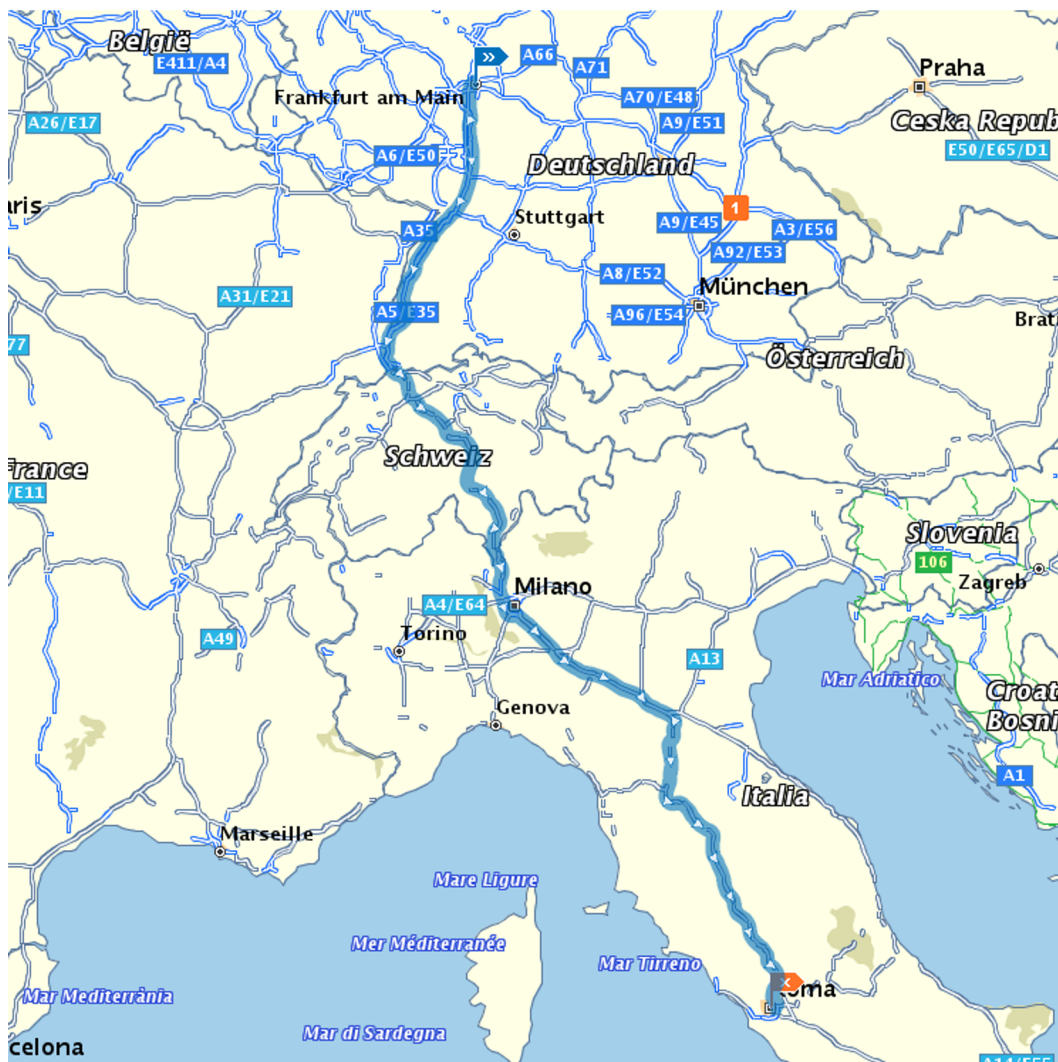









Abb. 32: Beispielausgabe des Routenplaners (Kopfteil mit Karte)


Im Kopfteil der Ausgabe wurden die Abfahrts- und die Zieladresse wiederholt. Darunter waren die Angaben zur Abfahrts- und Ankunftszeit angeordnet. Die

Ankunftszeit wurde stets mit den eingeplanten Pausen berechnet. Abschließend wurden in der Zusammenfassung der Eckdaten der Route die Fahrstrecke, sowie die Fahrzeit mit und ohne Pausen aufgelistet.

Unterhalb des Kopfteils wurden die Anweisungen der Routenbeschreibung aufgelistet. Abb. 33 zeigt den unteren Teil der Routenbeschreibung für die Fahrt von Frankfurt nach Rom.

1233km 20:05Uhr		Wechseln Sie auf die A90 in Richtung L'Aquila, Teramo, Tangentiale Est. Folgen Sie der A90 für 9,6km.
1243km 20:14Uhr		Wechseln Sie auf die A24 in Richtung Roma, Tangentiale Est. Folgen Sie der A24 für 6,5km.
1250km 23:19Uhr		Verlassen Sie an der Ausfahrt Portonaccio die A24.
1250km 23:19Uhr		Sie passieren jetzt die Ortseinfahrt von Rom.
1250km 23:19Uhr		Biegen Sie rechts ab in die Via Galla Placidia. Folgen Sie dem Straßenverlauf für 1,2km.
1251km 23:21Uhr		Biegen Sie links ab in die Via Tiburtina. Folgen Sie dem Straßenverlauf für 1,8km.
1253km 23:24Uhr		Biegen Sie rechts ab in die Viale delle Provincie. Sie haben Ihr Fahrziel erreicht.








Abb. 33: Beispielausgabe des Routenplaners (Route) für Bedienkonzept A

Die Routenbeschreibung war in drei Spalten gegliedert. In der ersten Spalte wurde die seit der Abfahrt zurückgelegte Strecke sowie die Uhrzeit, zu der die entsprechende Stelle passiert werden würde, angegeben. In der zweiten Stelle wurde durch Icons graphisch die wichtigste Information zum Routenverlauf an der entsprechenden Stelle dargestellt. In der dritten Spalte wurden die Informationen zum Routenverlauf ausführlicher in Textform gegeben.

Unter dem Ende der Routenbeschreibung waren drei Schaltflächen angeordnet. Im Bedienkonzept A waren diese Schaltflächen nur durch Icons gekennzeichnet. Die erste Schaltfläche war mit einem „X“ markiert und eine Wiederholung der Schaltfläche aus der Eingabemaske des Routenplaners. Sie löste einen Rücksprung zur Eingabemaske des Routenplaners aus und löschte vorhandene Eingaben. Das Icon der zweiten Schaltfläche zeigte die beiden Flaggen, die in der Übersichtskarte den Start und den Zielort markierten. Dazwischen waren zwei Pfeile, von denen der obere von links nach rechts, der untere in die umgekehrte Richtung zeigten. Diese Schaltfläche sprang ebenfalls zur Eingabemaske des Routenplaners zurück und vertauschte die vorhandenen Eingaben der Abfahrts- und der Zieladresse. Mit dieser

Schaltfläche konnte eine Rückreise zu einer bereits ausgegebenen Route geplant werden. Die dritte Schaltfläche war mit einem Stift als Icon versehen. Mit ihr konnte zum Editieren in die Eingabemaske zurückgesprungen werden, wobei die vorhandenen Eingaben nicht gelöscht wurden, sondern zum Bearbeiten in der Eingabemaske stehen blieben.

Im Gegensatz zu den Bedienkonzepten existierender Online-Routenplaner (wie z. B. www.mappy.de, www.reiseplanung.de oder www.viamichelin.de) blieb im Bedienkonzept A die Eingabemaske mit den eingegebenen Daten zugänglich. Die Eingabemaske blieb immer am Kopf der Seite, so dass jederzeit Änderungen oder Korrekturen an den Eingaben vorgenommen werden konnten. Bedienkonzept A erfüllte hierdurch die Forderungen der EN ISO 9241-10 nach Steuerbarkeit des Dialogverlaufs sowie nach Fehlertoleranz besser als existierende Online-Routenplaner. Die Kennzeichnung der Schaltflächen ausschließlich durch Icons war allerdings hinsichtlich der Selbstbeschreibungsfähigkeit der Bedienoberfläche bewusst nicht optimal gestaltet. Mit dieser Auslegung wurde von Bedienkonzept A eine nicht optimale Erlernbarkeit aufgrund der mangelnden Selbstbeschreibungsfähigkeit erwartet. Andererseits sollte Bedienkonzept A durch die konsequente Umsetzung eines vom Nutzer steuerbaren Dialogverlaufs sowie den Möglichkeiten zur Fehlerkorrektur Vorteile für erfahrene Nutzer bieten.

Abb. 34 zeigt die Eingabemaske von Bedienkonzept B. Dieses Bedienkonzept sollte durch eine stärkere Benutzerführung eine bessere Erlernbarkeit bieten. Gleichwohl wurde eine schlechtere Benutzbarkeit durch erfahrene Benutzer erwartet.

The image shows a user interface for a routing application. It consists of three main sections:

- Startadresse:** A header bar with a dark blue background and white text. Below it, a light blue input area contains five fields: 'PLZ' (a small square), 'Ort' (a wide text box), 'Straße' (a wide text box), 'Nr.' (a small square), and 'Land' (a dropdown menu).
- Zieladresse:** An identical header and input area to the 'Startadresse' section.
- Buttons:** Three dark blue buttons with white text are positioned below the input areas: 'Neue Route', 'Station einfügen', and 'Route berechnen'.

Abb. 34: Bedienkonzept B mit starker Benutzerführung und geringer Flexibilität

Die Adresseingabe erfolgte in Bedienkonzept B in jeweils fünf unterschiedliche Felder je Adresse. Postleitzahl, Ort, Straßename, Hausnummer und Land mussten jeweils separat eingegeben werden. Damit war für den Benutzer genau zu erkennen, in welcher Form die Adresseingabe zu erfolgen hatte. Allerdings entstand auch zusätzlicher Aufwand durch das notwendige Wechseln der Formularfelder. Die getrennte Eingabe von Postleitzahl und Ortsname bzw. Straßename und Hausnummer war technisch nicht notwendig, da die Auswertung der Eingaben durch

denselben Algorithmus erfolgte wie für Bedienkonzept A. Entsprechend war für Bedienkonzept B die Angabe der Postleitzahl oder des Ortsnamens ausreichend. Auch in Bedienkonzept B wurde die Hausnummer bei der Routenplanung nicht berücksichtigt.⁶

Die Abfolge der Eingabefelder (Postleitzahl, Ortsname, Straßename, Hausnummer und Land) war abweichend von der üblichen Darstellung von Adressen gewählt. In einer nicht publizierten, früheren Studie unserer Arbeitsgruppe zur Usability von Online-Routenplanern wurde beobachtet, dass dies häufig zu Fehleingaben führte, insbesondere wenn die Adressen schriftlich in der üblichen Abfolge (Straßenname mit Hausnummer, Postleitzahl mit Ortsname und darunter der Landesname) dargestellt waren.

Im Listenauswahlfeld zur Auswahl des Landes war keine Vorauswahl getroffen, so dass die Benutzer hier für jede Adresse manuell eine Eingabe vornehmen mussten. Des Weiteren war eine Eingabe von Zwischenstationen direkt nicht möglich. Hierfür musste der Benutzer zunächst durch Betätigen der „Station einfügen“-Schaltfläche eine weitere Adresseingabemöglichkeit anfordern.

Unterhalb des Eingabebereichs waren drei beschriftete Schaltflächen angeordnet. Die erste Schaltfläche (*Neue Route*) löschte vorhandene Formulareingaben. Die zweite Schaltfläche (*Station einfügen*) deckte ein weiteres Eingabefeld für die Adresseingabe eines Zwischenstopps auf. Dieses Eingabefeld wurde zwischen den Eingabefeldern für die Abfahrts- und die Zieladresse platziert. Die dritte Schaltfläche (*Route berechnen*) leitete den Benutzer weiter zur Zeitplanung.

Konnten die eingegebenen Adressen bekannten Routen zugeordnet werden, wurden die Adressinformationen zusammengefasst und eine Eingabemaske zur Eingabe der Zeitplanung präsentiert. Dies ist exemplarisch in Abb. 35 zu sehen.

⁶ Der Prototyp folgt mit der – unsinnigen – Abfrage der Hausnummer existierenden Online-Routenplanern, die vielfach diese Information erfragen, anschließend aber nicht berücksichtigen.

Route	
Start Glauburgstraße 60318 Frankfurt Deutschland	Ziel Viale delle Provincie 00162 Rom Italien

Zeitplanung	
<input checked="" type="radio"/> Abfahrt	<input type="text" value="06"/> : <input type="text" value="00"/> Uhr
<input type="radio"/> Ankunft	

Zurück	Route berechnen
---------------	------------------------

Abb. 35: Eingabemaske für die Zeitplanung in Bedienkonzept B

Im Kopfteil des Eingabefensters wurden die Abfahrts- und die Zieladresse wiederholt. Darunter war der Eingabebereich für die Zeitplanung. Über zwei Radiobuttons konnte zwischen der Eingabe einer Abfahrts- oder Ankunftszeit ausgewählt werden. Die Zeiteingabe selbst erfolgte über zwei Listenauswahlfelder. Im Unterschied zu Bedienkonzept A war sowohl bezüglich der Art der Zeiteingabe als auch in den beiden Listefeldern keine Vorauswahl getroffen. Der Routenplaner akzeptierte keine fehlenden Zeitangaben, weshalb die Benutzer in alle Eingabebereiche Werte eingeben mussten.

Unterhalb des Eingabebereichs waren zwei Schaltflächen. Die erste Schaltfläche (*Zurück*) führte zurück zur Eingabemaske für die Adresseingabe. Die zweite Schaltfläche (*Route berechnen*) löste die Planung der Route aus.

Auf fehlerhafte Eingaben wurde in Bedienkonzept B in einem eigenen Fenster aufmerksam gemacht. Eine solche Fehlermeldung ist in Abb. 36 zu sehen.

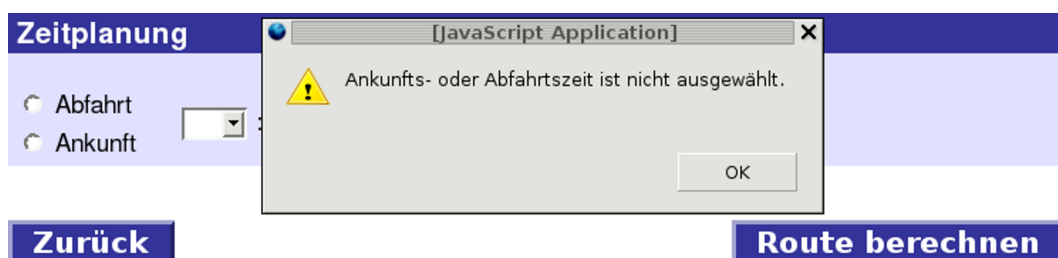


Abb. 36: Fehlermeldung der Zeitplanung von Bedienkonzept B

Diese Art der Fehlermeldung war auffälliger. Allerdings musste sie von den Benutzern bestätigt werden, bevor eine Korrektur des Fehlers möglich war. Auch konnte in dem Fenster nur auf genau einen Fehler hingewiesen werden. Dies war der erste bei der Überprüfung der Daten gefundene Fehler. Im Falle einer Eingabe mit mehreren Fehlern wurde der Benutzer zunächst nur auf einen Fehler

hingewiesen, erst nach dessen Korrektur bekam er den zweiten Fehler gemeldet. Zur Verdeutlichung dient Abb. 37, die eine zweite Fehlermeldung derselben Zeiteingabemaske zeigt, nachdem die Art der Zeitangabe ausgewählt worden ist, aber keine Zeit eingegeben wurde.

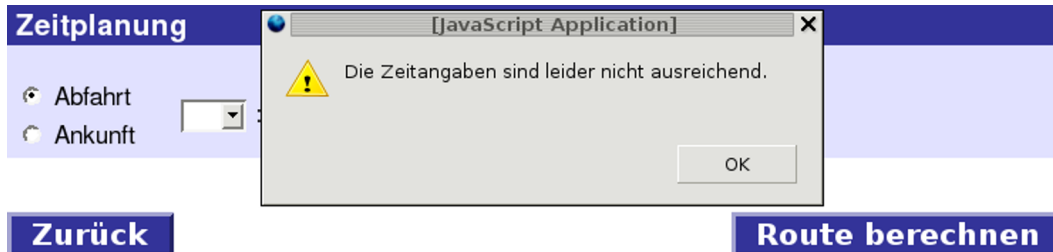


Abb. 37: Weitere Fehlermeldung der Zeitplanung von Bedienkonzept B

Die Ausgabe der Route erfolgte genau wie für Bedienkonzept A. Auch die ausgegebenen Routen an sich waren identisch, ebenso die Pausenvorschläge des Alertness-Managements. Funktional unterschieden sich der Prototyp in den beiden Bedienkonzept-Varianten also nicht. Allerdings nutzte Bedienkonzept B für die Darstellung der Route eine neue Seite, so dass vorherige Eingaben nicht mehr zugänglich waren. Auch die Schaltflächen unterhalb der Routenausgabe unterschieden sich von Bedienkonzept A, wie in Abb. 38 zu erkennen ist.

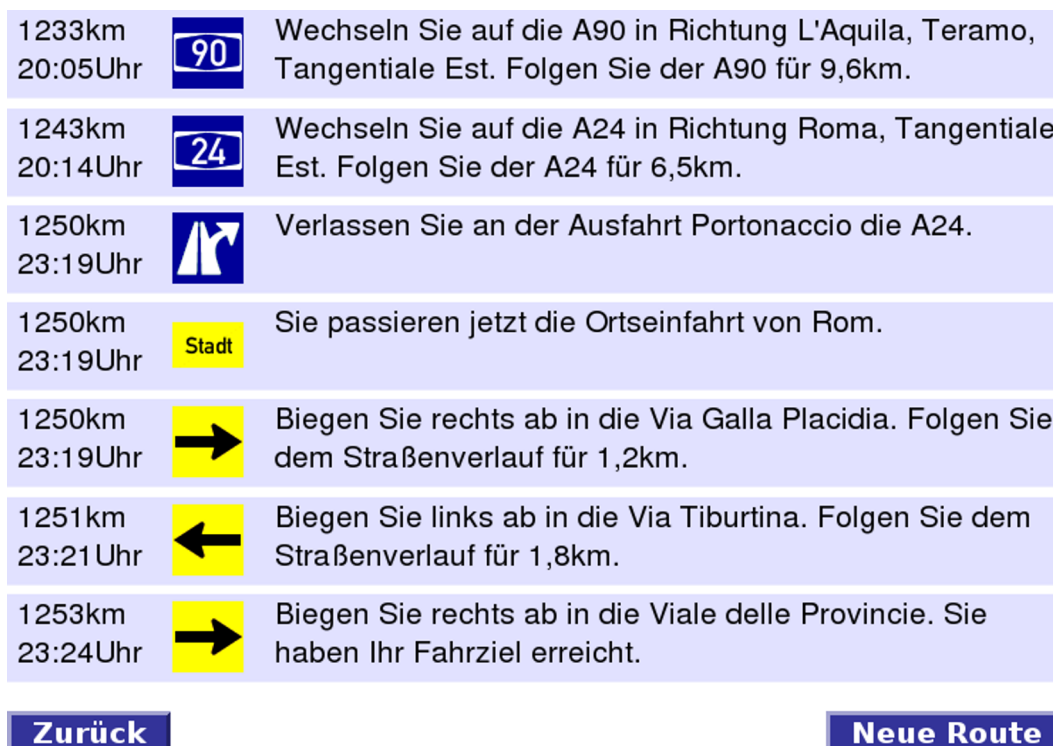


Abb. 38: Beispielausgabe des Routenplaners (Route) für Bedienkonzept B

Die erste Schaltfläche (*Zurück*) führte die Benutzer zurück zur Zeitplanung. Die zweite Schaltfläche (*Neue Route*) startete eine neue Routenplanung, wobei die

getätigten Eingaben verworfen wurden.

Alle Routen, die der Prototyp darstellen konnte, sind in Anhang I aufgelistet. Insgesamt konnte der Prototyp 75 Routen darstellen. In den 75 Routen waren auch die Rückreisen enthalten, die der Prototyp zu jeder Route anzeigen konnte. Routen mit einer Zwischenstation konnte der Prototyp auch als zwei Routen, also zunächst von der Startadresse bis zur Zwischenstation und dann von der Zwischenstation zur Zieladresse anzeigen. Die 75 verfügbaren Strecken ergaben sich aus 17 Routen ohne und 8 Routen mit Zwischenstopp. Für alle längeren Routen waren mehrere Pausenpläne im Prototypen gespeichert. Die Pausenpläne unterschieden sich je nach Abfahrtszeit, weil der Müdigkeitsverlauf während einer Fahrt von der Tageszeit abhängig ist.

Untersuchung

Die Skalen wurden im Rahmen einer Studie zweimal erhoben. Das Konstrukt PU wurde mit der Übersetzung der sechs Items aus Davis (1989) Studie erhoben. Die Messung des Konstrukts PEOU erfolgte durch Zusammenfassung der Skalen für PL und PEUEU. Die Nutzungsintention wurde mit zwei hierfür formulierten Items erfasst.

Um Reihenfolgeeffekte durch die Abfolge der Items zu vermeiden, wurden drei Varianten des Fragebogens mit unterschiedlicher Abfolge der Items eingesetzt. Jede dieser Abfolgen war zufällig. Die Zuordnung der Fragebogenvariante zu den Probanden erfolgte ebenfalls zufällig. Eine Fragebogenversion ist in Anhang J dargestellt. Nachfolgend sind die drei Zufallsreihenfolgen der Fragebogenvarianten aufgelistet. Dabei sind die Item-Konstrukte abgekürzt, wobei PU-Items durch ein „U“, PL-Items durch ein „L“ und PEUEU-Items durch ein „E“ kodiert sind. Die beiden Items zur Erfassung der Nutzungsintention sind durch ein „I“ gekennzeichnet:

- L09, U01, E07, E15, I01, L08, L06, L07, E02, U05, PE17, U03, L11, I02, L13, E11, L12, E12, U06, E13, U04, E14, L02, U02
- U01, E12, U05, L08, U02, L13, U03, E15, E17, U06, L12, E02, I01, L06, E14, L02, E11, U04, L09, E07, E13, I02, L07, L11
- U06, E12, E14, E11, E17, L06, I02, L13, L08, L12, L09, U04, L02, I01, U05, U02, E07, E13, L07, U03, E15, E02, U01, L11

Die Probanden wurden zufällig in zwei Gruppen eingeteilt, die jeweils mit einem der beiden Bedienkonzepte arbeiteten. Jeder Proband arbeitete während der gesamten Studie ausschließlich mit einer der beiden Bedienoberflächen. Ein Wechsel zwischen den Bedienoberflächen war nicht sinnvoll möglich, da erwartet werden musste, dass es zu einer nicht kontrollierbaren Übertragung von Erfahrungen mit der einen Bedienoberfläche auf die andere gekommen wäre.

Den Probanden wurde der Ablauf der Studie erläutert, dann füllten alle Probanden einen allgemeinen Fragebogen zur Person aus. Im Anschluss sollten sie die erste von insgesamt 19 Routen mit dem Prototypen planen. Um eine

Bearbeitung der Aufgaben sicherzustellen und um den Probanden einen plausiblen Grund für die Studie zu geben, sollten die Probanden zu jeder geplanten Route eine Kontrollfrage beantworten. Die zu planenden Routen und die zugehörige Kontrollfrage wurden den Probanden auf Karten ausgehändigt. In Abb. 39 ist die Karte mit der ersten Aufgabe abgebildet. Eine vollständige Auflistung der Planungsaufgaben mit den zugehörigen Kontrollfragen befindet sich in Anhang K.

Aufgabe 1

Sie wollen nach der Uni Ihr Auto in die Autowerkstatt bringen. Bitte planen Sie, wie Sie zum Autohaus Krull fahren.

Startort:
Wilschenbrucher Weg 84
21335 Lüneburg

Zielort:
Hamburger Str. 37
21339 Lüneburg

Frage: Wie lange dauert die Fahrt laut Routenplaner?

Abb. 39: Karte mit der ersten Planungsaufgabe und einer Kontrollfrage

Für die Beantwortung der Kontrollfragen stand den Probanden ein eigener Fragebogen zur Verfügung. Auf diesem wurde die Kontrollfrage wiederholt und konnte darunter beantwortet werden. Es erfolgte keine inhaltliche Auswertung der Antworten, da diesbezüglich keine Hypothesen existierten. Allerdings wurde die Anzahl der bearbeiteten Aufgaben – ungeachtet der Richtigkeit der Antworten – ausgezählt.

Nach der Bearbeitung der ersten Aufgabe und Eintrag der Antwort auf die Kontrollfrage füllten die Probanden einen Fragebogen mit den Skalen für PU, PL, PEUEU und der Nutzungsintention aus. Zu diesem Messzeitpunkt wurde ein möglichst früher Eindruck des Prototypen erhoben. Eine frühere Messung mit geringerer Interaktion erschien sinnvoll nicht möglich.

Nach dem Ausfüllen der Skalen sollten die Probanden so viele Aufgaben wie möglich innerhalb von 30 Minuten bearbeiten. Dabei sollte die Reihenfolge der Aufgaben eingehalten werden, ein Überspringen von Aufgaben war also nicht zulässig. Die 18 Aufgaben waren teilweise komplexer als die erste Aufgabe und erforderten eine detaillierte Beschäftigung mit den Ausgaben des Routenplaners. Die Aufgaben mit den zugehörigen Kontrollfragen sind in Anhang J vollständig aufgelistet.

Nach der 30-minütigen Arbeitsphase bewerteten die Probanden den Prototypen ein zweites Mal mit denselben Skalen. Die Erfahrung der Probanden im Hinblick auf den Routenplaner-Prototypen wurde nicht gemessen, da hierfür keine geeigneten

Möglichkeiten existierten. Vielmehr wurde angenommen, dass die Arbeitsphase ausreichend lang war, um mit der vorhandenen Funktionalität des Prototypen vertraut zu werden. Der Ablauf der Erhebung ist in Abb. 40 dargestellt.

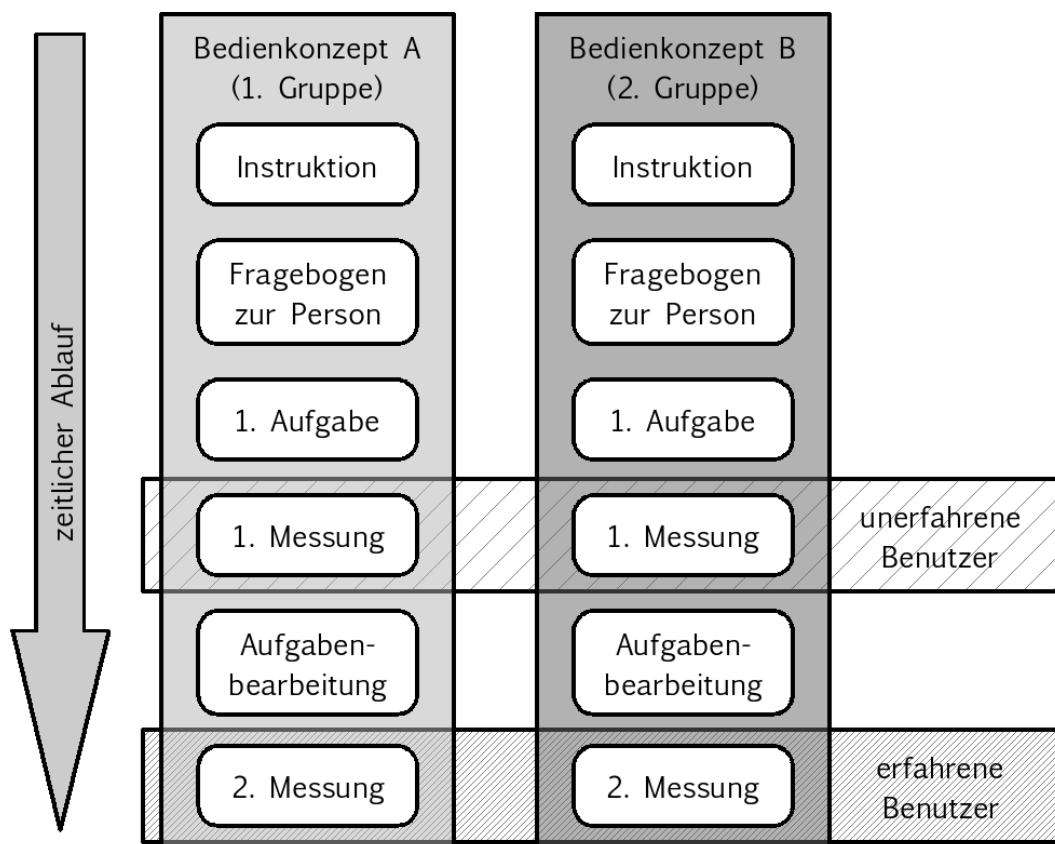


Abb. 40: Zeitlicher Ablauf der Erhebung

Stichprobe

An der Hauptuntersuchung nahmen 84 Psychologiestudenten/innen teil. Das Alter lag zwischen 18 und 36 Jahren mit einem Median von 23 Jahren. 45 der 84 Versuchspersonen waren weiblich. Alle Versuchspersonen hatten Erfahrung im Umgang mit einem Computer und dem Internet seit mindestens 3 Jahren. Der Median der Computererfahrung beträgt 10 Jahre und der Median der Erfahrung mit dem Internet 7 Jahre. Alle Versuchspersonen gaben an, über Zugang zum Internet zu verfügen.

Auswertungen

Zunächst wurden die Eigenschaften der neu entwickelten Skalen für PL und PEUEU analysiert. Hierfür wurden die Antwortverteilungen für die einzelnen Items sowie für die gemittelten Skalenwerte betrachtet. Vor allem die gebildeten Skalenwerte sollten als Voraussetzung für die weiteren statistischen Analysen nicht von der Normalverteilung abweichen. Auch für die Antworthäufigkeiten der

Einzelitems war ein deutliches Abweichen von der Normalverteilung nicht erwünscht. Eine Analyse der Korrelationen zwischen den Items diente der Bewertung der Homogenität der Skalen sowie deren Abgrenzung voneinander. Zusätzlich wurde eine Faktorenanalyse durchgeführt, um die Homogenität der Skalen weiter zu überprüfen. Als gängiges Maß für die Reliabilität der Skalen wurde Cronbachs α berechnet.

Im Anschluss an die Überprüfung der neuen Skalen wurden die statistischen Tests zur Überprüfung der Hypothesen der vorliegenden Arbeit gerechnet. Hierzu wurden die erhobenen Daten durch gemischte Modelle modelliert. Entsprechend des Designs mit sowohl abhängig wie auch unabhängig erhobenen Messwerten erfolgte die Modellierung durch gemischte Modelle. Zunächst wurden die Daten entsprechend des TAM modelliert. Eine zweite Modellierung bildete die Daten nach dem postulierten Prozessmodell ab. Ein statistischer Vergleich der Anpassungsgüte der beiden Modellierungen, bei dem berücksichtigt wurde, dass das Prozessmodell mit mehr Parametern eine aufwendigere Modellierung darstellt, beantwortete die Frage, ob das Prozessmodell dem TAM überlegen ist.

Verteilungsform der Itemantworten

Da in der statistischen Auswertung Verfahren zum Einsatz kamen, die eine Normalverteilung der Werte voraussetzen, wurde die Verteilungsform der Itemantworten auf Abweichung von der Normalverteilung untersucht. Hierbei wurde wie in der zweiten Itemauswahl (Kapitel Zweite Itemselektion – Beurteilung von Routenplanern) verfahren.

Für die nicht stetig verteilten Antworten der einzelnen Items wurde der χ^2 -Test zur Prüfung auf Abweichung von einer vorgegebenen Verteilung eingesetzt. Die vorgegebene Verteilung war eine Normalverteilung mit aus den Antworten geschätzten Werten für den Mittelwert und die Standardabweichung. Die erwarteten Häufigkeiten für die Antwortkategorien der Skalen wurden als Quantilen dieser Verteilung errechnet. Die Grenzen der Quantilen bildeten die arithmetischen Mittelwerte zwischen zwei Skalenstufen.

Insgesamt wurden 24 Items jeweils zweimal erhoben. Eine Zusammenfassung der Werte der beiden Messzeitpunkte war nicht zulässig. Hierbei wären abhängig (wiederholt an derselben Person) erhobene und unabhängig (an verschiedenen Personen) erhobene Messwerte vermischt worden. Der durchgeführte χ^2 -Test hätte dies nicht berücksichtigt, was zu nicht interpretierbaren Ergebnissen geführt hätte. Entsprechend wurde die Verteilung der Antworthäufigkeiten zweimal für jedes Item überprüft.

Für die resultierenden 48 Einzeltests wurde insgesamt ein α -Niveau von 5 % angesetzt, dass für die einzelnen Tests nach der Methode von Holm (1979) korrigiert wurde. Die Ergebnisse für die erste Messung sind in Tab 9 und für die zweite Messung in Tab 10 dargestellt. Die Aufteilung auf zwei Tabellen wurde gewählt, da die getrennte Darstellung von Verteilungen des ersten Messzeitpunkts

und Verteilungen des zweiten Messzeitpunkts vorteilhaft erschien. Die Korrektur des α -Niveau wurde über alle 48 Einzeltests und somit über beide Tabellen hinweg durchgeführt. Als Konsequenz ergibt sich keine kontinuierliche Reihenfolge der korrigierten α -Werte innerhalb einer Tabelle, sondern nur über beide Tabellen hinweg.

Item	χ^2 -Wert	p-Wert	α -Wert (nach Holm)	Ergebnis
L08	6,99	0,1366	0,0011	n. sig.
E17	6,82	0,1456	0,0012	n. sig.
E14	6,82	0,1457	0,0012	n. sig.
L06	5,92	0,2051	0,0013	n. sig.
U03	5,84	0,2113	0,0013	n. sig.
U04	5,66	0,2264	0,0014	n. sig.
L09	5,30	0,2577	0,0014	n. sig.
U06	5,21	0,2668	0,0015	n. sig.
E07	4,98	0,2894	0,0015	n. sig.
E11	4,81	0,3078	0,0016	n. sig.
E15	4,31	0,3663	0,0019	n. sig.
I01	4,25	0,3735	0,0020	n. sig.
U01	4,24	0,3744	0,0021	n. sig.
L07	4,01	0,4049	0,0022	n. sig.
L13	3,88	0,4219	0,0024	n. sig.
U05	2,64	0,6206	0,0029	n. sig.
U02	2,26	0,6889	0,0036	n. sig.
L12	1,88	0,7575	0,0050	n. sig.
L02	1,27	0,8661	0,0063	n. sig.
I02	1,15	0,8860	0,0071	n. sig.
E12	0,91	0,9230	0,0083	n. sig.
E13	0,53	0,9700	0,0167	n. sig.
E02	0,49	0,9750	0,0250	n. sig.
L11	0,28	0,9913	0,0500	n. sig.

Tab 9: χ^2 -Tests der Antwortverteilung zum ersten Messzeitpunkt ($df=4$, α -Grundniveau=5 %))

Die Auflistung der Items in Tab 9 ergibt eine absteigende Reihenfolge der χ^2 -Werte und somit eine aufsteigende Reihenfolge der p-Werte. Die korrigierten α -Werte ergeben keine kontinuierliche Reihenfolge innerhalb von Tab 9, da die

Korrektur der α -Werte über beide Messzeitpunkte gleichzeitig erfolgte und sich die korrigierten α -Werte somit auf Tab 9 und Tab 10 aufteilen. Von den – in Tab 9 dargestellten – Werten für die erste Messung zeigt das oberste Item L08 (Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen) die deutlichste Abweichung von einer Normalverteilung. Die Verteilung der Antworten für dieses Item zum ersten Messzeitpunkt ist in Abb. 41 dargestellt.

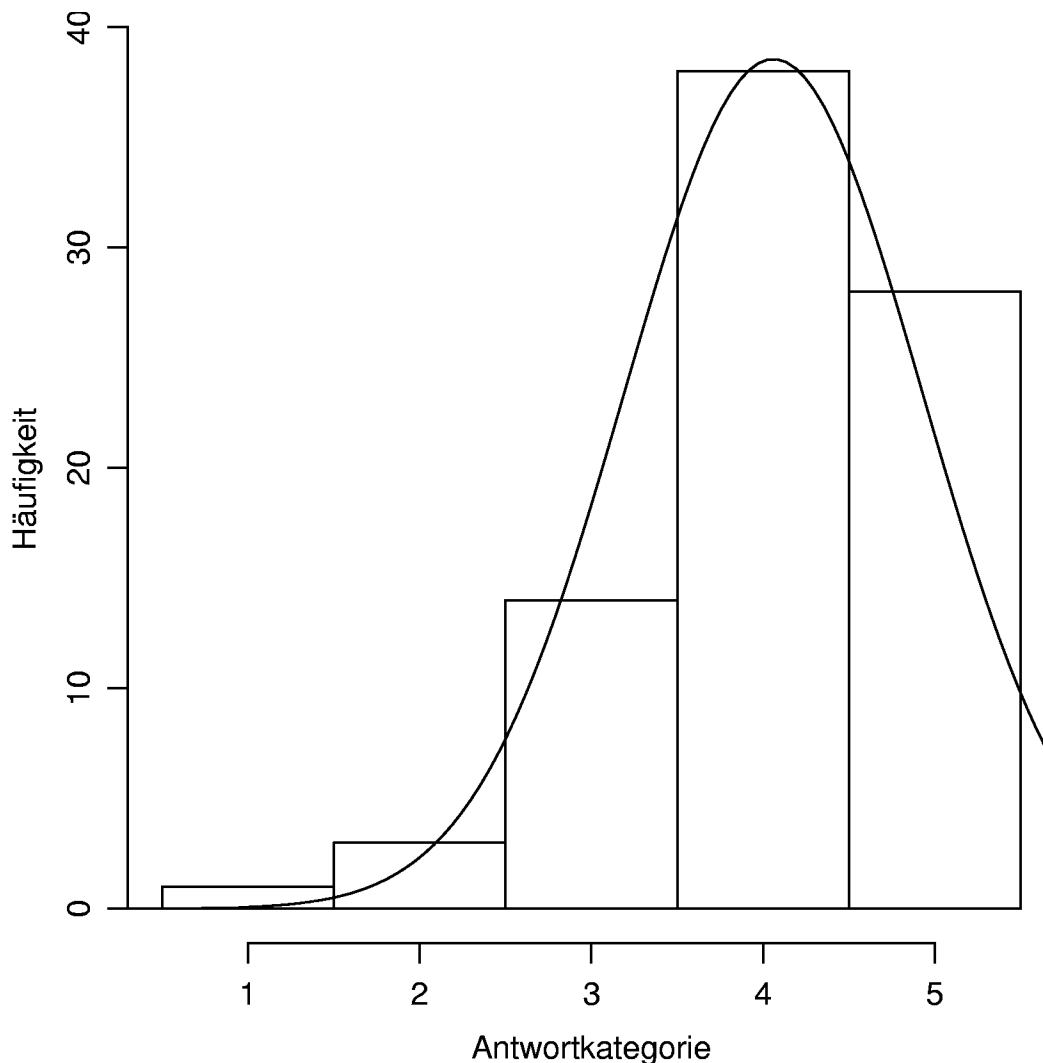


Abb. 41: Verteilung der Antworten für Item L08 zum ersten Messzeitpunkt

Die vierte Antwortkategorie wurde am häufigsten ausgewählt. Die Verteilung hat eine Schiefe von $-0,874$, ist also linksschief. Der Exzess der Verteilung beträgt $0,778$; die Verteilung ist also steilgipflig. Die Abweichung von der zum Vergleich dargestellten Normalverteilung mit gleichem Mittelwert und gleicher Standardabweichung ist statistisch nicht signifikant.

Über beide Messzeitpunkte hinweg betrachtet weichen jedoch die drei Items (L07, E11 und L06) beim zweiten Messzeitpunkt noch stärker von einer Normalverteilung ab als Item L08 bei der ersten Messung. Dies sind die drei obersten Items in der nachfolgenden Tab 10. Item L11 (Die Oberfläche von

SOFTWARE wirkt auf mich gut strukturiert) zeigt mit einem χ^2 -Wert von 0,28 bei der ersten Messung die beste Übereinstimmung mit einer Normalverteilung über beide Messzeitpunkte betrachtet. Alle Abweichungen von einer Normalverteilung sind statistisch nicht signifikant.

Item	χ^2 -Wert	p-Wert	α -Wert (nach Holm)	Ergebnis
L07	13,34	0,0097	0,0010	n. sig.
E11	10,55	0,0321	0,0011	n. sig.
L06	10,24	0,0366	0,0011	n. sig.
U01	7,69	0,1037	0,0011	n. sig.
I02	6,11	0,1911	0,0012	n. sig.
U03	5,88	0,2082	0,0013	n. sig.
L02	5,66	0,2258	0,0014	n. sig.
E07	4,72	0,3172	0,0016	n. sig.
E12	4,69	0,3200	0,0017	n. sig.
E15	4,42	0,3525	0,0017	n. sig.
E17	4,33	0,3633	0,0018	n. sig.
U06	4,26	0,3722	0,0019	n. sig.
L12	3,88	0,4219	0,0023	n. sig.
L08	3,33	0,5038	0,0025	n. sig.
L13	2,81	0,5893	0,0026	n. sig.
L11	2,81	0,5898	0,0028	n. sig.
L09	2,32	0,6774	0,0031	n. sig.
U02	2,28	0,6842	0,0033	n. sig.
I01	2,22	0,6955	0,0038	n. sig.
E02	2,06	0,7251	0,0042	n. sig.
U05	1,91	0,7524	0,0045	n. sig.
E14	1,56	0,8159	0,0056	n. sig.
U04	0,67	0,9553	0,0100	n. sig.
E13	0,67	0,9553	0,0125	n. sig.

Tab 10: χ^2 -Tests der Antwortverteilung zum zweiten Messzeitpunkt ($df=4$, α -Grundniveau=5 %)

Wie in Tab 9 sind auch in Tab 10 die Items so angeordnet, dass sich eine absteigende Reihenfolge der χ^2 -Werte und entsprechend eine aufsteigende Reihenfolge der p-Werte ergibt. Über beide Messzeitpunkte (und somit über Tab 9

und Tab 10) hinweg zeigt sich für Item L07 (Die Funktion von allen Bedienelementen ist offensichtlich) zum zweiten Messzeitpunkt mit einem χ^2 -Wert von 13,34 die schwächste Übereinstimmung mit einer Normalverteilung. Die Antworthäufigkeiten von Item L07 sind in Abb. 42 dargestellt. Die beste Übereinstimmung ergibt sich für den zweiten Messzeitpunkt bei Item E13 (Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben) mit einem χ^2 -Wert von 0,67. Auch für den zweiten Messzeitpunkt sind alle Abweichungen statistisch nicht signifikant.

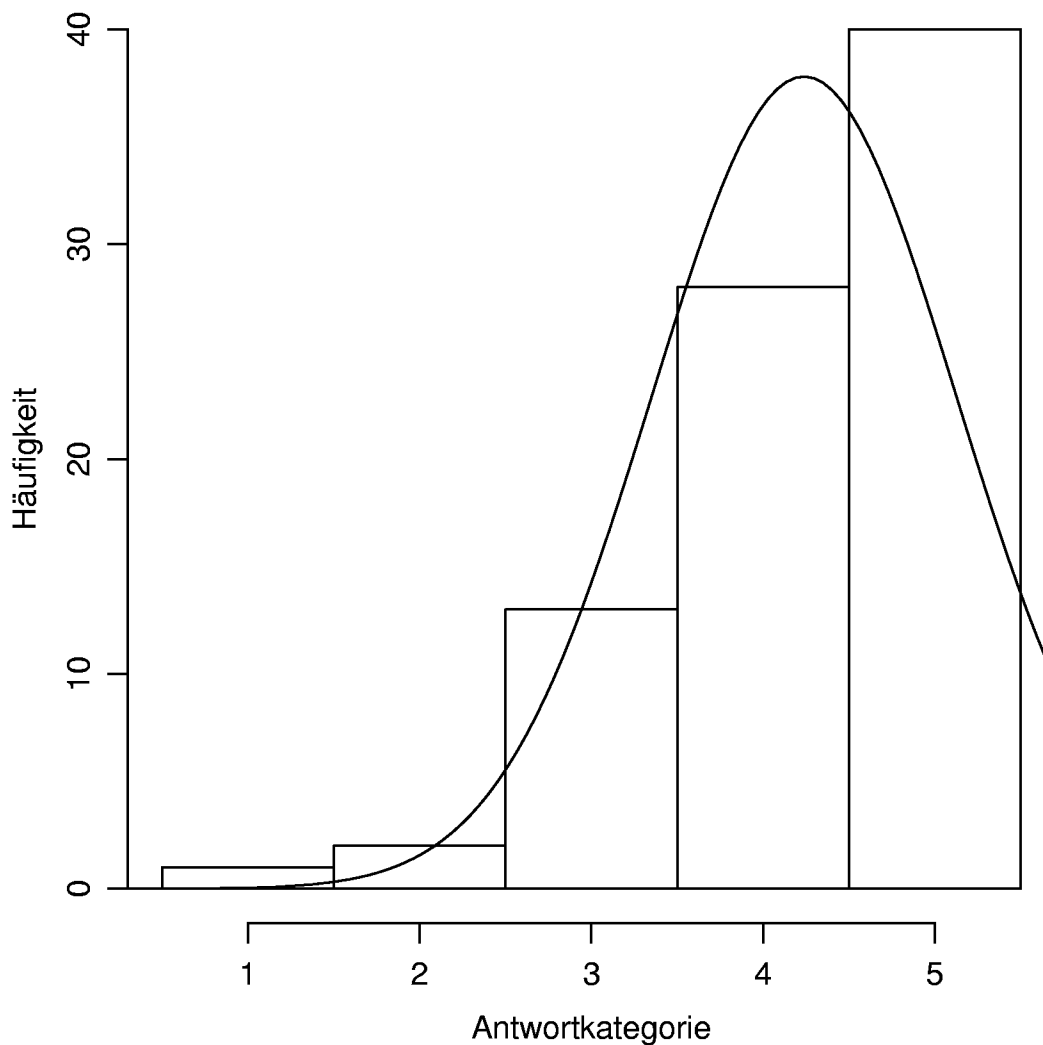


Abb. 42: Verteilung der Antworten für Item L07 zum zweiten Messzeitpunkt

Bei der zweiten Messung ist die fünfte Antwortkategorie von Item L07 am häufigsten besetzt. Die Verteilung ist linksschief mit einer Schiefe von -1,089. Der Exzess der Verteilung beträgt 0,936; die Verteilung ist also steilgipflig. Die Abweichung von der zum Vergleich dargestellten Normalverteilung mit gleichem Mittelwert und gleicher Standardabweichung ist statistisch nicht signifikant. Die Verteilungsformen der Antworthäufigkeiten der übrigen Items sind in Anhang H dargestellt.

Analyse der Interitemkorrelationen

Die inhaltlichen Zusammenhänge zwischen Items innerhalb einer Skala sowie die Abgrenzungen zwischen Items verschiedener Skalen wurden anhand einer Maximum-Likelihood-Faktorenanalyse betrachtet. Diese wurde über eine gemeinsame Korrelationsmatrix aller PU-, PL- und PEUEU-Items gerechnet. Die Ladungsmatrix wurde nach dem Varimax-Kriterium rotiert, was zu einer besseren Interpretierbarkeit der Faktoren führt (Bortz, 1993, S. 507). Bei der Faktorenanalyse ist die Auswahl einer geeigneten Anzahl an Faktoren grundsätzlich schwierig. Hierfür existieren unterschiedliche Verfahren (z. B. Kaiser-Guttman-Kriterium und Scree-Test), die regelmäßig zu abweichenden Ergebnissen führen.

Die Daten der ersten und der zweiten Messung konnten für die Faktorenanalyse nicht sinnvoll zusammengefasst werden, weil durch das Versuchsdesign eine Abhängigkeit zwischen Werten der ersten und der zweiten Messung gegeben war, die statistisch in einer Faktorenanalyse nicht berücksichtigt werden konnte. Entsprechend mussten zwei getrennte Faktorenanalysen durchgeführt werden. Sowohl nach dem Kaiser-Guttman-Kriterium als auch nach der Scree-Methode ergab sich eine unterschiedliche Anzahl bedeutsamer Faktoren für die beiden Messzeitpunkte. Zur Veranschaulichung sind nachfolgend Scree-Plots für die beiden Messzeitpunkte in Abb. 43 und Abb. 44 abgebildet.

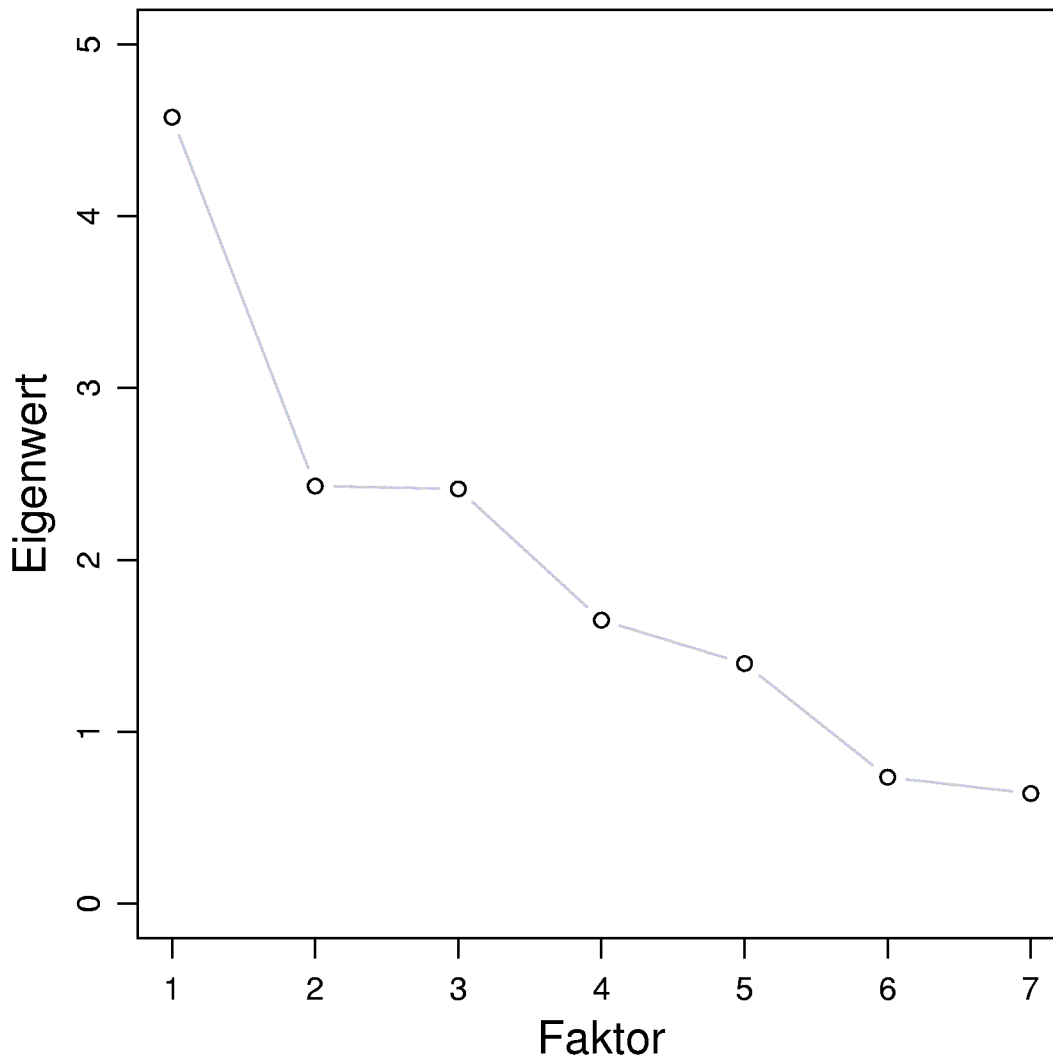


Abb. 43: Scree-Plot der Eigenwerte für die erste Messung

Im Scree-Plot sind die Eigenwerte der Faktoren in absteigender Reihenfolge auf der Ordinate abgetragen. Die Abszisse stellt die Faktoren in gleichen Abständen dar. Nach dem Kaiser-Guttman-Kriterium sind alle Faktoren mit einem Eigenwert größer als eins bedeutsam. Für den ersten Messzeitpunkt wären dies die ersten fünf Faktoren.

Beim Scree-Test wird versucht, einen ausgeprägten Knick in der Kurve, die die Punkte verbindet, zu erkennen. Alle Faktoren vor dem Knick werden als bedeutsam betrachtet, die übrigen Faktoren als nicht mehr bedeutsam. In Abb. 43 sind drei Knicke zu erkennen. Das erste, deutliche Abknicken erfolgt am zweiten Faktor. Dementsprechend wäre nur der erste Faktor bedeutsam. Alternativ kann auch der vierte und der sechste Datenpunkt als Knick betrachtet werden.

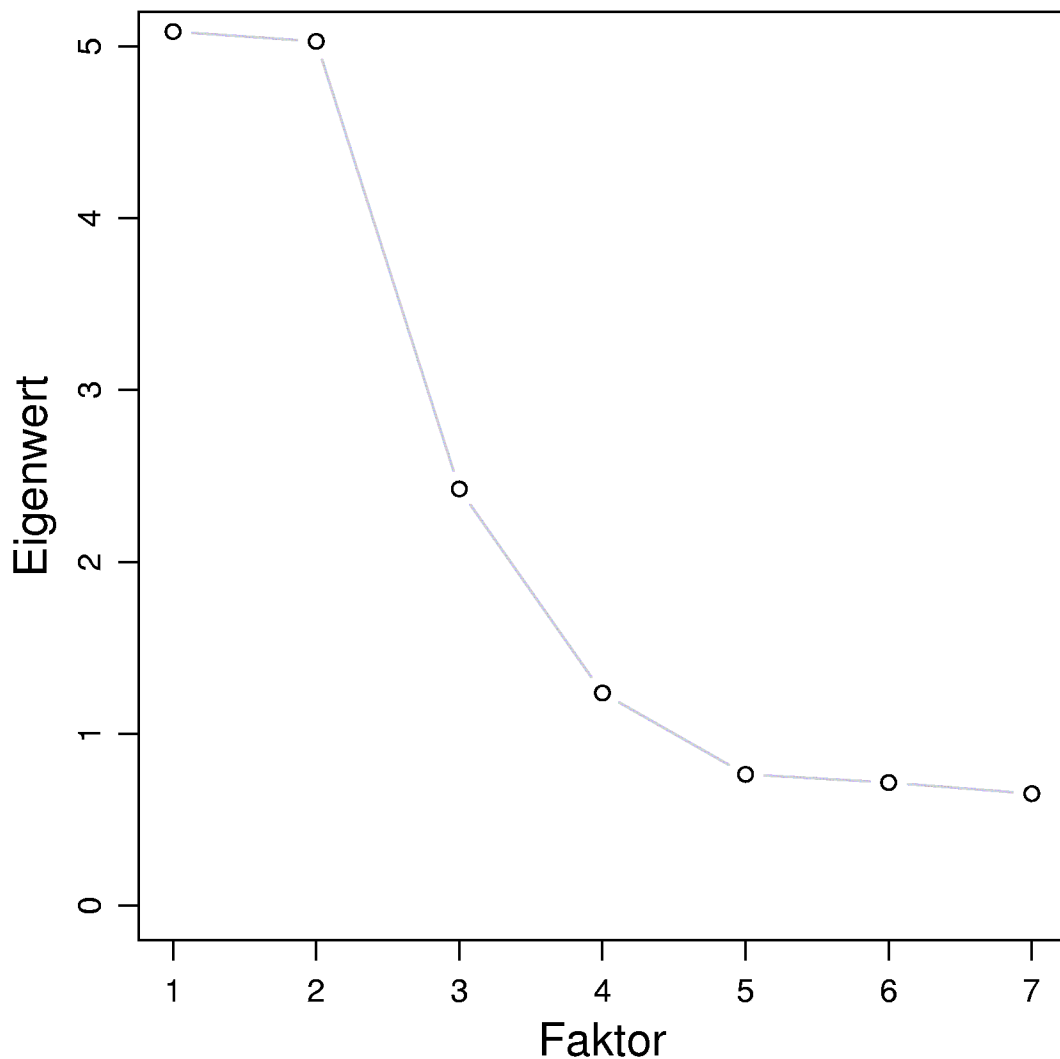


Abb. 44: Scree-Plot der Eigenwerte für die zweite Messung

Für den zweiten Messzeitpunkt sind nach dem Kaiser-Guttman-Kriterium die ersten vier Faktoren bedeutsam. Die übrigen Faktoren besitzen einen Eigenwert kleiner als eins.

In Abb. 44 ist nicht eindeutig, wo die Kurve im Sinne des Scree-Tests einen Knick aufweist. Mögliche Punkte sind für den dritten und den vierten Faktor zu erkennen. Dementsprechend wären die ersten beiden bzw. die ersten drei Faktoren bedeutsam.

Die Anzahl bedeutsamer Faktoren nach dem Kaiser-Guttman-Kriterium wurde verworfen. Zum Kaiser-Guttman-Kriterium weist Bortz (1993, S. 503) darauf hin, „dass die Anzahl der bedeutsamen Faktoren nach dieser Regel meistens überschätzt wird“. Nach dem Scree-Test war es möglich, die ersten drei Faktoren für beide Messzeitpunkte als bedeutsam zu betrachten. Diese Anzahl entsprach den drei verwendeten Skalen und wurde für die nachfolgenden Faktorenanalysen eingesetzt. Die Ladungen der Items auf die drei Faktoren für den ersten Messzeitpunkt sind Tab 11 zu entnehmen. Tab 12 enthält die Ladungen der Items für die zweite

Messung. In beiden Tabellen ist die jeweils höchste Ladung eines Items hervorgehoben.

Item	Faktor 1	Faktor 2	Faktor 3
L02	0,523	-0,102	0,304
L06	0,853	0,115	0,026
L07	0,768	0,072	-0,024
L08	0,669	-0,006	0,129
L09	0,765	0,076	0,074
L11	0,771	-0,059	0,130
L12	0,793	0,115	0,067
L13	0,649	0,117	0,348
E02	0,015	0,469	0,096
E07	0,016	0,638	0,271
E11	0,202	0,818	0,137
E12	-0,146	0,337	0,301
E13	0,156	0,504	0,252
E14	-0,166	0,344	0,322
E15	0,048	0,653	-0,035
E17	0,350	0,128	0,060
U01	0,200	0,333	0,582
U02	0,173	0,120	0,609
U03	0,155	0,385	0,590
U04	0,258	0,146	0,832
U05	0,301	0,469	0,432
U06	0,106	0,545	0,439

Tab 11: Ladungen der Items aller Skalen auf den ersten drei Faktoren (erste Messung)

Die PL-Items laden bei der ersten Messung auf dem ersten Faktor zwischen 0,523 und 0,853. Die Ladungen auf den zweiten Faktor bewegen sich zwischen -0,102 und 0,117. Die geringste Ladung der PL-Items auf dem dritten Faktor beträgt -0,024 und die höchste Ladung 0,348. Die PL-Items laden hauptsächlich auf dem ersten Faktor, der entsprechend mit Perceived Learnability bezeichnet werden kann.

Die PEUEU-Items laden zwischen -0,166 und 0,350 auf dem ersten und zwischen 0,337 und 0,818 auf dem zweiten Faktor. Die Ladungen auf dem dritten Faktor bewegen sich zwischen -0,035 und 0,322. Abgesehen von Item E17 laden alle

PEUEU-Items am höchsten auf dem zweiten Faktor. Der zweite Faktor entspricht *Perceived Ease of Use for Experienced Users*.

Die hohe Ladung von Item E17 (Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren) auf dem ersten Faktor bei diesem Messzeitpunkt entspricht nicht der erwarteten Zuordnung der Items zu den Skalen. Item E17 scheint bei der ersten Messung inhaltlich Aspekte der Perceived Learability zu erfassen. Mit 0,350 fällt die Ladung auf dem ersten Faktor allerdings niedrig aus.

Bei der ersten Messung liegen die Ladungen der PU-Items auf dem ersten Faktor zwischen 0,106 und 0,301. Auf dem zweiten Faktor ist die geringste Ladung bei Item U02 0,120. Die höchste Ladung eines PU-Items beträgt 0,545. Die Ladungen auf dem dritten Faktor bewegen im Bereich zwischen 0,432 und 0,832. Vier der sechs PU-Items laden am höchsten auf dem dritten Faktor, der somit Perceived Usefulness entspricht.

Die beiden Items U05 und U06 laden ähnlich hoch auf dem zweiten und dem dritten Faktor, wobei beide Items auf dem zweiten Faktor (*Perceived Ease of Use for Experienced Users*) höher laden. Dieses Ladungsmuster der Items weist darauf hin, dass die Items sowohl Aspekte des Konstrukts *Perceived Ease of Use for Experienced Users* wie auch Aspekte des Konstrukts *Perceived Usefulness* erfassen. Eine nähere Betrachtung der beiden Konstrukte und ihrer Abgrenzung zueinander könnte hierzu Klarheit schaffen, war aber nicht Gegenstand dieser Arbeit.

Item	Faktor 1	Faktor 2	Faktor 3
L02	0,645	0,027	0,421
L06	0,902	-0,112	0,035
L07	0,880	0,041	-0,019
L08	0,596	0,021	0,255
L09	0,781	-0,011	-0,099
L11	0,641	0,343	0,126
L12	0,875	-0,008	0,096
L13	0,777	0,122	0,229
E02	0,105	0,665	0,250
E07	-0,036	0,651	0,320
E11	0,189	0,737	0,253
E12	-0,067	0,632	0,442
E13	0,014	0,734	0,400
E14	-0,099	0,804	0,276
E15	0,056	0,745	0,149
E17	0,026	0,587	0,014
U01	0,058	0,337	0,386
U02	0,048	0,448	0,599
U03	0,255	0,331	0,564
U04	0,095	0,302	0,925
U05	0,306	0,426	0,561
U06	0,298	0,413	0,557

Tab 12: Ladungen der Items aller Skalen auf den ersten drei Faktoren (zweite Messung)

Bei der zweiten Messung luden die PL-Items zwischen 0,641 und 0,902 auf dem ersten Faktor. Für den zweiten Faktor bewegten sich die Ladungen zwischen -0,112 und 0,343. Die Ladungen auf dem dritten Faktor bewegten sich zwischen -0,099 und 0,421. Wie schon bei der ersten Messung luden die Items am höchsten auf dem ersten Faktor. Auch für die zweite Messung wurde der erste Faktor mit Perceived Learnability bezeichnet.

Bei der zweiten Messung lagen die Ladungen der PEUEU-Items auf dem ersten Faktor zwischen -0,099 und 0,189. Die geringste Ladung auf dem zweiten Faktor war 0,587 und die höchste Ladung 0,804. Die Ladungen auf dem dritten Faktor bewegten sich im Bereich zwischen 0,014 und 0,442. Alle PEUEU-Items luden am höchsten auf dem zweiten Faktor, der somit – wie schon bei der ersten Messung –

Perceived Ease of Use for Experienced Users entsprach.

Die PU-Items luden zwischen 0,048 und 0,306 auf dem ersten und zwischen 0,302 und 0,448 auf dem zweiten Faktor. Die Ladungen auf dem dritten Faktor bewegten sich zwischen 0,386 und 0,925. Alle PU-Items luden am höchsten auf dem dritten Faktor. Der dritte Faktor entsprach Perceived Usefulness.

Bei der ersten Messung luden die drei Items E17, U05 und U06 anders als erwartet. E17 (Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren) lud am höchsten auf dem Faktor Perceived Learnability, während die beiden PU-Items am höchsten auf dem PEUEU-Faktor luden. Item E17 war bei der ersten Itemselektion nicht aus dem Itempool ausgeschlossen worden (S. 119), weil es als einziges Item den Aspekt der Fehlerkorrektur erfasste.

Über die Daten des zweiten Messzeitpunktes gerechnet luden alle Items wie theoretisch erwartet. Die PL-Items luden gemeinsam auf einem Faktor, die PEUEU-Items luden gemeinsam auf dem zweiten Faktor und die PU-Items luden gemeinsam auf dem dritten Faktor. Die inhaltlichen Zusammenhänge zwischen den Items waren damit im Wesentlichen wie erwartet.

Bildung der Skalenwerte

Die Skalenwerte der Skalen für PL- und PEUEU wurden durch arithmetische Mittelung der einzelnen Itemwerte gebildet. Dies war problemlos möglich, da alle Items die gleiche Polung aufwiesen, d.h. ein hoher Wert auf einem Item hat inhaltlich immer dieselbe Bedeutung wie ein hoher Wert auf einem anderen Item.

Davis Konstrukt PEOU differenziert nicht in die Aspekte PL und PEUEU. Entsprechend wurde die Skala für PEOU durch Mittelung der beiden Skalenwerte der PL- und der PEUEU-Skala gebildet⁷. Wie die einzelnen Items wiesen auch die beiden Skalen eine identische Polung auf.

Die Skala für PU wurde ebenfalls durch arithmetische Mittelung der Skalenwerte gebildet. Gegenüber der Berechnung von Summenwerten als Skalenwert bestand der Vorteil einer arithmetischen Mittelung darin, dass die Skalenwerte direkt mit der fünfstufigen Antwortskala für die Items vergleichbar war. Summenwerte sind von der Anzahl der Items abhängig.

Die Messung der Nutzungsintention erfolgte durch die beiden Items I01 (Falls ROUTENPLANER für mich verfügbar wäre, würde ich es benutzen) und I02 (Wenn möglich würde ich ROUTENPLANER häufig nutzen). Auch diese beiden Items wurden gemittelt, um zu einem stabileren Messwert der Nutzungsintention zu gelangen.

⁷ Da sowohl die PL- wie auch die PEUEU-Skala aus jeweils 8 Items besteht, führt dieses Vorgehen zum selben Ergebnis wie eine gemeinsame arithmetische Mittelung der PL- mit den PEUEU-Items.

Verteilungsform der Skalenwerte

Wichtiger als die Verteilungsform der Einzelitems war die Verteilung der Antworten der gebildeten Skalenwerte. Die Skalenwerte wurden durch arithmetische Mittelung der einzelnen Itemwerte gebildet. Für die PEOU-Skala wurden sowohl die PL- wie auch die PEUEU-Items zusammengefasst. Die PU-, PL- und PEUEU-Skalen wurden durch Mittelung der jeweiligen Skalenitems gebildet.

Für die Skalenwerte ergab sich durch die Mittelung eine deutlich feinere Abstufung im Vergleich mit den ganzzahligen Itemwerten. Hiermit waren die Voraussetzungen zur Durchführung eines Kolmogorov-Smirnov-Anpassungstest auf Normalverteilung erfüllt. Im streng mathematischen Sinne sind auch die gemittelten Werte nicht stetig verteilt, hinsichtlich der Voraussetzungen für den Kolmogorov-Smirnov-Anpassungstest kann diese Bedingung trotzdem als erfüllt betrachtet werden.

Der normale Kolmogorov-Smirnov-Test unterschätzt die Abweichung von der Normalverteilung, weil die Verteilungsparameter (Mittelwert und Standardabweichung) der Vergleichsnormalverteilung aus der zu prüfenden Verteilung geschätzt werden. Lilliefors (1967) hat eine Variante des Kolmogoroff-Smirnov-Test vorgeschlagen, der diesem Umstand Rechnung trägt und die Wahrscheinlichkeiten entsprechend anpasst.

Für die Überprüfung der Verteilungsform der Skalen wurde die Variante mit Lilliefors-Schranken eingesetzt. Das α -Gesamtniveau wurde auf 5 % gesetzt. Da die Daten der prae- und der post-Messung für die 4 gebildeten Skalen jeweils getrennt betrachtet werden mussten, waren insgesamt 8 Verteilungen auf Abweichung von der Normalverteilung zu prüfen. Das α -Niveau der Einzeltests wurde wieder nach der Methode von Holm (1979) korrigiert. Die Prüfgrößen und die Grenzwerte sind in Tab 13 dargestellt.

Skala	D-Wert	p-Wert	α -Wert (nach Holm)	Ergebnis
PL (prae)	0,1160	0,00712	0,00625	nicht sig.
PL (post)	0,1104	0,01311	0,00714	nicht sig.
PEOU (prae)	0,1022	0,03386	0,00833	nicht sig.
PU (prae)	0,0986	0,04454	0,01000	nicht sig.
PE (post)	0,0922	0,08153	0,01250	nicht sig.
PE (prae)	0,0767	0,27400	0,01666	nicht sig.
PEOU (post)	0,0700	0,41370	0,02500	nicht sig.
PU (post)	0,0643	0,53160	0,05000	nicht sig.

Tab 13: Kolmogorov-Smirnov-Tests der Verteilung der Skalenwerte (α -Grundniveau=5 %)

Keine der Abweichungen von der Normalverteilung ist statistisch signifikant. Die Skalen sind in Tab 13 so aufgelistet, dass sich eine absteigende Reihenfolge der

D-Werte des Kolmogorov-Smirnov-Test ergibt. Dies entspricht einer aufsteigenden Reihenfolge für die p-Werte. Die PL-Skala weicht insgesamt am deutlichsten von einer Normalverteilung ab. Zum ersten Messzeitpunkt ist die Abweichung größer als zum zweiten Messzeitpunkt. Der D-Wert des Kolmogorov-Smirnov-Test beträgt für den ersten Messzeitpunkt 0,1160 ($p=0,00712$).

Für einen visuellen Vergleich zweier Verteilungsformen eignen sich so genannte Quantile-Quantile-Plots (kurz Q-Q-Plots). Bei dieser Darstellung werden die Quantilen zweier Verteilungen gegeneinander abgetragen. Für die visuelle Begutachtung der Abweichung einer empirischen Verteilung von einer theoretischen Verteilung – hier der Normalverteilung – werden die Quantilen der Skalenwerte gegen die Quantilen der Normalverteilung abgetragen. Stammen die empirisch ermittelten Werte tatsächlich aus einer Grundgesamtheit mit der angenommenen Verteilungsform, so ergibt sich eine gute Übereinstimmung zwischen den Quantilen der empirischen und der theoretischen Verteilung. In diesem Fall liegen die Punkte auf einer Geraden bzw. für z-standardisierte Werte auf der Diagonalen. Die folgenden Abbildungen stellen Q-Q-Plots der gebildeten Skalen jeweils für die beiden Messzeitpunkte dar. Die empirischen Skalenwerte wurden z-transformiert, da dies die visuelle Begutachtung der Diagramme erleichtert.

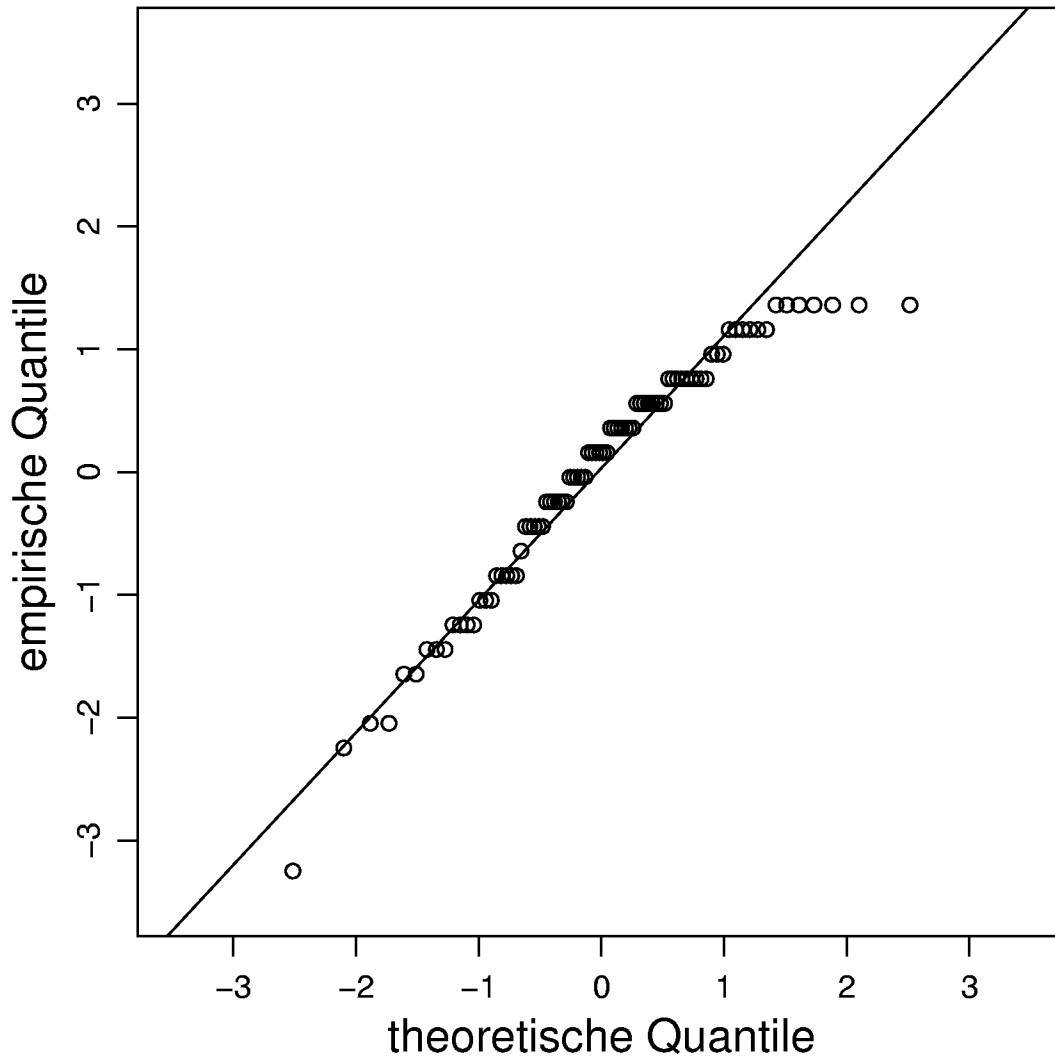


Abb. 45: Q-Q-Plot der PL-Skala (z-transformiert) für die erste Messung

Die Gerade im Diagramm geht durch das erste und vierte Quartil und ermöglicht eine optische Orientierung. Die stärksten Abweichungen ergeben sich für am oberen Ende der Verteilung. Hier flachen die empirischen Werte ab, während die theoretische Normalverteilung weiter ansteigt.

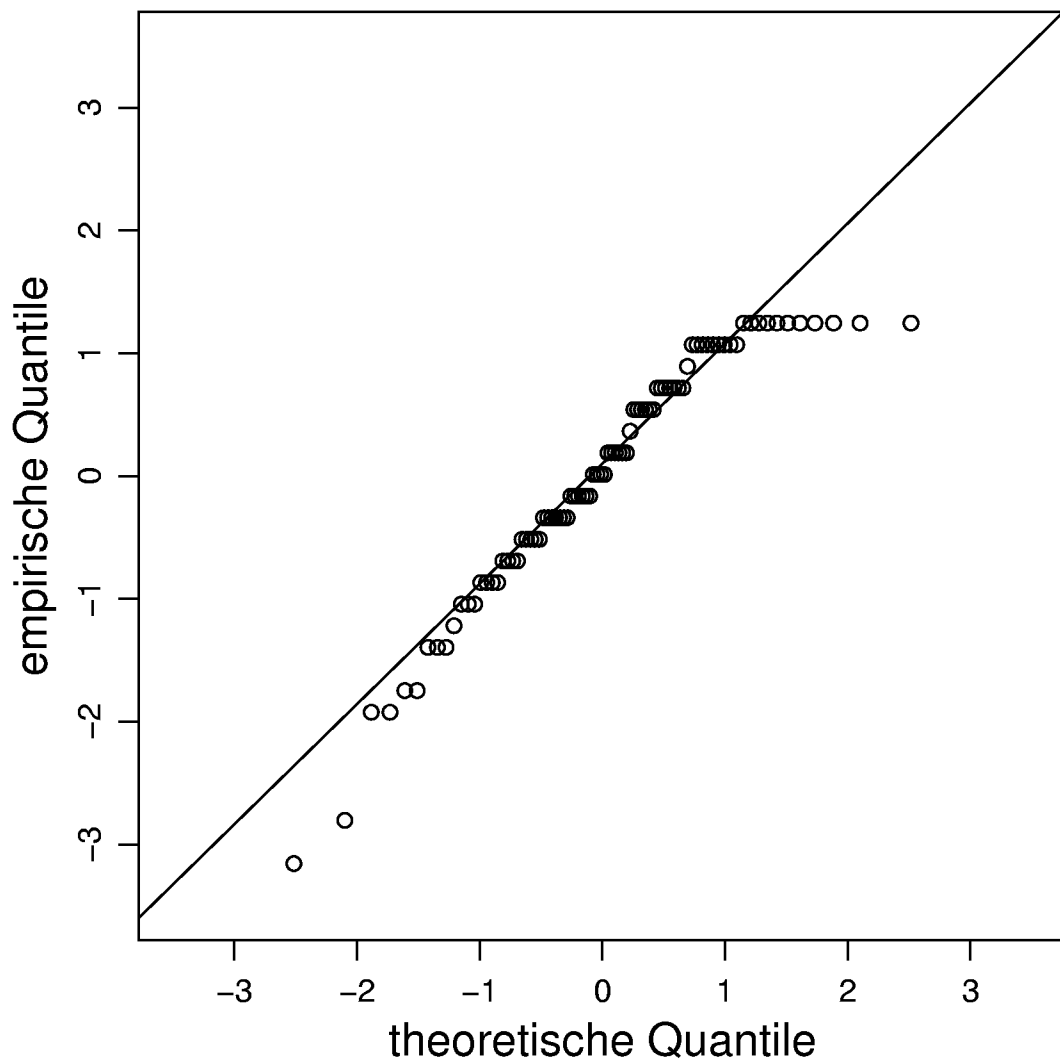


Abb. 46: Q-Q-Plot der PL-Skala (z-transformiert) für die zweite Messung

Auch bei der zweiten Messung ergeben sich für die PL-Skala die deutlichsten Abweichungen am oberen Ende der Verteilung. Die Sprünge der empirischen Quantilen zeigen, dass die empirischen Werte keine stetige Verteilung im engen Sinne aufwiesen. Durch die Mittelung der Itemwerte bei der Skalenbildung entstand eine feinere Abstufung der Skalenwerte im Vergleich zu den fünfstufigen Antwortskalen der Items.

Die Q-Q-Plots der andern drei Skalen sind in Anhang M dargestellt. Wie die PL-Skala zeigt auch die übersetzte PU-Skala die ausgeprägtesten Abweichungen von der Normalverteilung im oberen Bereich. Für die Q-Q-Plots der PEUEU- und der PEOU-Skala lässt sich nicht eindeutig erkennen, wo die deutlichsten Abweichungen auftreten.

Alle gebildeten Skalen weichen nicht signifikant von einer Normalverteilung ab. Damit sind die statistischen Voraussetzungen für eine varianzanalytische Auswertung der Daten erfüllt.

Im Vergleich mit der Normalverteilung zeigen sich für die PL-Skala

Abweichungen im Bereich der oberen Quantilen. Dies zeigt, dass die empirische Skala häufig ihren Maximalwert erreichte, was auf tendenziell zu leichte Items hinweist.

Modellierung der Nutzungsintention

Die neu entwickelten Skalen ermöglichten die Messung der theoretischen Konstrukte PL und PEUEU, welche sich als Prädiktoren für die Nutzungsintention eignen sollten. Die Bestimmung der Zusammenhänge zwischen den unabhängigen Variablen PL, PEUEU, PU sowie PEOU und der Nutzungsintention wurde im Rahmen einer linearen Modellierung durchgeführt. Alle Modellparameter wurden nach der LogLikelihood-Methode geschätzt. Dabei wird eine Wahrscheinlichkeitsfunktion formuliert, welche die Wahrscheinlichkeit beschreibt, mit der abhängig von den gewählten Modellparametern die beobachteten Werte auftreten. Die Parameter des Modells werden so festgelegt, dass sich für die durch diese Parameter bestimmte Wahrscheinlichkeitsfunktion eine maximale Wahrscheinlichkeit ergibt, dass die beobachteten Werte eintreten würden.

Im Versuchsdesign wurden Variablen sowohl an unabhängigen wie auch abhängigen Stichproben erhoben. Als unabhängig werden hierbei Stichproben bezeichnet, bei denen kein systematischer Zusammenhang zwischen den beiden Stichproben besteht. Dies ist in der Studie für die beiden zufällig zugeordneten Gruppen gegeben, die entweder an Bedienkonzept A oder an Bedienkonzept B gearbeitet haben. Abhängig sind Stichproben, wenn ein systematischer Zusammenhang zwischen den beiden Stichproben besteht. Es ist offensichtlich, dass dies für Messwiederholungen gegeben ist, da die Zusammensetzung der zweiten Stichprobe durch die erste Stichprobe festgelegt ist.

Für die Interpretierbarkeit ergeben sich Unterschiede zwischen abhängig und unabhängig gemessenen Stichproben. Bei einer unabhängigen Messung gibt es keine Möglichkeit, Rückschlüsse über spezifische Eigenheiten einzelner Elemente (hier Probanden) der Stichprobe zu ziehen.

Bewertet beispielsweise eine Person bei einer einmaligen Messung ihre Nutzungsintention für eine Informationstechnologie hoch, so kann diese Person generell moderner Technik gegenüber sehr aufgeschlossen sein – was eine Eigenschaft der Person wäre – oder die Technologie, die betrachtet wird, ist sehr ansprechend. Dies könnte bei einem neuartigen Bedienkonzept wie Spracheingabe der Fall sein, was einem über die Einzelperson hinaus gehenden Einfluss auf die gesamte Population entsprechen würde. Eine einmalige Messung der Nutzungsintention erlaubt es nicht, die individuelle Technikaffinität der einzelnen Probanden zu schätzen. Dies würde durch eine wiederholte Messung mit unterschiedlichen Programmen bzw. Technologien möglich. Bewertet hierbei eine Person wiederholt überdurchschnittlich hoch, so kann dies auf die individuelle Technikaffinität dieser Person zurückgeführt werden.

Einflüsse mit Wirkung auf die gesamte Population, wie das neuartige Bedienkonzept der Spracheingabe im vorigen Beispiel, nennt man feste Effekte. In der Regel sind es diese Effekte, denen das Interesse von wissenschaftlichen Studien gilt. Die Ausprägung von Eigenschaften auf individueller Ebene nennt man zufällige Effekte, da sie durch die zufällige Ziehung der Untersuchungsstichprobe aus der gesamten Population entstehen. Die zufälligen Effekte wirken in der Regel als Störeffekte in wissenschaftlichen Untersuchungen. Dabei überlagern sie sich allerdings mit den festen Effekten und maskieren diese hierdurch. Besteht die Möglichkeit, zufällige Effekte sinnvoll zu schätzen, kann man den Einfluss der zufälligen und der festen Effekte statistisch trennen und so zu genaueren Aussagen über die festen Effekte gelangen. So genannte gemischte Modell ermöglichen es, feste und zufällige Effekte in eine gemeinsame Modellierung aufzunehmen (Pinheiro & Bates, 2004).

Für das TAM nach Davis ergibt sich entsprechend ein Modell mit folgenden unabhängigen Variablen:

- feste Effekte
 - PU
 - PEOU
- zufälliger Effekt
 - Proband (VP)

Als abhängige Variable vorhergesagt werden soll:

- abhängige Variable
 - Nutzungsintention (I)

Angelehnt an die Syntax von R ergibt sich folgende Formulierung für dieses Modell:

$I = \text{PU PEOU, random} = \text{VP}$

Diese Modellierung der festen Effekte mit den unabhängigen Variablen PU und PEOU sowie der abhängigen Variablen I ist in Abb. 47 dargestellt. Die grau dargestellten Komponenten wurden in dieser Untersuchung nicht erhoben und konnten entsprechend auch nicht modelliert werden.

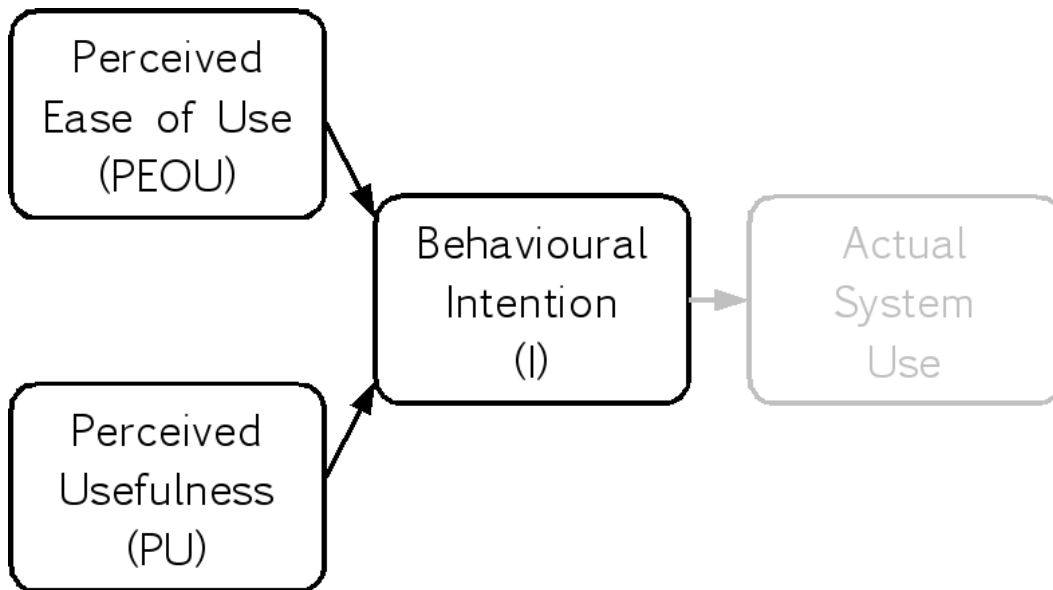


Abb. 47: Modellierung der festen Effekte des TAM (nicht erhobene Modellkomponenten sind grau dargestellt)

Für das Prozessmodell ergeben sich die folgenden unabhängigen Variablen:

feste Effekte

- PU
- Interaktion zwischen PL und Messzeitpunkt (PL*time)
- Interaktion zwischen PEUEU und Messzeitpunkt (PEUEU*time)
- zufälliger Effekt
 - Proband (VP)

Als abhängige Variable soll vorhergesagt werden:

- abhängige Variable
 - Nutzungsintention (I)

Angelehnt an die Syntax von R ergibt sich folgende Formulierung für dieses Modell:

$I = PU \text{ PL*time } PEUEU*time, \text{ random} = VP$

Abb. 48 zeigt die Modellierung der festen Effekte der unabhängigen Variablen PU, PEUEU, PL und des Messzeitpunkts im Prozessmodell.

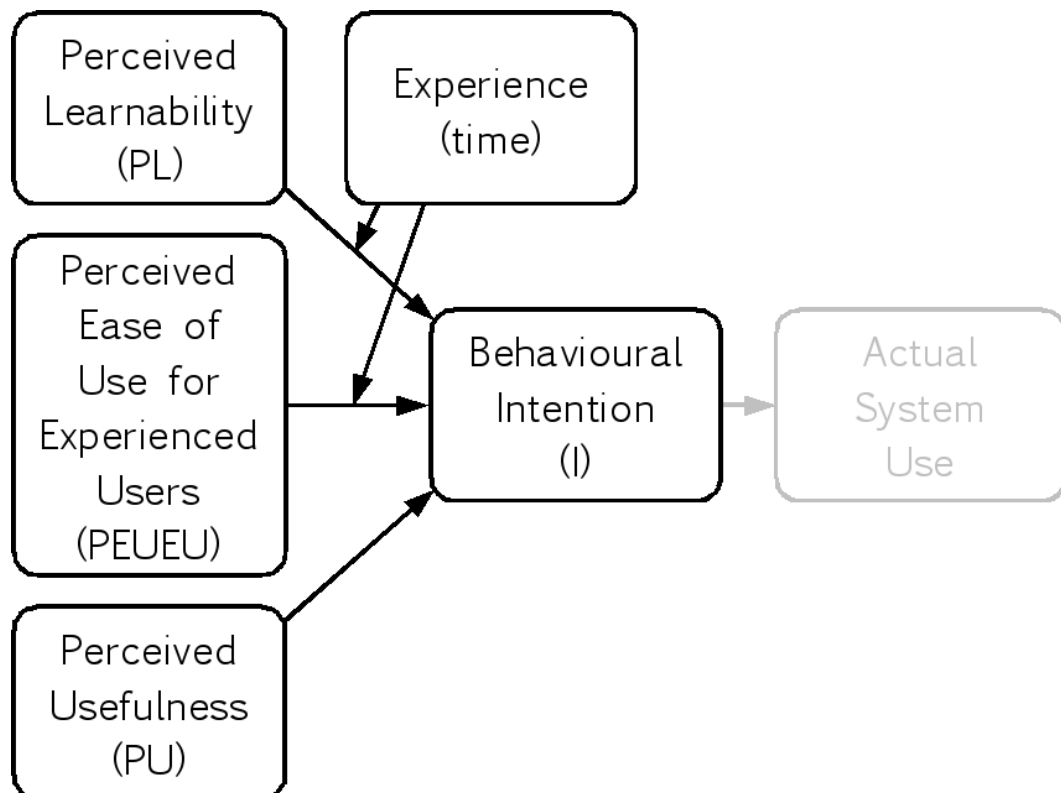


Abb. 48: Modellierung der festen Effekte des Prozessmodells (nicht erhobene Modellkomponenten sind grau dargestellt)

Um die Güte der Modellierung umfassend beurteilen zu können, wurden noch für weitere Modellierungen Parameter berechnet und mit dem TAM und dem Prozessmodell verglichen. Insbesondere wurde beim Prozessmodell überprüft, ob es auch Haupteffekte von PL bzw. PEUEU gibt. Für diese Überprüfung wurden folgenden Effekte modelliert:

feste Effekte

- PU
- PL
- PEUEU
- Interaktion zwischen PL und Messzeitpunkt (PL*time)
- Interaktion zwischen PEUEU und Messzeitpunkt (PEUEU*time)
- zufälliger Effekt

- Proband (VP)

Als abhängige Variable soll vorhergesagt werden:

- abhängige Variable
 - Nutzungsintention (I)

Angelehnt an die Syntax von R ergibt sich folgende Formulierung für dieses Modell:

$I = PU + PL + PEUEU + PL*time + PEUEU*time, random = VP$

Auch eine Modellierung, in der keine der unabhängigen Variablen einen Einfluss auf die Nutzungsintention hat, wurde betrachtet.

Der Vergleich der Anpassungsgüte der Modellierungen erfolgte auf Basis des *Bayesian Information Criterion* (BIC, Schwartz, 1978). Das *Akaike Information Criterion* (AIC, Sakamoto, Ishiguro & Kitagawa, 1986) hätte ein alternatives Kriterium zur Beurteilung der Anpassungsgüte dargestellt. Die beiden Kriterien sind sehr ähnlich, bei der Berechnung des BIC wird die Anzahl der Modellparameter mit dem natürlichen Logarithmus der Probandenzahl gewichtet, während beim AIC die Modellparameter mit einem festen Faktor von 2 gewichtet werden. Generell sollten Modelle so einfach wie möglich sein (Burnham und Anderson, 2002, S.29-36). Eine Erhöhung der Anzahl der Modellparameter wird absolut immer zu einer verbesserten Anpassung an empirische Daten führen. Die Verbesserung sollte jedoch überzufällig ausfallen, um die Aufnahme eines weiteren Parameters in eine Modellierung zu rechtfertigen. Insofern muss der Gewinn an Abbildungsqualität des Modells gemeinsam mit der Modellkomplexität betrachtet werden. Dies leisten Kennwerte wie das AIC und das BIC.

Eine höhere Anzahl an Modellparametern wirkt sich dabei in beiden Kennwerten negativ auf die Bewertung der Anpassungsgüte eines Modells aus. In der vorliegenden Studie mit 82 vollständigen Datensätzen entspricht der $\ln(82)$ einem Wert von 4,407. Somit bewertet das BIC die Aufnahme zusätzlicher Parameter in die Modellierung kritischer als das AIC. Da die Anzahl der Modellparameter im Prozessmodell gegenüber dem TAM ansteigt, wurde das BIC gewählt, da es im Vergleich zum AIC konservativ zuungunsten des Prozessmodells entschied.

Modellvergleich

Von den erhobenen 84 Datensätzen waren zwei nicht vollständig und wurden deshalb aus der Auswertung ausgeschlossen. Die vollständigen Datensätze wurden zur Schätzung der Modellparameter genutzt. Die Kennwerte verschiedener Modellierungen sind in Tab 14 dargestellt.

Modell	(feste Effekte)	BIC	aufgeklärte Varianz
Basis	konstant	505,646	0,000 %
TAM	PU PEOU	433,628	39,050 %
Vollständiges Modell	PU PL PEUEU (PL*time) (PEUEU*time)	436,344	56,464 %
TAM mit alternativen Konstrukten	PU PL PEUEU	437,141	40,026 %
Prozessmodell	PU (PL*time) (PEUEU*time)	401,909	53,264 %
TAM mit Wechselwirkung	PU PEOU (PEOU*time)	437,222	39,670 %

Tab 14: Kennwerte verschiedener Modellierungen der Nutzungsintention I
(in allen Modellen ist als zufälliger Effekt der Einfluss der Versuchsperson enthalten)

Das *Basismodell* enthält keine unabhängigen Variablen zur Vorhersage der Nutzungsintention. Entsprechend wird die Nutzungsintention einfach durch ihren Mittelwert als besten Schätzwert vorhergesagt. Die durch diese Modellierung aufgeklärte Varianz beträgt entsprechend 0,000 % bei einem BIC von 505,642. Beim BIC wird eine schlechtere Anpassung der Modellierung an die Daten in höhere Kennwerte umgesetzt. Die Angabe der aufgeklärten Varianz erfolgt für alle Modellierungen ohne den zufälligen Effekt zur Modellierung des Einflusses der Versuchsperson. Dieser Effekt wird nicht berichtet, da er nur in die Modellierung aufgenommen werden muss, um die Messwiederholung statistisch zu berücksichtigen, aber an sich nicht von Interesse ist. Das Basismodell wurde als Vergleichsgrundlage für die anderen Modelle berechnet. Inhaltlich entspricht das Basismodell einer Modellierung, bei der keinerlei Vorhersage der Nutzungsintention auf Basis der gemessenen unabhängigen Variablen möglich ist.

Die zweite Modellierung mit zwei Haupteffekten entspricht dem *TAM*. Die beiden Haupteffekte sind der Einfluss von PU und PEOU auf die Nutzungsintention. Dieses Modell bildet die erhobenen Daten besser ab als das *Basismodell*, was durch einen geringeren BIC von 434,669 angezeigt wird. Die durch das *TAM* erklärte Varianz beträgt 39,050 % der gesamten Varianz der Daten. Eine Visualisierung der Güte der Modellanpassung kann man über ein Diagramm erzielen, in dem die vorhergesagten Werte gegen die empirisch beobachteten Werte aufgetragen werden. Abb. 49 zeigt eine solche Darstellung für die Modellierung der Daten durch das *TAM*.

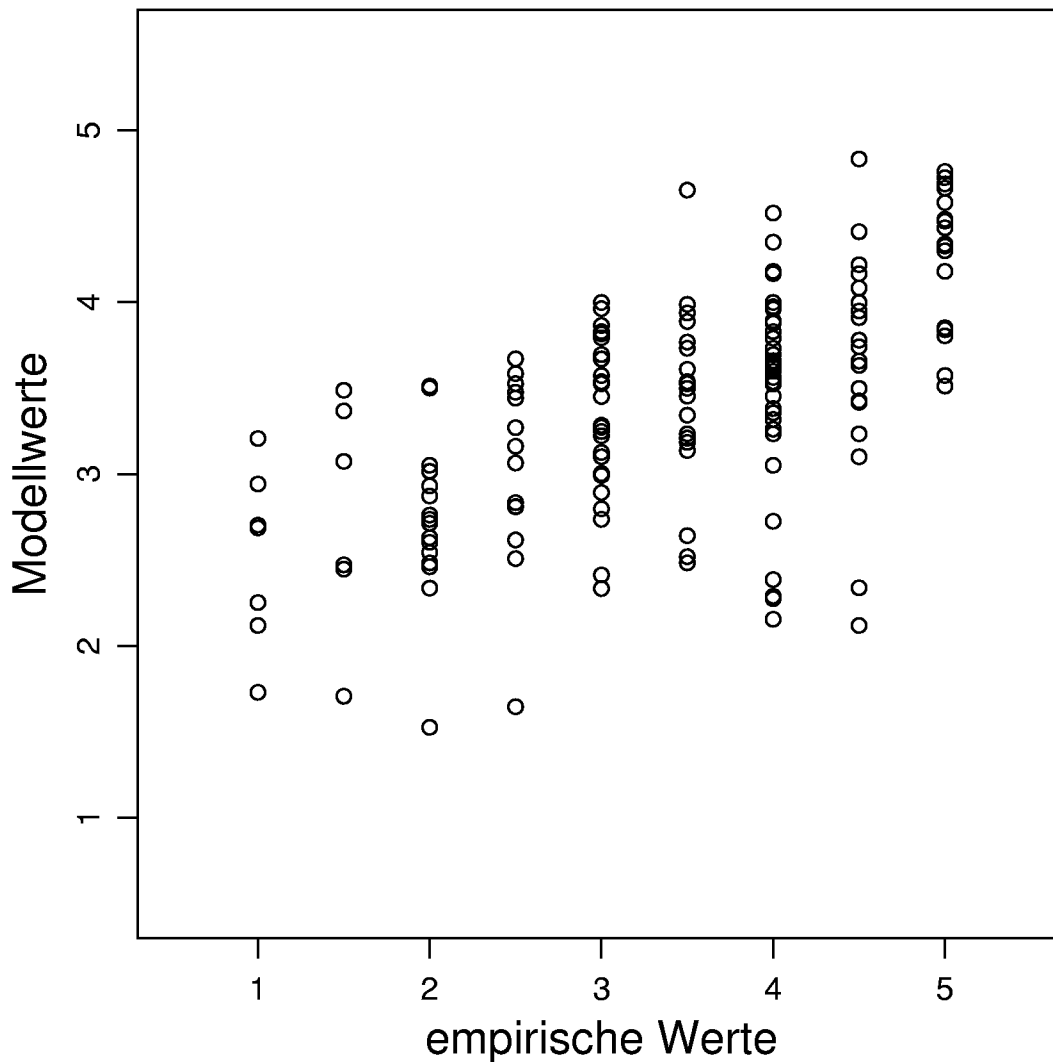


Abb. 49: Darstellung der Nutzungsintention zur Beurteilung der Anpassungsgüte (beobachtete vs. durch das TAM vorhergesagte Werte)

Abb. 49 stellt auf der Ordinate die durch das TAM vorhergesagten Werte der Nutzungsintention dar. Auf der Abszisse sind die in der Studie beobachteten Werte dargestellt. Da die Achsen den gleichen Maßstab haben, liegen perfekt vorhergesagte Werte auf der Diagonalen von links unten nach rechts oben. Je weiter die Punkte vertikal von der Diagonalen abweichen, desto größer ist die Abweichung der vorhergesagten von den beobachteten Werten. Für eine gute Modellierung ergibt sich eine Punkteschar nahe der Diagonalen. Die Stufen der empirischen Werte ergeben sich durch die Skalenbildung für die Nutzungsintention, bei der zwei fünfstufige Items gemittelt wurden.

In der dritten Modellierung wurde das TAM-Konstrukt PEOU durch die beiden neuen Konstrukte PL und PEUEU ersetzt. Gleichzeitig wurden für diese beiden Konstrukte Interaktionseffekte mit dem Messzeitpunkt aufgenommen. Der Messzeitpunkt entspricht durch das Versuchsdesign der Erfahrung der Benutzer. Diese Modellierung wurde als *vollständiges Modell* bezeichnet. Das *vollständige Modell* erzielt eine höhere Varianzaufklärung als das TAM. Sie steigt durch die

veränderte Modellierung auf 56,464 % an. Für das *vollständige Modell* ergibt sich ein BIC von 436,344. Das vollständige Modell wurde gerechnet, weil es eine Obergrenze für die erreichbare Prognosegüte der Nutzungsintention anhand der erhobenen unabhängigen Variablen auf Kosten hoher Modellkomplexität darstellt. Ein Interaktionseffekt zwischen dem Messzeitpunkt und dem TAM-Konstrukt PU wurde – entgegen der Bezeichnung *vollständiges Modell* – nicht aufgenommen, da es aus bisherigen Studien zum TAM keine Hinweis auf das Vorliegen einer solche Interaktion gibt.

In einer weiteren Modellierung wurden gegenüber dem *vollständigen Modell* die Interaktionseffekte entfernt. Dieses Modell entspricht dem TAM, wobei das Konstrukt PEOU durch die neuen Konstrukte PL und PEUEU ersetzt wurde. Dieses, mit TAM *mit alternativen Konstrukten* bezeichnete Modell kann 40,026 % der Varianz aufklären. Das BIC dieses Modells beträgt 437,141. Die Berechnung des TAM *mit alternativen Konstrukten* erfolgte zur Quantifizierung der Verbesserung der Modellierung durch die Differenzierung des TAM-Konstrukts PEOU in die neuen Konstrukte PL und PEUEU.

Die fünfte Modellierung mit einem Haupteffekt des Konstrukts PU und zwei Wechselwirkungen ist die in dieser Arbeit vorgeschlagene Veränderung des TAM zum *Prozessmodell*. Die beiden Wechselwirkungen stellen die Interaktion der Konstrukte PL und PEUEU mit dem Messzeitpunkt dar. Der Messzeitpunkt ist stellvertretend für die Erfahrung der Nutzer zu betrachten. Somit würde sich eine Interaktion von PL und PEUEU zur Erfahrung der Benutzer ergeben. Diese Modellierung weist ein BIC von 401,909 auf und erklärt 53,264 % der gesamten Varianz. Abb. 50 zeigt eine Darstellung der beobachteten Werte gegenüber den durch das *Prozessmodell* vorhergesagten Werte.

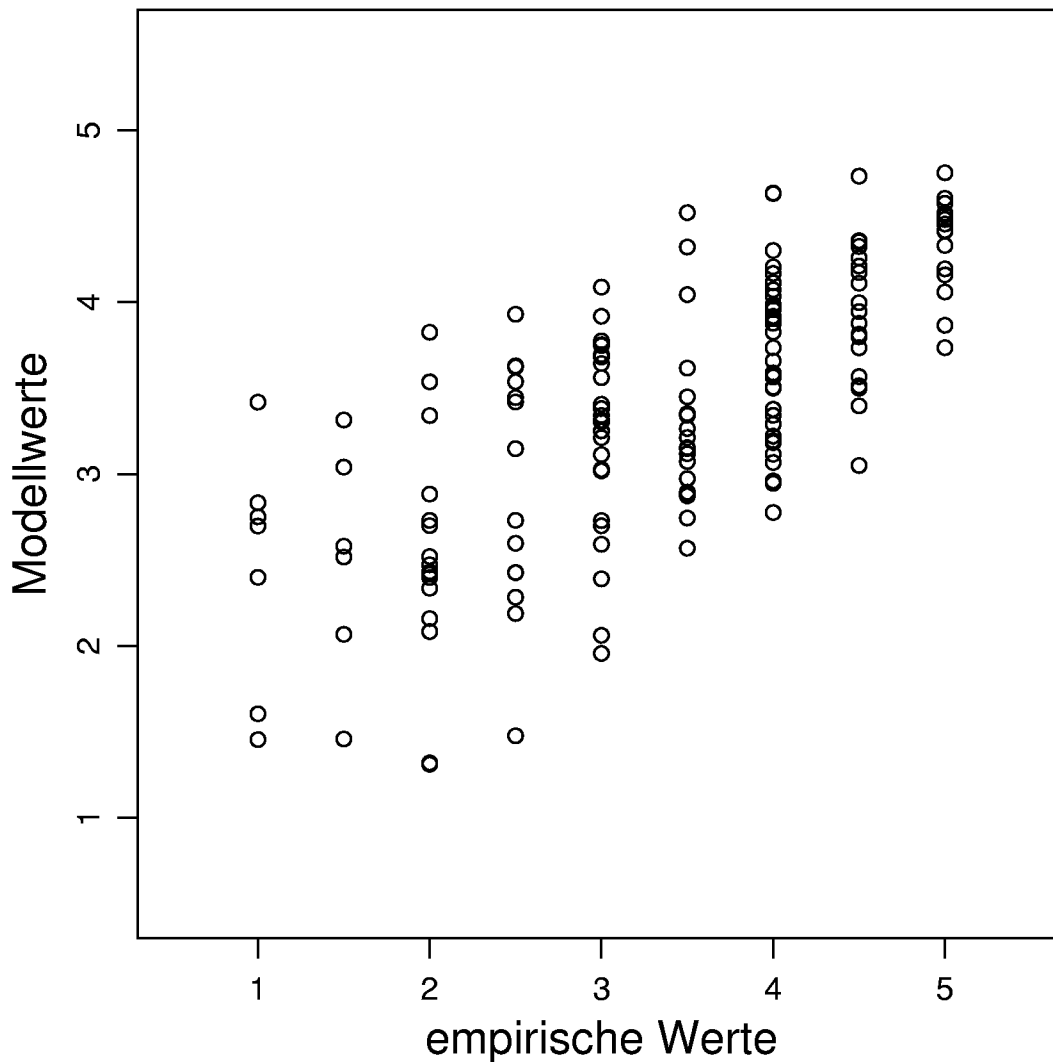


Abb. 50: Darstellung der Nutzungsintention zur Beurteilung der Anpassungsgüte (beobachtete vs. durch das Prozessmodell vorhergesagte Werte)

Wie in Abb. 49 für das TAM, werden in Abb. 50 die durch das *Prozessmodell* vorhergesagten Werte graphisch in Bezug zu den beobachteten Werten gesetzt. Die vorhergesagten Werte der Nutzungsintention sind auf der Ordinate, die beobachteten Werte auf der Abszisse dargestellt.

In der sechsten Modellierung wurde versuchsweise analysiert, wie sich die Aufnahme der Wechselwirkung von PEOU mit der Erfahrung ins *TAM* auswirkt. Die Berechnung dieses Modells diente der Überprüfung, ob die verbesserte Anpassung des *Prozessmodells* allein auf die Aufnahme eines Interaktionseffekts mit der Benutzererfahrung zurückgeführt werden konnte. Diese, mit *TAM mit Wechselwirkung* bezeichnete Modellierung kann 39,670 % der Varianz der Daten erklären. Damit erhöht sich – wie durch die Aufnahme eines zusätzlichen Modellparameters generell zu erwarten – die aufgeklärte Varianz gegenüber dem *TAM*. Das BIC des *TAM mit Wechselwirkung* beträgt 437,222.

Das vorgeschlagene *Prozessmodell* mit einem Haupteffekt von PU und zwei Interaktionseffekten zwischen PL bzw. PEUEU und dem Messzeitpunkt weist eine

deutlich höhere Varianzaufklärung als das TAM auf. Auch der zugehörige Parameter für die Beurteilung der Güte der Modellanpassung BIC ist mit 401,909 gegenüber 434,669 für das TAM deutlich kleiner. Kleinere Werte des BIC zeigen eine bessere Anpassung einer Modellierung an.

Die statistische Bedeutsamkeit der Verbesserung der Modellanpassung kann durch einen Likelihood-Ratio-Test überprüft werden (Davison, 2003, S.126-140). Ein Vergleich des TAM mit dem *Prozessmodell* ergibt mit einem Quotienten der Likelihoods von 42,959 ($p < 0,01$) einen signifikanten Unterschied der Anpassungsgüte der beiden Modelle. Da zusätzlich das BIC des *Prozessmodells* kleiner ist als das BIC des TAM, bildet das *Prozessmodell* die erhobenen Daten signifikant besser ab als das TAM.

Abgesehen vom *Basismodell*, in dem keine unabhängigen Variablen enthalten sind, und dem *Prozessmodell* sind alle Unterschiede in der Abbildungsgüte zwischen den übrigen Modellen (*TAM*, *vollständiges Modell*, *TAM mit alternativen Konstrukten* und *TAM mit Wechselwirkung*) nicht signifikant. Das *vollständige Modell* erhöht die aufgeklärte Varianz gegenüber dem TAM von 39,050 % auf 56,464 %. Diese Verbesserung der Vorhersagequalität wird durch die Erhöhung der Anzahl der Modellparameter von zwei auf fünf erreicht. Das BIC dieser beiden Modellierungen unterscheidet sich mit 434,669 zu 436,344 nur unwesentlich. Das Eliminieren der Interaktionseffekte des vollständigen Modells führt zum *TAM mit alternativen Konstrukten*. Gegenüber dem *vollständigen Modell* sinkt die Anzahl der Modellparameter auf drei (für die Haupteffekte) ab. Die durch das Modell erklärte Varianz verringert sich gegenüber dem *vollständigen Modell* auf 40,026 % bei einem BIC von 437,141. Damit erreicht auch das *TAM mit alternativen Konstrukten* eine ähnliche Anpassungsgüte wie das TAM (BIC=434,669), ist aber komplexer. Auch das *TAM mit Wechselwirkung*, bei dem das TAM um einen Interaktionseffekt zwischen PEOU und dem Messzeitpunkt erweitert wurde, verbessert die Modellanpassung nicht wesentlich. Die aufgeklärte Varianz steigt zwar gegenüber dem TAM von 39,050 % auf 39,670 % an, jedoch verschlechtert sich zugleich das BIC aufgrund der erhöhten Modellkomplexität von 434,669 auf 437,222.

Das *Prozessmodell* mit einem Haupteffekt von PU und den beiden Interaktionseffekten zwischen PL bzw. PEUEU und dem Messzeitpunkt führt zu einer insgesamt verbesserten Abbildungsgüte. Die aufgeklärte Varianz steigt von 39,050 % auf 53,264 % an. Die Anzahl der Modellparameter erhöht sich von zwei auf drei. Das BIC sinkt von 434,669 auf 401,909 ab, wobei niedrigere Werte einer besseren Anpassung entsprechen. Dieser Unterschied ist in einem Likelihood-Ratio-Test mit einem Quotienten von 42,959 statistisch signifikant.

Analyse des Prozessmodells

Das Prozessmodell erzielt eine signifikant bessere Modellierung der empirischen Daten zur Nutzungsintention. Hierfür kommt ein Modellparameter mehr im Vergleich

zum TAM zum Einsatz. Im folgenden werden die Parameter der Modellierung eingehend betrachtet.

Wie schon im TAM ist auch im Prozessmodell PU ein wichtiger Prädiktor der Nutzungsintention. PU hat einen Haupteffekt auf die Nutzungsintention, ist also von der Erfahrung des Benutzers unabhängig. Graphisch ist der Zusammenhang zwischen PU und der vorhergesagten Nutzungsintention in Abb. 51 dargestellt.

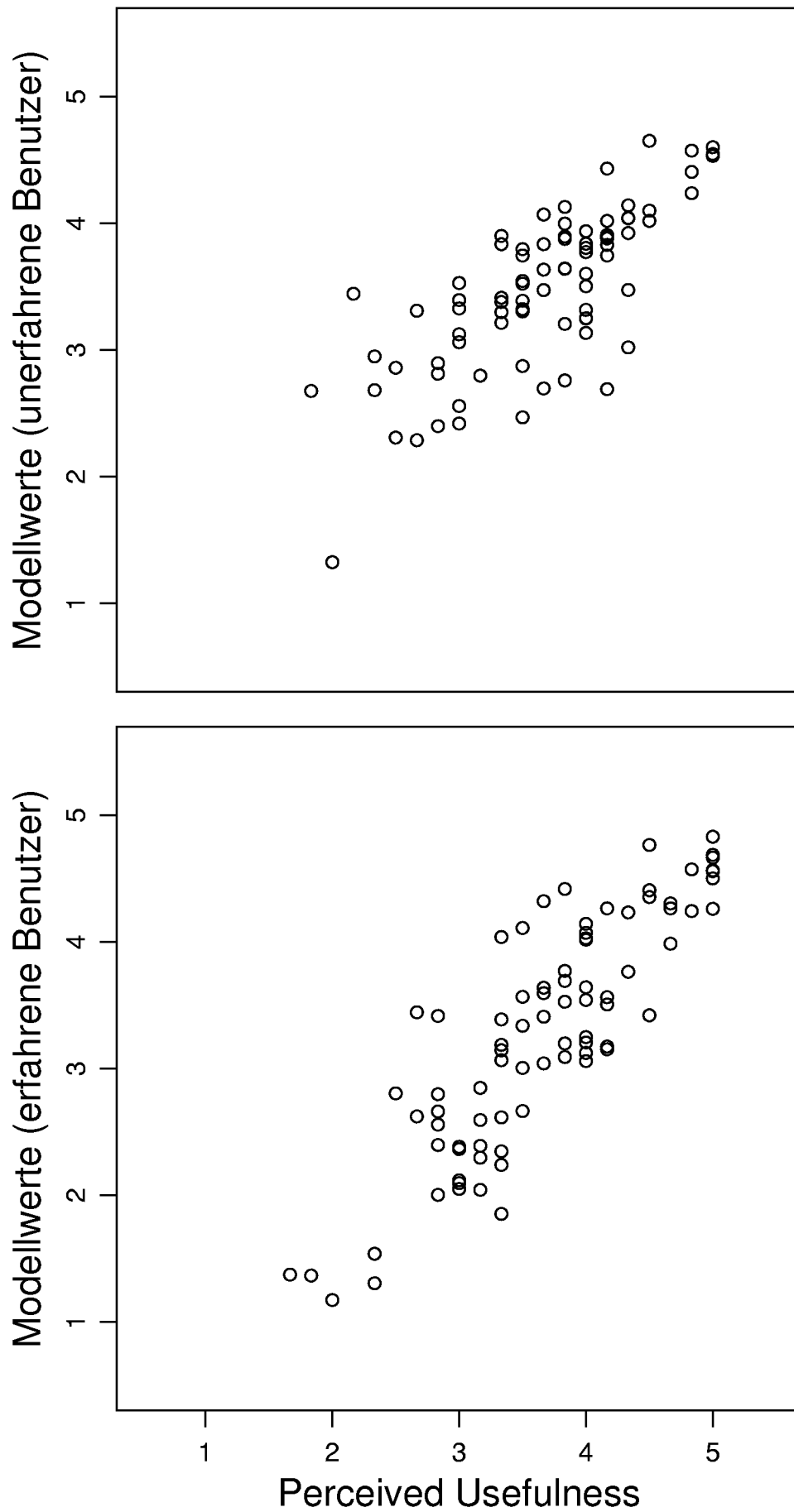


Abb. 51: Einfluss von PU auf die vorhergesagte Nutzungsintention

Auf den Ordinaten sind in Abb. 51 die durch das Prozessmodell für die Nutzungsintention vorhergesagten Werte dargestellt. Dabei sind die Daten für den ersten Messzeitpunkt (unerfahrene Benutzer) und den zweiten Messzeitpunkt (erfahrene Benutzer) in jeweils eigenen Diagrammen dargestellt. Diese Darstellungsart wurde gewählt, um die Vergleichbarkeit zu den nachfolgenden Diagrammen für PL und PEUEU sicher zu stellen. In diesen Diagrammen muss zwischen den beiden Messzeitpunkten unterschieden werden, da es sich für PL und PEUEU um Interaktionseffekte handelt.

Die Abszisse stellt die für PU gemessenen Werte dar. In beiden Diagrammen wurde für die Achsen der gleiche Maßstab gewählt. Wäre PU der einzige Prädiktor einer perfekten Modellierung, so müssten die vorhergesagten Werte auf einer Geraden liegen. Die Abweichungen dieser Geraden kommen durch den Einfluss der zusätzlichen Prädiktoren zustande. Je stärker der Einfluss von PU auf die vorhergesagten Modellwerte ist, desto näher liegen die Punkte in den Diagrammen an dieser gedachten Geraden.

Der Regressionskoeffizient (β -Gewicht) von PU im Prozessmodell beträgt 0,559. Dieser Regressionskoeffizient trägt mit einem t-Wert von 5,160 ($p < 0,001$; $df = 77$) hochsignifikant zur Vorhersage der Nutzungsintention bei.

Neben dem Haupteffekt von PU wird die Nutzungsintention im Prozessmodell durch Wechselwirkungen zwischen der Erfahrung des Benutzers und PL bzw. PEUEU vorhergesagt. Abb. 52 zeigt den Zusammenhang zwischen PL und den vorhergesagten Werten für die Nutzungsintention.

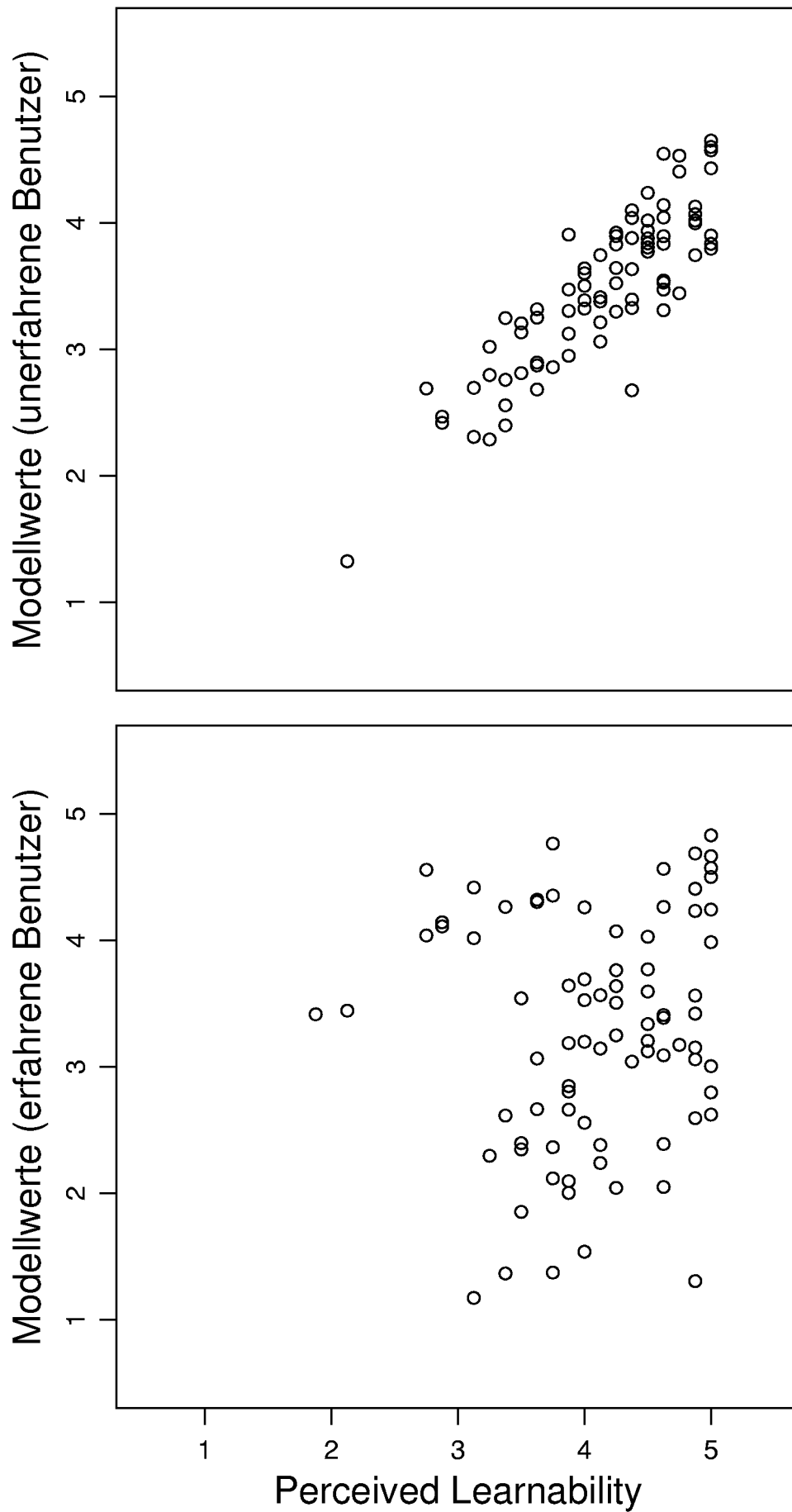


Abb. 52: Einfluss von PL auf die vorhergesagte Nutzungsintention

Wie in Abb. 51 sind auch in Abb. 52 die durch das Prozessmodell vorhergesagten Werte auf den Ordinaten der Diagramme dargestellt. Auf der Abszisse sind die gemessenen Werte für PL abgetragen. Im oberen Diagramm sind die Werte für unerfahrene Benutzer (erster Messzeitpunkt) dargestellt. Das untere Diagramm zeigt die Werte für erfahrene Benutzer (zweiter Messzeitpunkt).

Je stärker der Zusammenhang zwischen PL und der Nutzungsintention ausfällt, desto mehr ordnen sich die Punkte entlang einer Geraden an, die den Zusammenhang modelliert. Die Einflüsse der weiteren Modellparameter führen zu Abweichungen von dieser gedachten Geraden.

Im oberen Diagramm – also für unerfahrene Benutzer – ordnet sich die Punkteschar erkennbar entlang einer ansteigenden Geraden an. Hier ergibt sich ein ausgeprägter Zusammenhang zwischen PL und der Nutzungsintention. Mit einem ansteigenden Skalenwert für PL, also mit einer positiv wahrgenommenen Erlernbarkeit, steigt auch die Nutzungsintention an. Der Regressionskoeffizient für diesen Zusammenhang beträgt 0,672 und ist mit einem t-Wert von 5,577 ($p < 0,001$; $df = 77$) hochsignifikant.

Im unteren Diagramm ist der Zusammenhang für erfahrene Benutzer, also zum zweiten Messzeitpunkt, dargestellt. Hier bilden die Punkte eine Wolke, in der visuell keine Systematik zu erkennen ist. Der Regressionskoeffizient zwischen PL und der Nutzungsintention beträgt für erfahrene Benutzer -0,172. Diesem β -Gewicht entspricht ein t-Wert von -1,557, womit dieser Zusammenhang nicht signifikant zur Vorhersage der Nutzungsintention beiträgt ($p = 0,124$; $df = 77$).

Der dritte Prädiktor für die Nutzungsintention im Prozessmodell ist der Interaktionseffekt zwischen der Erfahrung der Benutzer und PEUEU. Dieser Zusammenhang ist in Abb. 53 graphisch dargestellt.

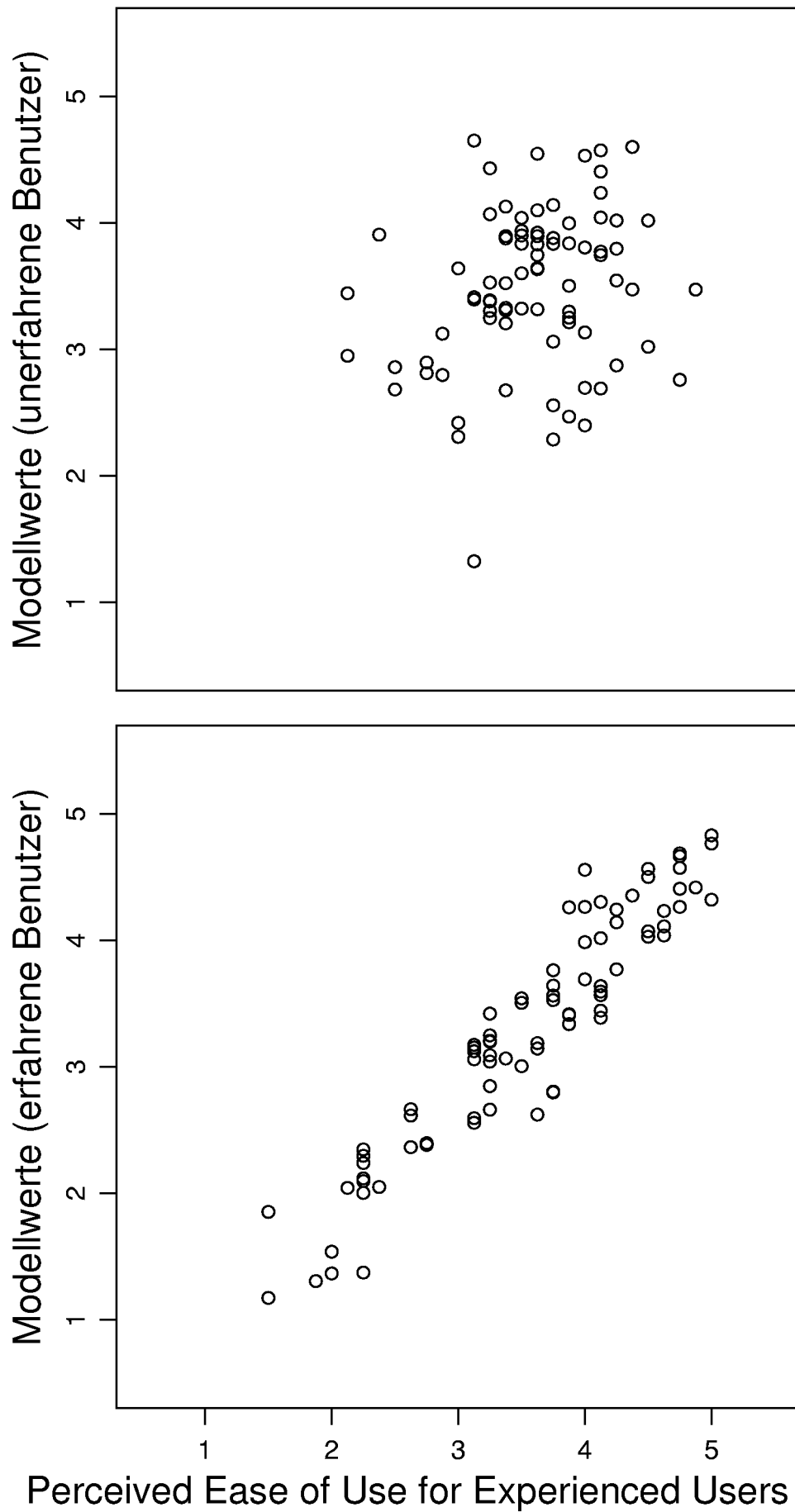


Abb. 53: Einfluss von PEUEU auf die vorhergesagte Nutzungsintention

In Abb. 53 sind – wie schon in Abb. 51 und Abb. 52 – die vorhergesagten Werte des Prozessmodells in beiden Diagrammen jeweils auf der Ordinate abgetragen. Entlang der Abszisse sind die für PEUEU gemessenen Werte dargestellt. Das obere Diagramm zeigt die Werte für unerfahrene Benutzer (erster Messzeitpunkt). Die Werte für erfahrene Benutzer (zweiter Messzeitpunkt) sind im unteren Diagramm dargestellt.

In dieser Darstellung zeigt sich ein deutlicher Zusammenhang zwischen PEUEU und den durch das Modell vorhergesagten Werten dadurch, dass sich die Punkte entlang einer gedachten Linie anordnen. Je stärker der Einfluss des Konstrukts PEUEU auf die vorhergesagte Nutzungsintention, desto stärker ordnen sich die Punkte in den Diagrammen entlang einer Linie an.

Im oberen Diagramm, also für unerfahrene Benutzer, bilden die Punkte eine Wolke. Ein eindeutiger Zusammenhang zwischen PEUEU und der Nutzungsintention ist nicht zu erkennen. Die statistischen Kennwerte für den Regressionskoeffizienten bestätigen den optischen Eindruck. Das β -Gewicht für PEUEU beträgt zum ersten Messzeitpunkt $-0,264$. Dieser Zusammenhang wird mit einem t-Wert von $-1,798$ statistisch nicht signifikant ($p=0,076$; $df=77$).

Im unteren Diagramm ist der Zusammenhang zwischen PEUEU und der Nutzungsintention für erfahrene Benutzer dargestellt. Hier ist zu erkennen, dass höhere Werte von PEUEU zu einer höheren Nutzungsintention führen. Der Regressionskoeffizient dieses Zusammenhangs beträgt $0,659$. Der Einfluss von PEUEU auf die Nutzungsintention erfahrener Benutzer ist damit hochsignifikant ($t=6,024$; $p<0,000$; $df=77$).

Zusammenfassung

Das postulierte Prozessmodell modelliert die Nutzungsintention signifikant besser als das TAM. Die durch das Modell aufgeklärte Varianz steigt von $39,050\%$, die durch das TAM erklärt werden, auf $53,264\%$ an. Diese Steigerung der Abbildungsgüte rechtfertigt die Erhöhung der Anzahl der Modellparameter von drei auf fünf.

Das Prozessmodell übernimmt das Konstrukt PU unverändert aus dem TAM. Zusätzlich wird die Nutzungsintention durch zwei Wechselwirkungen vorhergesagt. In beiden Wechselwirkungen spielt die Erfahrung eines Nutzers mit der Software eine Rolle. Bei unerfahrenen Benutzern hat die wahrgenommenen Erlernbarkeit (PL) einen signifikanten Einfluss auf die Nutzungsintention. Für erfahrene Benutzer wird dieser Zusammenhang unbedeutend. Umgekehrt verhält es sich für PEUEU. Dieses Konstrukt hat bei unerfahrenen Benutzern keinen bedeutsamen Einfluss auf die Nutzungsintention, wird aber für erfahrene Benutzer relevant.

Diskussion

Ziel dieser Arbeit war es, die Dynamik der Bewertung moderner Informationsverarbeitungstechnologien in Abhängigkeit von der Erfahrung des Benutzers zu erforschen. Hierzu wurde Davis (1989) TAM zu einem Prozessmodell weiterentwickelt. Im folgenden werden die Ergebnisse vor dem Hintergrund der für diese Arbeit formulierten Hypothesen betrachtet. Weiterhin wird diskutiert, inwiefern eine Verallgemeinerung der Ergebnisse dieser Arbeit an Grenzen stößt.

Die Ergebnisse im Hinblick auf die Hypothesen

Die ersten beiden Studien, in denen die Meßskalen für die neuen Konstrukte des Prozessmodells entwickelt wurden, bereiteten die dritte Untersuchung vor, die der Überprüfung der Hypothesen dieser Arbeit diene.

In der ersten Studie wurden die Items sowohl nach inhaltlicher Ähnlichkeit als auch nach deren Übereinstimmung mit den Konstruktdefinitionen sortiert. Die Auswertung dieser Studie resultierte in einer inhaltlichen Gliederung der Items. Diese bestätigte die Annahme der beiden getrennten Konstrukte PL und PEUEU und diene einer ersten Auswahl geeigneter Items.

In der zweiten Studie wurden die Skalen zur Messung von PL und PEUEU zweier Online-Routenplaner eingesetzt. Die dabei erhobenen Daten ermöglichten eine parametrische Analyse der Item- und Skaleneigenschaften. Wie schon die erste Studie zeigte auch diese Untersuchung, dass die Annahme der beiden getrennten Konstrukte PL und PEUEU statt des Konstrukts PEOU sinnvoll war. Eine Betrachtung der Interkorrelationsmatrix der Items zeigte zufriedenstellend hohe Korrelationen innerhalb der Items eines Konstrukts und niedrige Korrelationen zu Items der anderen Skala. Insgesamt bestand der Schwerpunkt der beiden ersten Studien in der Entwicklung der benötigten neuen Meßskalen und nicht auf der Überprüfung der Hypothesen dieser Arbeit.

In der dritten Studie wurden alle für die Nutzungsintention relevanten Variablen an unerfahrenen und erfahrenen Benutzern eines Online-Routenplaners erhoben. Diese Daten dienten der Schätzung von Modellparametern für die Modellierung der Nutzungsintention auf Basis der unabhängigen Variablen. Hierdurch wird eine Vorhersage der Nutzungsintention anhand der unabhängigen Variablen möglich.

Es ergaben sich deutliche Unterschiede zwischen unterschiedlichen Modellierungen zur Erklärung der Nutzungsabsicht. Bei einer solchen Modellierung gilt es stets abzuwägen zwischen einem möglichst einfachen Modell und einer möglichst hohen Anpassung des Modells an die empirisch erhobenen Daten. Das beste Modell ist also nicht jenes, welches die Daten absolut am besten widerspiegelt, sondern das Modell, welches einen guten Kompromiss zwischen der Komplexität des Modells und der Abbildungsgüte erzielt.

Dieser Kompromiss gelang dem in dieser Arbeit vorgeschlagenen Prozessmodell

am besten. Nach diesem Modell hängt die Absicht, eine Informations- oder Kommunikationstechnologie zu benutzen, zunächst von deren wahrgenommener Nützlichkeit ab. Dieser Zusammenhang gilt gleichermaßen für alle Benutzer. Zusätzlich zeigen sich zwei weitere Abhängigkeiten, die von der Erfahrung des Benutzers abhängen. Für unerfahrene Benutzer zeigt sich ein positiver Zusammenhang zwischen der Erlernbarkeit und der Nutzungsintention. Eine Software wird also eher eingesetzt, wenn sie leicht zu erlernen ist. Für erfahrene Benutzer verliert dieser Zusammenhang an Bedeutung und der Einfluss der Erlernbarkeit wird abgelöst durch den zunehmenden Einfluss der Benutzbarkeit, wie sie sich einem erfahrenen Benutzer darstellt. Es ergibt sich also eine Abhängigkeit der Nutzungsintention von den beiden neuen Konstrukten PL und PEUEU, die systematisch mit der Nutzungserfahrung der Benutzer variieren.

Die Nutzungserfahrung wurde in der Studie durch den Messzeitpunkt gezielt variiert. Da die eingesetzte Software allen Probanden der Studie unbekannt war, konnten die Ergebnisse der ersten Messung als Bewertung durch unerfahrene Benutzer aufgefasst werden. Für diesen Zeitpunkt zeigte sich ein deutlicher Einfluss der PL auf die Nutzungsintention. Zusätzlich wurde die Nutzungsintention von PU bestimmt, während PEUEU nur einen geringen Einfluss auf die Nutzungsintention ausübte.

Nach dem die Probanden durch Bearbeiten mehrerer Aufgaben mit zunehmenden Schwierigkeitsgrad Erfahrung mit der eingesetzten Software erlangt hatten, wurde bei der zweiten Messung am Ende der Untersuchung die Bewertung durch erfahrene Benutzer erhoben. Auch zu diesem Messzeitpunkt hatte PU einen signifikanten Einfluss auf die Nutzungsintention. Bezüglich der beiden neuen Konstrukte PL und PEUEU fand ein Wechsel des Einflusses statt: PL beeinflusste die Nutzungsintention nur noch in geringem Maß, während PEUEU einen bedeutsamen Einfluss auf die Nutzungsintention ausübte.

Diese Befunde bestätigen die grundlegenden Annahmen dieser Arbeit, dass die Bewertung einer Software sich mit der Erfahrung eines Benutzers verändert. Für unerfahrene Benutzer spielt die Erlernbarkeit eine wichtige Rolle. Dieser Einfluss nimmt mit zunehmender Erfahrung ab. Gleichzeitig gewinnt die Wahrnehmung der Bedienbarkeit, wie sie sich einem erfahrenen Benutzer darstellt, an Bedeutung.

Das in dieser Arbeit vorgeschlagene Prozessmodell kann die erhobenen Daten besser abbilden als das TAM von Davis (1989). Die aufgeklärte Varianz erhöht sich auf 53,264 %, was eine statistisch signifikante Verbesserung der Modellanpassung um 14,214 Prozentpunkte darstellt.

Die beiden neuen Konstrukte PL und PEUEU ermöglichen ein genaueres Verständnis der Prozesse, die zur Bildung einer Nutzungsintention führen. Für die beiden neuen Konstrukte wurden zwei getrennte Skalen konstruiert. Im Prozess der Skalenentwicklung bestätigte sich die angenommene inhaltliche Trennung in die neuen Konstrukte PL und PEUEU.

Durch die differenziertere Betrachtung der Aspekte der Benutzbarkeit des Prozessmodells gegenüber dem TAM ergibt sich eine deutlich verbesserte Vorhersage der Nutzungsintention. Die Annahme zweier unabhängiger Aspekte der Benutzbarkeit hat sich somit für die Modellierung der Nutzungsintention bewährt und trägt zu einem verfeinerten Verständnis der Ausbildung von Technologieakzeptanz bei. Inhaltlich erfassen die beiden Konstrukte gemeinsam Davis Konstrukt PEOU, trennen dabei aber zwischen den beiden unterschiedlichen Prozessen.

Validität der Arbeit

Eine häufige Problematik humanwissenschaftlicher Untersuchungen besteht in der Frage, inwieweit die Ergebnisse einer Studien generalisiert, also auf andere Situationen übertragen, werden können. Die konkreten Ergebnisse einer Studie sind nur insofern von übergeordneter Bedeutung, als sie Rückschlüsse auf allgemeine Regeln ermöglichen. Inwiefern dies für die Ergebnisse der vorliegenden Arbeit zutrifft und welche Einschränkungen dabei zu berücksichtigen sind, wird im Folgenden diskutiert.

Validität der Skalen

Die Konstrukte psychologischer Theorien sind meist nicht direkt beobachtbar. Deshalb kommen in psychologischen Untersuchungen Verfahren zum Einsatz, die eine indirekte Messung dieser Konstrukte erlauben. In der vorliegenden Arbeit wurden fünfstufige Antwortskalen für die Messung der Konstrukte des Prozessmodells entwickelt. Die Qualität der Aussagen über die Konstrukte einer psychologischen Theorie hängt direkt von der Güte deren Messung ab. Eine unzuverlässige Messung der Konstrukte verhindert validen Aussagen über die zugrunde liegende Theorie.

Validität der neuen Skalen für PL und PEUEU

Die Entwicklung der Skalen zur Messung der beiden neuen Konstrukte erfolgte in einem mehrstufigen Prozess. Ziel dieses Vorgehens war die Konstruktion reliabler Skalen, um die neu definierten Konstrukte des Prozessmodells zu erfassen. Hierzu wurde zunächst eine inhaltlich breite Itemsammlung erstellt, die möglichst alle denkbaren Aspekte der Konstrukte erfassen sollte. Anschliessend erfolgte eine zweistufige Auswahl der geeignetsten Items auf der Grundlage zweier Studien. Die Auswahl erfolgte nach verschiedenen statistischen Parametern, wobei stets inhaltliche Aspekte mitberücksichtigt wurden. Dieses Vorgehen wurde von Anastasi (1986) für die Entwicklung reliabler Skalen mit einer hohen Validität empfohlen.

Zunächst wurde auf der Grundlage einer detaillierten Betrachtung ähnlicher Konzepte Definitionen für die beiden Konstrukte formuliert. Um einen Überblick über für diese Definitionen relevante Eigenschaften einer Software zu erhalten, wurden 12

Personen befragt. Die resultierende Sammlung relevanter Eigenschaften bildete zusammen mit der Analyse ähnlicher Skalen die Grundlage für die Formulierung des ursprünglichen Itempools für die beiden neuen Skalen. Clark und Watson (1995) betonen, dass dieser ursprüngliche Itempool bereits die maximal erreichbare Qualität einer Skala determiniert. Da der weitere Prozess der Skalenkonstruktion keine Methode zur systematischen Ergänzung des Itempools durch zusätzliche Items vorsieht, bleiben Aspekte, die die ursprüngliche Itemsammlung nicht abdeckt, unberücksichtigt.

Die für die Sammlung der Aspekte befragten Personen hatten alle Erfahrung mit der Benutzung von Computern. Insofern sind die Vorstellungen von Personen ohne jegliche Erfahrung im Umgang mit Computer nicht berücksichtigt. Das Alter der Befragten bewegte sich zwischen 28 und 59. Es sind also keine Schüler und auch keine Rentner berücksichtigt worden. Alle befragten Personen nutzten Computer beruflich.

Es ist nicht auszuschließen, dass junge Personen wie Schüler oder Studenten den Lernprozess moderner Informationsverarbeitungssysteme anderes erleben, als die befragten Personen. Auch für ältere Personen sind diesbezüglich systematische Unterschiede vorstellbar. Ob die in dieser Arbeit entwickelten Skalen die Konstrukte auch für Altersgruppen, die nicht in der Befragung berücksichtigt wurden, messen können, ist nicht sichergestellt.

An der zweiten Studie zur Auswahl geeigneter Items nahmen Psychologiestudenten als Probanden teil. Dies erfolgte aus praktischen Gründen, beschränkt jedoch die Validierung streng genommen auf die Population der Psychologiestudenten. Praktisch kann die Skala auch für vergleichbare Populationen eingesetzt werden. Für abweichende Populationen, also beispielsweise Personen mit einer geringeren formalen Bildung, muss die Itemauswahl neu überprüft werden.

Generell sind die Größen der an der Skalenentwicklung beteiligten Untersuchungsstichproben als eher klein zu betrachten. Im Hinblick auf einen weiteren Einsatz der Skalen wäre es daher wünschenswert, wenn weitere Erhebungen mit den Skalen – vor allem auch an abweichenden Stichproben – durchgeführt würden, um die Validität der Skalen auf ein breiteres Fundament zu stellen.

Validität der Skala für PEOU

Die in dieser Arbeit eingesetzte Skala für PEOU entspricht nicht der von Davis entwickelten. Statt dessen wurden die beiden neu gebildeten Skalen für PL und PEUEU gemittelt. Dieses Vorgehen wurde aufgrund verschiedener Überlegungen gewählt.

Zum einen existiert keine ins Deutsche übersetzte Version der PEOU-Skala von Davis, die validiert wäre. Zwar wurden im deutschen Sprachraum bereits Untersuchungen zum TAM durchgeführt, hierbei kamen allerdings durchweg nicht validierte Skalen zum Einsatz (z. B. Mahlke, 2005). Zusätzlich wurden in den meisten

Untersuchungen – auch bei Mahlke (2005) – einzelne Items verändert oder ausgetauscht, ohne anschließende Überprüfung der vorgenommenen Veränderungen. Für den Einsatz in dieser Arbeit erschien eine nicht validierte, übersetzte Skala nicht geeignet. Eine Übersetzung mit anschließender Validierung der PEOU-Skala von Davis (1989) war im Rahmen dieser Arbeit nicht zu leisten.

Zum anderen erfassen die beiden neu entwickelten Skalen für PL und PEUEU die Aspekte des Konstrukts PEOU in differenzierter Form. Die beiden Konstrukte PL und PEUEU stellen eine feinere Aufgliederung des Konstrukts PEOU dar. Hierdurch konnte davon ausgegangen werden, dass bereits eine für die Messung von PEOU geeignete Auswahl an Items in Form der PL- und PEUEU-Items bestand.

Bezüglich dieser Itemauswahl gelten naturgemäß alle Vorbehalte bezüglich der Validität weiter, die im vorigen Kapitel schon im Hinblick auf die PL- bzw. die PEUEU-Skala genannt wurden. Zusätzlich ist keine Validierung der durch Mittelung gebildeten PEOU-Skala für das Konstrukt PEOU erfolgt. Implizit kann jedoch eine Validität dieser Skala entsprechend der PL- und PEUEU-Skalen angenommen werden.

Validität der Skalen für PU und der Nutzungsintention

Die eingesetzte Skala für das Konstrukt PU ist eine ins Deutsche übersetzte Version von Davis (1989) PU-Skala. Dabei mussten einigen Anpassungen an der Formulierung der Items vorgenommen werden.

Die bei Davis (1989) abgedruckten Skalen sind für eine prospektive Befragung im Futur formuliert. Für die vorliegende Arbeit mussten die Zeitform der Items ins Präsens geändert werden.

Zusätzlich musste der Kontext der Items angepasst werden. Davis hatte die Skalen für die Untersuchung der Computernutzung in einem beruflichen Umfeld formuliert. In der vorliegenden Arbeit wurden Online-Routenplaner eingesetzt. Entsprechend wurde auch hier eine Anpassung der Itemformulierungen vorgenommen.

Die PU-Skala kann aufgrund dieser Anpassungen nicht mehr als validiert betrachtet werden. Eine validierte deutschsprachige Version der PU-Skala existiert nicht. Die von Mahlke (2005) eingesetzte Version ist im Vergleich zum Original von Davis ebenfalls stark verändert und ebenfalls nicht validiert.

Der Einsatz einer unvalidierten Skala für das Konstrukt PU schränkt die Interpretierbarkeit der Ergebnisse bezüglich des Konstrukts PU ein. Dies erschien für diese Arbeit akzeptabel, da die Befunde hinsichtlich des Zusammenhangs zwischen PU und der Nutzungsintention eindeutig sind und dieser Zusammenhang nicht von zentraler Bedeutung für die Fragestellung dieser Arbeit war.

Repräsentativität der Stichprobe

Die Daten für den Vergleich des postulierten Prozessmodells mit dem TAM

wurden an 84 Psychologiestudenten erhoben. Diese Stichprobe repräsentiert in ihrer Zusammensetzung keinesfalls die Gesamtbevölkerung. Allerdings kann davon ausgegangen werden, dass Psychologiestudenten im Hinblick auf Computernutzung Personen mit einem vergleichbaren Bildungsniveau ähneln.

Das Alter der Probanden lag zwischen 18 und 36 Jahren. Ein Verallgemeinerung der Ergebnisse dieser Arbeit über diesen Altersbereich hinaus ist kritisch zu bewerten. Jüngere Personen wachsen mit einem anderen Verhältnis zu Informationsverarbeitungstechnologien auf. Auch für ältere Personen können sich systematische Abweichungen bei der Ausbildung einer Nutzungsintention ergeben. In anderen Studien zur Integration eines Alertness-Managements, die in unserer Arbeitsgruppe durchgeführt wurden, ist subjektiv der Eindruck entstanden, dass ältere Benutzer einer nicht optimal gestalteten Bedienoberfläche weniger tolerant begegnen als jüngere Benutzer. Was von jüngeren Probanden als Herausforderung im Sinne einer Problemlöseaufgabe betrachtet wurde, wurde hierbei häufig von älteren Benutzern als unzumutbar abgelehnt.

Die Probanden verfügten alle über mindestens drei Jahre Erfahrung mit dem Internet und entsprechend auch dem Computer. Inwieweit die Bewertungsprozesse von der allgemeinen Computervorerfahrung abhängen, wurde im Rahmen dieser Arbeit nicht betrachtet. Da sich allerdings eine Veränderung der Bewertungsprozesse in Abhängigkeit von der konkreten Erfahrung mit einer Software zeigt, muss auch eine allgemeine Abhängigkeit der Bewertungsmechanismen von der allgemeinen Erfahrung mit dem Computer oder vergleichbaren Technologien erwartet werden. Die Ergebnisse dieser Arbeit sind insofern nicht unmittelbar auf Personen ohne oder mit nur wenig Erfahrung im Umgang mit Computern übertragbar. Allerdings lässt sich auch für diese Personengruppe erwarten, dass der Einfluss der Erlernbarkeit mit zunehmender Erfahrung im Umgang mit der Software schwindet, während gleichzeitig der Einfluss der Benutzbarkeit zunimmt.

Repräsentativität der eingesetzten Software

In allen Studien dieser Arbeit wurden Online-Routenplaner als Software eingesetzt. Mit dem Web 2.0 erleben Online-Applikationen aktuell einen Aufschwung. Gegenüber einer Software, die unabhängig von einer Internetverbindung vollständig lokal läuft, ergeben sich für Online-Applikationen systematische Unterschiede.

Die Reaktion auf Benutzereingaben erfolgt bei Online-Applikationen häufig verzögert, je nachdem, ob die Verarbeitung lokal (beispielsweise durch JavaScript) oder auf dem zuständigen Server erfolgt. Bei klassische Programme wie beispielsweise einer Textverarbeitung treten solche Unterschiede nicht auf.

Die Ergebnisse dieser Arbeit sind in einem engen Sinn nur für Online-Routenplaner gültig. Es besteht allerdings kein Grund anzunehmen, ähnliche Software, also zunächst alle Online-Applikationen, die Nutzereingaben erfordern und entsprechende Ergebnisse zurückmelden, würde anders wahrgenommen oder

bewertet werden. Auch für lokal laufende Software werden sich wahrscheinlich keine Unterschiede im Einfluss der Konstrukte des Prozessmodells auf die Nutzungsintention ergeben.

In den im theoretischen Teil dieser Arbeit betrachteten Untersuchungen zum TAM zeigten sich keine systematischen Veränderungen des TAM in Abhängigkeit von der eingesetzten Software. Das TAM zeigt sich somit robust bezüglich Veränderungen der Art der eingesetzten Software. Da das Prozessmodell aus dem TAM weiterentwickelt wurde, lässt sich vermuten, dass auch das Prozessmodell nicht empfindlich auf die Art der untersuchten Software reagiert. Insofern wäre das Prozessmodell auch für Modellierung der Nutzungsintention von Desktop-Anwendungen oder ähnlicher Kommunikations- und Informationstechnologie geeignet. Dennoch empfiehlt sich gezielte Überprüfung dieser Annahme.

Operationalisierung der Erfahrung der Benutzer

In dieser Arbeit wurde auf eine direkte Messung der Erfahrung eines Benutzers im Umgang mit einer speziellen Software verzichtet. Eine solche Messung wäre sehr komplex und die Entwicklung geeigneter Messinstrumente war im Rahmen dieser Arbeit nicht möglich.

Die Erfahrung der Probanden wurde durch das Studiendesign gezielt manipuliert. Der eingesetzte Routenplaner war ein neu entwickelter Prototyp, den die Probanden nicht kennen konnten. Somit waren alle Probanden am Beginn der Studie unerfahrene Benutzer. Durch die Bearbeitung der gestellten Aufgaben wurden die Probanden dazu gebracht, mit dem Routenplaner zu arbeiten. Im Verhältnis zu der geringen Komplexität der Bedienoberfläche erschien die Arbeitsphase mit einer Dauer von 30 Minuten ausreichend lange bemessen, um zu einem erfahrenen Benutzer zu werden. Diese Annahme konnte im Rahmen dieser Arbeit nicht explizit überprüft werden.

Für eine Messung der Erfahrung eines Benutzers mit einer spezifischen Softwareanwendung stehen bislang keine Messinstrumente zur Verfügung. Zusätzlich stellt sich die Frage, ob eine objektive Messung oder eine subjektive Messung im Hinblick auf den Einfluss der Erfahrung im Rahmen des Prozessmodells sinnvoller wäre.

Eine objektive Messung könnte sich an den Möglichkeiten der Software orientieren. Die Erfahrung eines Benutzers könnte als Anteil der von einem Benutzer benutzten Möglichkeiten operationalisiert werden. Ein Problem dieser Art der Operationalisierung besteht darin, dass häufig mehrere Lösungswege für eine Aufgabenstellung bestehen.

So kann beispielsweise eine Reiseroute mit einem Zwischenstop mit Routenplanern, die dies unterstützen, in einem Durchgang geplant werden. Eine alternative Möglichkeit besteht darin, zunächst die Route vom Startort bis zum

Zwischenstop zu planen und anschliessend eine zweite Route vom Zwischenstop zum Endziel der Reise. Hierbei wäre zu beachten, dass die zweite Strategie eine größere Unabhängigkeit von den Möglichkeiten des eingesetzten Routenplaners bieten würde, also robuster gegenüber den Möglichkeiten der verfügbaren Software ist. Dies könnte durchaus ein Hinweis auf einen erfahrenen Benutzer sein. Eine solche Nutzungsstrategie wäre allerdings weniger effizient verglichen mit einem auf die spezifischen Möglichkeiten einer speziellen Software optimierten. Die Bewertung der Erfahrung müsste also berücksichtigen, ob ein Benutzer häufig verschiedenen Anwendungen zur Lösung desselben Aufgabentyps benutzt oder nicht. Insgesamt erscheint eine objektive Erfassung der Erfahrung eines Benutzers schwierig.

Eine subjektive Messung der Erfahrung eines Benutzers würde auf der Selbsteinschätzung eines Benutzers basieren. Stark vereinfacht würde ein Benutzer dann als erfahren gelten, wenn er sich selbst als erfahren einschätzt. Für eine Verwendung im Rahmen des Prozessmodells wie auch des TAM erscheint diese Art der Operationalisierung vorteilhaft, da in beiden Modellen die unabhängigen Konstrukte nicht in Bezug auf einen objektiven Maßstab, sondern als subjektive Wahrnehmungskorrelate definiert sind. Für das subjektive Erleben der Interaktion eines Benutzers mit einer Software ist es unerheblich, ob der Benutzer die Software sehr gut kennt oder ihm nur ein Bruchteil der Möglichkeiten der Software vertraut sind. Entscheidend ist vielmehr, ob ein Benutzer sich selbst noch am Anfang eines Lernprozesses sieht oder diesen als für sich eher abgeschlossen betrachtet.

In dieser Arbeit wurde die Erfahrung der Benutzer nicht gemessen, sondern durch das experimentelle Design manipuliert. Bedingt durch dieses Vorgehen konnten nur unerfahrene mit erfahrenen Benutzern verglichen werden; es gab nur zwei mögliche Ausprägungen der Benutzererfahrung. Der Erwerb von Erfahrung kann im Allgemeinen als kontinuierlicher Vorgang betrachtet werden. Ein verbessertes Verständnis des variablen Einflusses der beiden neuen Konstrukte PL und PEUEU erfordert eine fein abgestufte Erfassung der Erfahrung. Eine interessante Fragestellung wäre hierbei, ob die Veränderung des Einflusses der beiden Konstrukte mit zunehmender Erfahrung eher kontinuierlich verläuft oder ob es zu einem sprunghaften Wechsel kommt.

Die genannten Vorbehalte bezüglich der Generalisierbarkeit der Ergebnisse dieser Arbeit sind für wissenschaftliche Arbeiten üblich. Sie wurden weniger deshalb diskutiert, um die Aussagekraft dieser Arbeit einzuschränken, sondern vielmehr, um aufzuzeigen, wo ergänzender Forschungsbedarf besteht und wie die Erkenntnisse dieser Arbeit sinnvoll weiterentwickelt und verfeinert werden können.

Ausblick: Einfluss der Erfahrung auf den Bewertungsprozess

In dieser Arbeit konnte gezeigt werden, wie sich die Ausbildung der

Nutzungsintention mit zunehmender Benutzererfahrung verändert. Dabei bezog sich die Benutzererfahrung konkret auf eine spezifische Software. Vor dem Hintergrund mentaler Modelle als Grundlage für die Interaktion zwischen einem Benutzer und einem Informationsverarbeitungssystem ist anzunehmen, dass auch der allgemeine Erfahrungshintergrund eines Benutzers den Bewertungsprozess beeinflusst.

Die Ausbildung mentaler Modelle zur Interaktion mit einer unbekanntem Software basiert auf den vorhandenen Schemata des Benutzers. Verfügt ein Benutzer über ein breites allgemeines Wissen über Software und möglicher Bedienkonzepte, so stehen ihm entsprechend viele Möglichkeiten zur Analogiebildung zur Verfügung. Es ist zu erwarten, dass die Bedienung von Software, die zumindest in ihren Grundzügen auf verbreiteten Konzepten aufbaut (z. B. hierarchische Menüstruktur), leichter von einem allgemein erfahrenen Computerbenutzer erlernt werden können. Andererseits ist es vorstellbar, dass erfahrene Benutzer einer tiefgreifenden Veränderung des Bedienkonzepts ablehnender gegenüberstehen, da sie nicht nur neue Konzepte erlernen müssen, sondern eventuell auch bereits Erlerntes verlernen müssen. Dieser Effekt zeigte sich in einer Untersuchung von Whiteside, Jones, Levy und Wixon (1985). In dieser Studie wurden textbasierte mit graphischen Bedienkonzepten verglichen. Die Einstellung der Probanden zu den unterschiedlichen Bedienkonzepten wurde teilweise erheblich durch ihre Vorerfahrung mit anderen Bedienkonzepten beeinflusst. Die Bewertung eines neuen Systems fiel positiver aus, wenn die Benutzer bereits über Erfahrung mit einem ähnlichen System verfügten.

Der Einfluss der Erfahrung eines Benutzers auf dessen Wahrnehmung einer Software beschränkt sich nicht auf Erfahrungen mit der Bedienung von Informationsverarbeitungssystemen. Auch Wissen und Erfahrungen außerhalb des Bereichs der digitalen Informationsverarbeitung kann hilfreich bei der Analogiebildung und damit bei der Entwicklung eines adäquaten mentalen Modells sein. In vielen Gestaltungsrichtlinien wird beispielsweise empfohlen, sich sprachlich an das Vokabular des Benutzers anzupassen. Entsprechend ergibt ein anderer Sprachstil für die Bedienoberfläche je nach fachlichem Hintergrund der anvisierten Benutzergruppe. Das Benutzerinterface einer Software für professionellen Textsatz kann also durchaus Fachausdrücke benutzen, die es in einer universellen Textverarbeitung zu vermeiden gilt. Gängige Gestaltungsrichtlinien betonen die notwendige Abstimmung einer optimalen Gestaltung auf das Wissen, die Erfahrungen und das Vokabular der anvisierten Benutzer. Wie die vorhandene Erfahrung zukünftiger Benutzer festgestellt werden kann und welche Erfahrungsdomänen relevant sind, ist allerdings weitgehend unerforscht. Hier hängt es bislang vom Einfühlungsvermögen der für die Systemgestaltung verantwortlichen Personen ab, inwieweit auf Anhieb eine passende Gestaltung implementiert wird.

Insbesondere der Erfahrungshintergrund eines Benutzers aus dem Anwendungsbereich einer Software wird sich auf die Wahrnehmung derselben

auswirken. Eine Bürofachkraft wird beispielsweise die „Ordner“-Metapher zur Dateiverwaltung gängiger Betriebssysteme als verständlicher bewertet verglichen mit einem im Produktionsbereich eingesetzten Facharbeiter, der nur selten mit in Ordnern sortierten Unterlagen in Berührung kommt. Diese beiden fiktiven Personen unterscheiden sich bezüglich der für sie verfügbaren Schemata, wie Inhalte sortiert werden können. Die Vertrautheit mit dem Konzept „Ordner“ bestimmt, in welchem Ausmaß die Analogie den Lernprozess erleichtert.

Aus diesen Überlegungen lässt sich ein allgemeiner Einfluss des Bildungsniveaus auf den Bewertungsprozess von Software erwarten. Die Ergebnisse einer Untersuchung von Chismar und Wiley (2003) können in dieser Weise interpretiert werden. Die Autoren fanden in einer Befragung von Ärzten zur Nutzungsintention zukünftiger internetbasierter Gesundheitssysteme keinen Einfluss des TAM-Konstrukts PEOU auf die Nutzungsintention. Die Autoren führen dies mitunter darauf zurück, dass für Ärzte überdurchschnittliche kognitive und intellektuelle Fähigkeiten vorausgesetzt werden können und Ärzte – durch ihre Ausbildung – gewohnt sind, neue Fähigkeiten zu erlernen. Eine Verallgemeinerung dieser Annahme führt zu der Hypothese, dass die Bedeutung der Erlernbarkeit einer Software generell mit zunehmender allgemeiner Lernerfahrung abnehmen sollte.

Eine zweite Einflussgröße, die die Erfahrung mit der Bedienung von Software systematisch variiert, ist das Alter einer Person. Junge Menschen wachsen auf umgeben von einer Vielzahl von Informations- und Unterhaltungstechnologien. Sie lernen die Interaktion mit diesen Systemen in einem anderen Lebensabschnitt als Erwachsene. Es ist vorstellbar, dass sich hierdurch grundlegend andere Einstellungen und Bewertungsstrategien gegenüber Informationsverarbeitungssystemen entwickeln. Im TAM wie auch im in dieser Arbeit postulierten Prozessmodell sind diese Einflüsse – zumindest teilweise – implizit in den Konstrukten enthalten, da diese als Wahrnehmungskorrelate der Personen konzipiert sind. Somit werden diese Veränderungen zum Teil bereits durch den Einfluss des unterschiedlichen Erfahrungshintergrunds verschieden alter Personen erklärt. Allerdings wäre es erstrebenswert, den Einfluss altersabhängiger Persönlichkeitsmerkmale auf die Ausbildung der Konstrukte des Prozessmodells besser zu verstehen. Hierbei ist anzunehmen, dass sich die altersbedingten Unterschiede nicht auf die Art und Anzahl der für die Modellbildung zur Verfügung stehenden Schemata beschränken.

Praktische Bedeutung der Ergebnisse

Bei der Planung einer Bedienoberfläche sind die Ergebnisse dieser Arbeit auf konzeptioneller Ebene von Bedeutung. In Abhängigkeit von der für das geplante Anwendungsfeld vertretbaren Einarbeitungszeit kann der Fokus stärker auf unerfahrene Benutzer gelegt und die Erlernbarkeit optimiert werden oder eine Ausrichtung eher auf erfahrene Benutzer mit einer optimierten Bedienbarkeit für diese erfolgen.

Im Hinblick auf konkrete Detailgestaltungen vermögen die Ergebnisse dieser Arbeit weniger zu helfen, da die Konstrukte des Prozessmodells nicht als objektive Softwareeigenschaften definiert werden können. Die Konzeption der Variablen des Prozessmodells – wie auch schon der TAM-Konstrukte – als Wahrnehmungskorrelate beschränkt die Möglichkeiten, die Ergebnisse dieser Arbeit zur Optimierung der Gestaltung von Software zu nutzen. Das Prozessmodell verdeutlicht zwar, dass für unerfahrene Benutzer eine hohe wahrgenommene Erlernbarkeit (PL) zu einer ausgeprägteren Nutzungsintention führt, lässt aber die Frage, durch welche Gestaltung sich die wahrgenommene Benutzbarkeit verbessern ließe, unbeantwortet. Diese Problematik trifft analog auch für das Konstrukt PEUEU zu. Um zu konkreten Empfehlungen hinsichtlich der Systemgestaltung zu gelangen, müssten die Determinanten der Wahrnehmungskonstrukte gemeinsam mit den zugehörigen Wahrnehmungsprozessen erforscht werden.

In diesem Zusammenhang erscheint es als erster Schritt viel versprechend, die Wirkung der Gestaltungsrichtlinien einschlägiger Usability-Normen, wie der EN-ISO 9241 auf die Wahrnehmung der Erlernbarkeit bzw. Benutzbarkeit durch erfahrene Benutzer zu untersuchen. Dringend sind hierzu Messverfahren notwendig, die eine objektive Erfassung von Eigenschaften einer Bedienoberfläche ermöglichen. Verfügbare Messinstrumente wie der ErgoNorm-Fragebogen (Dzida, Hofmann, Freitag, Redtenbacher, Baggen, Geis, Beimel, Zurheiden, Hampe-Neteler, Hartwig und Peters, 2000), der geeignet ist, subjektiv empfundene Abweichungen von den Empfehlungen der Norm festzustellen, leisten dies nicht. Für einen Entwickler in der Praxis ist es aber erstrebenswert, schon beim ersten Entwurf eines Bedienkonzepts über möglichst konkrete Richtlinien zu verfügen, anstatt in einem iterativen Trial-and-Error-Vorgehen solange erkannte Schwachstellen zu verändern, bis ein zufriedenstellendes Niveau erreicht ist.

Eine Herausforderung für zukünftige Forschung besteht in der Verbesserung des Verständnisses der an der Bewertung moderner Informations- und Kommunikationstechnologie beteiligten Prozesse. Eine dringliche Frage im Hinblick auf die Ergebnisse dieser Arbeit scheint der Zusammenhang zwischen objektiven Systemeigenschaften und den zugehörigen Wahrnehmungskorrelaten zu sein. Erkenntnisse über systematische Zusammenhänge in diesem Bereich würden die Weiterentwicklung von Gestaltungsrichtlinien auf einer soliden theoretischen Grundlage ermöglichen.

Literaturverzeichnis

- Adams, D. A., Nelson, R. R. & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: a replication. *MIS Quarterly*, 16 (2), 227-247.
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. New York: Prentice-Hall.
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Anderson, J. R. (1985). *Cognitive psychology and its implications* (2. Ed.). San Francisco: Freeman.
- Asher, J. J. (1963). Towards a neo-field theory of problem solving. *Journal of General Psychology*, 68, 3-8.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). *Multivariate Analysemethoden*. Berlin: Springer.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.
- Bartlett, F. C. (1932). *Remembering: a study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Bevan, N. & Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology*, 13, 132-145.
- Boehm, Brown & Lipow (1976). Quantitative Evaluation of Software Quality. *Proceedings of the 2nd International Conference on Software Engineering*, 592-605.
- Borg, I. & Groenen, P. J. F. (2005). *Modern multidimensional scaling* (2. Aufl.). New York: Springer.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Berlin: Springer.
- Bortz, J., Lienert, G. & Boehnke, K. (2000). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Bowen, W. (1986). The Puny Payoff from Office Computers. *Fortune*, May 26. Nachdruck in Forester, T. (Ed.) (1989), *Computers in the Human Context*, 267-271. Blackwell.

- Bundesministerium für Wirtschaft (1996). *Info 2000: Deutschlands Weg in die Informationsgesellschaft*, Bonn.
- Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodal inference: a practical information-theoretic approach* (2. Aufl.). New York: Springer.
- Calisir, F. & Calisir, F. (2004). The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems. *Computers in Human Behavior*, 20, 505-515.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56 (2), 81-105.
- Carroll, J. M. & Mack, R. L. (1984). Learning to use a word processor: by doing, by thinking, and by knowing. In: J. C. Thomas & M. L. Schneider (Eds.): *Human Factors in Computer Systems*. Norwood: Ablex Publishing.
- Cattell, R. B. & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales. *Educational and Psychological Measurement*, 1, 3-30.
- Chismar, W. G. & Wiley-Patton, S. (2003). Does the extended technology acceptance model apply to physicians. *Proceedings of the 36th Hawaii international conference on system sciences* (HICSS'03).
- Chung, J. & Tan, F. B. (2004). Antecedents of perceived playfulness an exploratory study on user acceptance of general information-searching websites. *Information & Management*, 41 (7), 869-881.
- Clark, L. A. & Watson, D. (1995). Constructing Validity: basic issues in objective scale development. *Psychological Assessment*, 7 (3), 309-319.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78 (1), 98-104.
- Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.
- Dalgaard, P. (2002). *Introductory statistics with R*. New York: Springer.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13 (3), 319-340.
- Davis, F. D., Bagozzi, R. P., Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, 35 (8), 982-1003.

- Davis, F. D., Bagozzi, R. P., Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of applied social psychology*, 22 (14), 1111-1132.
- Davison, A. C. (2003). *Statistical Models*. New York: Springer.
- De Kleer, J. & Brown, J. S. (1983). Assumptions and Ambiguities in mechanistic mental models. In: D. Gentner & A. L. Stevens (Eds.): *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates, 7-14.
- Douglas, S. A. & Moran, T. P. (1984). Learning text editor semantics by analogy. In A. Janda (Ed.): *Human factors in computing systems. Proceedings of the CHI '83 Conference*, Boston (207-211). Amsterdam: North-Holland.
- Draper, S. W. (1984). The nature of expertise in UNIX. *Proceedings of IFIP INTERACT '84 First International Conference Human-Computer Interaktion*, 465-471.
- Dutke, S. (1994). *Mentale Modelle: Konstrukte des Wissens und Verstehens*. Göttingen: Hogrefe.
- Dweck, C. S. & Legett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Dzida, W. (1983). Das IFIP-Modell für Benutzerschnittstellen. *Office Management*, Sonderheft 13.
- Dzida, W., Herda, S. & Itzfeld, W. D. (1978). *Factors of user-perceived quality of interactive systems*. Bonn: Gesellschaft für Mathematik und Datenverarbeitung.
- Dzida, W., Hofmann, B., Freitag, R., Redtenbacher, W., Baggen, R., Geis, T., Beimel, J., Zurheiden, C., Hampe-Neteler, W., Hartwig, R. & Peters, H. (2000). *Gebrauchstauglichkeit von Software – ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN ISO 8241 Teile 10 und 11* (Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Forschungsbericht 1693). Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- FAZ-Institut und Novomind AG (2004). *Kundenkompass Onlineshopping*. Frankfurt: FAZ.
- Fishbein, M. & Ajzen, I. (1980): Predicting and understanding consumer behavior: Attitude-behavior correspondence. In: I. Ajzen & M. Fishbein (Hrsg.): *Understanding Attitudes and Predicting Social Behavior*. New York: Prentice Hall.
- Gentner, D. & Gentner, D. R. (1983). Flowing water and teeming crowds: mental models of electricity. In D. Gentner & A. L. Stevens (Hrsg.): *Mental Models*. New York: Erlbaum.

- Goodhue, D. L. (1995). Understanding user evaluations of information systems. *Management Science*, 41 (12), 1827-1844.
- Green, B. F. (1978). In defense of measurement. *American Psychologist*. 33, 664-670.
- Grudin, J. (1992). Utility and usability: research issues and development contexts. *Interacting with Computers*, 4 (2), 209-217.
- Hacker, W. (1998). *Allgemeine Arbeitspsychologie*. Bern: Huber.
- Han, S. (2003) *Individual Adoption of IS in Organisations: A Literature Review of the Intention-Based Theories*. Technical Report 539, TUCS, Aug 2003. Verfügbar im Internet: <<http://www.tucs.fi/publications/insight.php?id=tHan03a>> [2005-05-25].
- Hasebrook, J. (1995). *Multimedia-Psychologie*, Heidelberg: Spektrum.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- Hsu,C.-L. & Lu, H.-P. (2004). Why do people play on-line games? An extended TAM with social influence and flow experience. *Information & Management*, 41 (7), 853-868.
- Johnson-Laird, P. N. (1983). *Mental models: towards a cognitive science of language, inferences, and consciousness*. Cambridge: Cambridge University Press.
- Karahanna, E. & Straub, D. W. (1999). The psychological origins of perceived usefulness and ease-of-use. *Information & Management*, 35, 237-250.
- Keil, M., Beranek, P. M. & Konsynski, B. R. (1995). Usefulness and ease of use: field study evidence regarding task considerations. *Decision Support Systems*, 13, 75-91.
- Klopping, I. M. & McKinney, E. (2004). Extending the technology acceptance model and the task-technology fit model to consumer e-commerce. *Information Technology, Learning, and Performance Journal*, 22 (1), 35-48.
- Krings, B.-J. (2004). Auswirkung der Informationstechnologien auf die Arbeitswelt. *NACHRICHTEN – Forschungszentrum Karlsruhe*, 36 (4), 243-247.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399-402.
- Mahlke, S. (2005). *An integrative approach to user experience of interaction*. ZMMS Berlin, Internal report. Verfügbar im Internet: <[http://www.zmms.tu-berlin.de/~sma/User %20experience %20research_Mahlke.pdf](http://www.zmms.tu-berlin.de/~sma/User%20experience%20research_Mahlke.pdf)> [2005-08-16].

- Manktelow, K. & Jones, J. (1987). Principles from the psychology of thinking and mental models. In: M. M. Gardiner & B. Christie (Eds.): *Applying Cognitive Psychology to User-Interface Design*. New York: John Wiley & Sons, 83-117.
- Marks, D. F. (1990). On the relationship between imagery, body, and mind. In P. J. Hampson, D. F. Marks & J. T. E. Richardson (Hrsg.): *Imagery: current developments*. London: Routledge.
- Mayer, R. E. (1991). *Thinking, problem solving, cognition* (2. Auflage). New York: Freeman.
- Moon, J.-W. & Kim, Y.-G. (2001). Extending the TAM for a world-wide-web context. *Information & Management*, 38, 217-230.
- Moore G. C. & Benbasat I. (1991). Combining utility usability and accessibility methods for universal access. *Information Systems Research*, 2 (3), 192-222.
- Neisser, U. (1976). *Cognition and reality: principles and implications of cognitive psychology*. San Francisco: Freeman.
- Neisser, U. (1979). *Kognition und Wirklichkeit: Prinzipien und Implikationen der kognitiven Psychologie*. Stuttgart: Klett-Cotta.
- Nielsen, J. (1993). *Usability Engineering*, San Diego: Academic Press.
- Nielsen, J. (2003). *Usability 101: Introduction to Usability*. Verfügbar im Internet: <<http://www.useit.com/alertbox/20030825.html>> [2007-04-16].
- Norman, D. A. (1983). Some observations on mental models. In: D. Gentner & A. L. Stevens (Eds.): *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates, 7-14.
- Penner, D. E. (2001). Cognition, Computers, and Synthetic Science: Building Knowledge and Meaning through Modeling. *Review of Research in Education*, 25, 1-35.
- Pinheiro, J. C. & Bates, D. M. (2004). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Rips, L. J. (1990). Reasoning. *Annual review of psychology*, 41, 321-353.
- Roberts, T. L. & Moran, T. P. (1983). The evaluation of text editors: methodology and empirical results. *Communications of the ACM*, 26 (4), 265-283.
- Robey, D. (1979). User Attitudes and Management Information System Use. *The Academy of Management Journal*, 22 (3), 527-538.
- Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222-245.

- Rouse, W .B. & Morris, N. M. (1986). On looking into the black box: prospects and limits in the search for mental models. *Psychological Bulletin*, 100 (3), 349-363.
- Rumelhart, D. E. (1980). Schemata: the building blocks of cognition. In: R. J. Spiro, B. C. Bruce & W. F. Brewer (Eds.): *Theoretical Issues in Reading Comprehension*. Hillsdale: Lawrence Erlbaum.
- Sakamoto, Y., Ishiguro, M. & Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. Dordrecht: KTK Scientific Publishers.
- Saadé, R. & Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: an extension of the technology acceptance model. *Information & Management*, 42 (2), 317-327.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6 (2), 461-464.
- Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning*, 1 (1), 59-85.
- Segars, A. H. & Grover, V. (1993). Re-examining perceived ease of use and perceived usefulness: a confirmatory factor analysis. *MIS Quarterly*, 17 (4). 517-525.
- Senghaas, M. (2004). *Personenbezogene Einflussfaktoren des Lernens aus Fehlern*. Unveröffentlichte Diplomarbeit, Universität Lüneburg.
- Sherif, M. & Sherif, C. W. (1967). The own category procedure in attitude research. In: M. Fishbein (Ed.): *Readings in Attitude Theory*. New York: Wiley.
- Shih, H.-P. (2004). Extended technology acceptance model of internet utilization behavior. *Information & Management*, 41 (6), 719-729.
- Shneiderman, B. (1980). *Software Psychology. Human Factors in Computer and Information Systems*. New York: Winthrop.
- Spinas, P., Waeber, D. & Strohm, O. (1990). Kriterien der benutzerorientierten Dialoggestaltung und partizipative Softwareentwicklung: eine Literaturlaufarbeitung. In: P. Spinas, M. Rauterberg, O. Strohm, D. Waeber & E. Ulich (Eds.), *Benutzerorientierte Softwareentwicklung und Schnittstellengestaltung*. Projektbericht Nr. 1. ETH Zürich: Institut für Arbeitspsychologie.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer.
- Steier, R. & Eid, M. (2001). *Messen und Testen*. Berlin: Springer.
- TNS-Infratest (2007). *(N)ONLINER Atlas 2007: Eine Topographie des digitalen Grabens durch Deutschland*. Verfügbar im Internet: <<http://www.nonliner-atlas.de>> [2007-07-16].

- Ulich, E. (1986). Aspekte der Benutzerfreundlichkeit. In: W. Remmler & M. Sommer (Hrsg.): *Arbeitsplätze morgen, Berichte des German Chapter of the ACM* (102-121). Stuttgart: Teubner.
- Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S*. New York: Springer.
- Venkatesh, V. & Davis, F. D. (1996). A Model of the Antecedents of Perceived Ease of Use: Development and Test. *Decision Sciences*, 27 (3), 451-481.
- Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management science*, 46 (2), 186-204.
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27 (3), 425-478.
- Veldhuyzen, W. & Stassen, H. G. (1976). The internal model: what does it mean in human control? In T. B. Sheridan & G. Johansson (Hrsg.): *Monitoring behavior and supervisory control*. New York: Plenum Press.
- Veldhuyzen, W. & Stassen, H. G. (1977). The internal model concept: an application to modeling human control of large ships. *Human Factors*, 19, 367-380.
- Vicente, K. J. (1997). Should an interface always match the operator's mental model? *CSERIAC Gateway*, 8 (1), 1-5.
- Wason, P. C. (1966). Reasoning. In: B. M. Foss (Ed.): *New horizons in psychology*. Harmondsworth: Pinguin.
- Wason, P.C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly journal of experimental psychology*, 23, 63-71.
- Whiteside, J., Jones, S., Levy, P. S. & Wixon, D. (1985). User performance with command, menu, and iconic interfaces. In: L. Borman & B. Curtis (Eds.): *Proceedings of the ACM CHI 85 Human Factors in Computing Systems Conference*. April 14-18, 1985, San Francisco, California.
- Wu, J.-H. & Wang, S.-C. (2005). What drives mobile commerce? an empirical evaluation of the revised technology acceptance model. *Information & Management*, 42 (5), 719-729.
- Yi, M. Y. & Hwang, Y. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies*, 59 (4), 431-449.
- Yousfi, S. (2005). Mythen und Paradoxien der klassischen Testtheorie (II). *Diagnostica*, 51 (2), 55-66.

Zhang, L., Helander, M. G. & Drury, C. G (1996). Identifying Factors of Comfort and Discomfort in Sitting. *Human Factors*, 38 (3), 377-389.

Anhang A

Stichwortsammlung zu den Konstrukten

Konstrukt: Perceived Learnability (PL)

Konstrukt-Definition: Der Grad zu dem eine Person glaubt, dass die Benutzung einer bestimmten Software leicht zu erlernen ist.

Für eine Gliederung wurden nachträglich Aspekte zu den Stichpunkten zugeordnet. Diese zugeordneten Aspekte stehen unter den genannten Stichpunkten.

- „selbsterklärend“
Aspekt: Selbsterklärend
- „logisches Vorgehen“
Aspekt: Verständlichkeit
- „strukturiert“
Aspekt: Übersichtlichkeit
- „Übersichtlichkeit“
Aspekt: Übersichtlichkeit
- „Transparenz“
Aspekt: Verständlichkeit
- „keine unnötigen/störenden Informationen“
Aspekt: Übersichtlichkeit
- „Klarheit“
Aspekt: Übersichtlichkeit
- „Prägnanz“
Aspekt: Übersichtlichkeit
- „Einfachheit“
Aspekt: Übersichtlichkeit/Verständlichkeit
- „gute Visualisierung“
Aspekt: Übersichtlichkeit
- „ansprechend“
Aspekt: Übersichtlichkeit
- „Nutzen erkennbar“
Aspekt: Übersichtlichkeit
- „Belohnung bei Benutzung“
Aspekt: Übersichtlichkeit
- learning by doing
Aspekt: Fehlerkorrektur

- Hilfefenster
Aspekt: Selbsterklärend
- „Menüs übersichtlich“
Aspekt: Übersichtlichkeit
- „Aufbau wie Microsoft-Programme“
Aspekt: Erwartungskonformität
- „Icons logisch“
Aspekt: Verständlichkeit/Erwartungskonformität
- „gutes Hilfeprogramm“
Aspekt: Selbsterklärend
- „nicht zu viele Funktionen (die ein Anfänger nicht braucht)“
Aspekt: Verständlichkeit/Übersichtlichkeit
- „Grundfunktionen über ein Icon mit einem Klick erreichbar“
Aspekt: Übersichtlichkeit
- „einfache Führung durch Funktionen, nicht so viele Wahlmöglichkeiten“
Aspekt: Übersichtlichkeit
- „einfache Sprache“
Aspekt: Verständlichkeit
- „vertraute Bedienelemente“
Aspekt: Erwartungskonformität
- „vertraut mit Anwendungen“
Aspekt: Erwartungskonformität
- „Anwendungsbeispiele“
Aspekt: Selbsterklärend
- „Konformität mit Standards“
Aspekt: Erwartungskonformität
- „Gruppierung und Trennung von Bedienelementen“
Aspekt: Übersichtlichkeit
- „Bubble-Help (MouseOver)“
Aspekt: Übersichtlichkeit/Selbsterklärend
- „übersichtliche Liste aller Befehle vorhanden“
Aspekt: Übersichtlichkeit
- „die wichtigsten Befehle können eingeblendet werden“
Aspekt: Übersichtlichkeit/Selbsterklärend
- „wenige, einfache Bedienoptionen“
Aspekt: Übersichtlichkeit
- „ermöglicht intuitives Arbeiten“

Aspekt: Übersichtlichkeit/Verständlichkeit/Erwartungskonformität

• „selbsterklärende Schaltflächenbeschriftung“

Aspekt: Verständlichkeit

• „informatives, gut strukturiertes Tutorial“

Aspekt: Selbsterklärend

• „schlichter Eingangsbildschirm, Konzentration auf wenige, wesentliche Punkte“

Aspekt: Übersichtlichkeit

• „Sinn und Zweck sollte schnell ersichtlich werden“

Aspekt: Verständlichkeit/Übersichtlichkeit/Selbsterklärend

• „beispielhafte Tour durch das Programm (Wizard)“

Aspekt: Selbsterklärend

• „wichtige Aufgaben iconographisch im Programm darstellen“

Aspekt: Verständlichkeit/Selbsterklärend

Konstrukt: Perceived Ease of Use for Experienced Users (PEUEU)

Konstrukt-Definition: Der Grad zu dem eine Person glaubt, dass die Nutzung einer bestimmten Software für den darin geübten Benutzer nicht aufwendig ist.

Für eine Gliederung wurden nachträglich Aspekte zu den Stichpunkten zugeordnet. Diese zugeordneten Aspekte stehen unter den genannten Stichpunkten.

• „keine Redundanz“

Aspekt: keine überflüssigen Arbeitsschritte notwendig

• „keine automatischen Hilfefenster/Instruktionen“

Aspekt: die Interaktion wird vom Benutzer kontrolliert

• „abkürzende Tastenkombinationen (Hotkeys)“

Aspekt: Effizienz

• „automatische Spracherkennung und -korrektur“

Aspekt: Effizienz

• „kurze Wege (z. B. Mausclicks)“

Aspekt: Effizienz/keine überflüssigen Arbeitsschritte notwendig

• „Eingabespeicher“

unklar/nicht klassifizierbar

• „effiziente Vorgehensweise“

Aspekt: Effizienz

• „nutzbringender Outcome“

Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)

• „schnell“

Aspekt: Effizienz

- „Produktivität“
passt nicht zur Definition
- „Grundfunktionen mit wenigen Klicks erreichbar“
Aspekt: Effizienz/keine überflüssigen Arbeitsschritte notwendig
- „erweiterte Funktionen leicht zu finden“
Aspekt: Übersichtlichkeit
- „häufig benötigte Items sind an prominenten Stellen“
Aspekt: Übersichtlichkeit/Effizienz
- „Anpassung des Programms an die Bedürfnisse des Benutzers“
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- „es passiert nichts, zu dem man nicht vorher zugestimmt hat“
Aspekt: die Interaktion wird vom Benutzer kontrolliert
- „keine unnötigen Hilfe-Angebote“
Aspekt: die Interaktion wird vom Benutzer kontrolliert
- „ökonomische Eingabe“
Aspekt: keine überflüssigen Arbeitsschritte notwendig
- „Ausgaben verwendbar“
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- „Ziel ist mit wenigen Mausklicks erreichbar“
Aspekt: keine überflüssigen Arbeitsschritte notwendig
- „Bedienelemente, die zur Durchführung einer Aufgabe dienen, gruppieren“
Aspekt: Übersichtlichkeit/Effizienz
- „die gesamte Aufgabe kann in einem Programm bearbeitet werden“
Aspekt: Effizienz/keine überflüssigen Arbeitsschritte notwendig
- „es nicht erforderlich ist, sich viele verschiedenen Tastenkombinationen zumerken“
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- „wiederkehrende Aufgaben sollten evtl. automatisierbar sein“*
Aspekt: zielt auf die Funktionalität einer Software
- „Vorlagen/Templates lassen sich erzeugen“
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- „Schnittstellen für Import und Export“
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- „gute History-Funktion“
Aspekt: Fehlerkorrektur
- „Auto-Vervollständigen“

Aspekt: Effizienz

Anhang B

Analyse des ErgoNorm-Fragebogens

Aufgabenangemessenheit

- Enthält das Programm alle für Ihre Aufgaben benötigten Funktionen?
Geht über die Bedienbarkeit von Software hinaus und legt den Fokus auf die Funktionalität. Für die neuen Skalen nicht relevant.
- Müssen Sie Eingaben oder Dialogschritte machen, die eigentlich überflüssig wären?
Aspekt: keine überflüssigen Arbeitsschritte notwendig
- Ist es Ihnen möglich, das wiederholte Eingeben von Daten oder Texten zu vereinfachen?
Aspekt: keine überflüssigen Arbeitsschritte notwendig
- Finden Sie, dass der erforderliche Aufwand für Ihr Arbeitsergebnis jeweils angemessen ist?
Aspekt: Effizienz
- Haben Sie das Gefühl, dass Sie Arbeiten machen müssen, die besser das Programm erledigen sollte?
Aspekt: keine überflüssigen Arbeitsschritte notwendig
- Müssen Sie Werte und Texte eingeben, die der Computer eigentlich wissen könnte?
Aspekt: keine überflüssigen Arbeitsschritte notwendig
- Müssen Sie sich mit Umwegen oder Tricks behelfen, um Ihre Arbeitsergebnisse so zu erzielen, wie Sie diese haben möchten?
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)
- Finden Sie in dem Programm Hilfetexte, die Ihnen auch tatsächlich weiter helfen?
Aspekt: Selbsterklärend
- Passt das Programm zu Ihren Formularen und bisherigen Formaten?
Aspekt: Ein- bzw. Ausgabe ist wie gewünscht (passt zur Aufgabe)

Selbstbeschreibungsfähigkeit

- Sind die Informationen, die zur Erledigung der Aufgabe notwendig sind, auf dem Bildschirm übersichtlich verfügbar?
Aspekt: Selbsterklärend/Übersichtlichkeit
- Können Sie bei der Arbeit mit dem Programm erkennen, welche Eingabe als nächstes von Ihnen erwartet wird?
Aspekt: Selbsterklärend
- Sind die Meldungen des Systems für Sie immer verständlich?

Aspekt: Verständlichkeit

- Werden Sie vor Aktionen, die nicht rückgängig gemacht werden können, von der Software gewarnt?

Aspekt: Fehlerkorrektur

- Hilft Ihnen die Hilfefunktion wirklich weiter, wenn einmal ein Dialogschritt oder Menüpunkt nicht ganz klar ist?

Aspekt: Selbsterklärend

- Müssen Sie oft Kollegen oder ein Handbuch konsultieren, um weiterarbeiten zu können?

Aspekt: Selbsterklärend (dabei werden zusätzliche Annahmen über verfügbare Informationsquellen gemacht)

Steuerbarkeit

- Können Sie Ihre Arbeitsschritte in der Reihenfolge erledigen, die Ihnen am sinnvollsten erscheint?

Aspekt: die Interaktion wird vom Benutzer kontrolliert

- Macht das Programm manchmal etwas, ohne dass Sie es zu dem Zeitpunkt wollen?

Aspekt: die Interaktion wird vom Benutzer kontrolliert

- Können Sie bei Bedarf eine Aufgabe unterbrechen und später wieder fortsetzen, ohne alles neu eingeben zu müssen?

Aspekt: die Interaktion wird vom Benutzer kontrolliert

- Können Sie einen Arbeitsschritt wieder zurücknehmen, wenn es für Ihre Aufgabenerledigung zweckmäßig ist?

Aspekt: die Interaktion wird vom Benutzer kontrolliert/Fehlerkorrektur

- Fühlen Sie sich in Ihrem Arbeitstempo durch das Programm manchmal gebremst, z. B. durch zu lange Wartezeiten?

Aspekt: Effizienz, die Frage zielt aber in hohem Maße auf Aspekte der Effizienz des Programmcodes im Zusammenspiel mit eingesetzten Hardware ab

Erwartungskonformität

- Finden Sie Menüpunkte oder Funktionen dort, wo Sie Ihrer Meinung nach auch sein sollten?

Aspekt: Erwartungskonformität

- Sind Sie sich bei Wartezeiten immer noch sicher, ob das Programm weiterarbeitet?

Aspekt: Selbsterklärend

- Sind Sie manchmal überrascht, wie das Programm auf Ihre Eingabe reagiert?

Aspekt: Erwartungskonformität

Fehlertoleranz

- Bekommen Sie bei fehlerhaften Eingaben Korrekturhinweise?
Aspekt: Fehlerkorrektur/Selbsterklärend
- Können Sie die Folgen einer fehlerhaften Eingabe mit geringem Aufwand beheben?
Aspekt: Fehlerkorrektur
- Arbeitet das Programm während der Ausführung Ihrer Aufgabe immer stabil und zuverlässig?
Aspekt: hat nichts mit der Bedienoberfläche zu tun

Individualisierbarkeit:

- Können Sie am Computer alles so einstellen, dass Ihnen das Lesen und Arbeiten leichter fällt?
Aspekt: Effizienz (geht aber über die Bedienoberfläche hinaus)

Lernförderlichkeit

- Ermöglicht Ihnen das Programm, auch einmal etwas gefahrlos auszuprobieren?
Aspekt: Fehlerkorrektur

Anhang C

Analyse von PEOU-Skalen

Davis (1989)

Definition: The degree to which a person believes that using a particular system would be free of effort.

- Learning to operate CHART-MASTER would be easy for me.
- I would find it easy to get CHART-MASTER to do what I want it to do.
- My interaction with CHART-MASTER would be clear and understandable.
- I would find CHART-MASTER to be flexible to interact with.
- It would be easy for me to become skillful at using CHART-MASTER.
- I would find CHART-MASTER easy to use.

Moore & Benbasat (1991)

- I believe that a PWS is cumbersome to use.
- My using a PWS requires a lot of mental effort.
- Using a PWS is often frustrating.
- My interaction with a PWS is clear and understandable.
- I believe that it is easy to get a PWS to do what I want it to do.
- Overall, I believe that a PWS is easy to use.
- Learning to operate a PWS is easy for me.

Davis, Bagozzi, Warshaw (1989, 1992)

- Learning to operate WriteOne would be easy for me.
- I would find it easy to get WriteOne to do what I want it to do.
- It would be easy for me to become skillful at using WriteOne.
- I would find WriteOne easy to use.

Keil, Beranek, Konsynski (1995)

- I believe that CONFIG is cumbersome to use.
- Learning to operate CONFIG is easy for me.
- Using CONFIG is often frustrating.
- Using CONFIG is clear and understandable.
- I believe that it is easy to get CONFIG to do what I want it to do.
- My using CONFIG requires a lot of mental effort.

- Overall, I believe that CONFIG is easy to use.

Karahanna & Straub (1999)

- comfortable using E-mail

Moon & Kim (2001)

Definition: The strength of one's belief that interacting with a WWW would be free of effort.

- It will be impossible to use WWW without expert help.
- Learning to operate WWW is easy for me.
- It is difficult to learn how to use WWW.
- I find it easy to get WWW to do what I want it to do.
- It takes too long time to learn to use WWW.
- It is easy to remember how to use WWW.
- Using WWW requires a lot of mental effort.
- My interaction with WWW is clear and understandable.
- It is easy for me to become skillful at using WWW.

Chismar & Wiley-Patton (2003)

Definition: the degree to which a person believes that using a particular innovation would be free of effort.

- My interaction with Internet-based health applications will be clear and understandable.
- Internet-based health applications will be easy to use.
- Interacting with Internet-based health applications will not require a lot of mental effort.
- It will be easy to get Internet-based health applications to do what I want them to do.

Venkatesh, Morris, Davis, Davis (2003)

- My interaction with the system would be clear and understandable.
- It would be easy for me to become skillful at using the system.
- I would find the system easy to use.
- It takes to long to learn how to use the system to make it worth the effort.

Calisir & Calisir (2004)

- I found it easy to get the system to do what I wanted it to do.

- It was easy for me to become skillful at using the system.
- I found the system easy to use.

Hsu & Lu (2004)

Definition: The extent to which the user believes that playing on-line games was effortless.

- It is easy for me to become skillful at playing on-line games.
- Learning to play an on-line game is easy for me.
- It is easy to play.

Klopping & McKinney (2004)

- It is difficult to learn how to use the Internet to do my shopping activities.
- It took a long time to learn to use the Internet to do my shopping activities.
- I often become confused when I use the Internet for my shopping activities.

Shih (2004)

- Learning to use the Internet is easy for me.
- I can use the Internet in a manner that allows me to obtain the information I want.
- My interaction with the Internet is clear and understandable.
- In general, I find the Internet easy-to-use.

Saade & Bahli (2005)

- Learning to use the Internet-based learning system is hard for me.
- I find that the process of using the Internet-based learning system was clear, understandable and straight forward.
- Navigating through the Internet-based learning system was easy for me.
- It would be easy to become skillful at using the Internet-based learning system.
- I find the Internet-based learning system easy to use.

Wang & Wu (2005)

Definition: The degree to which a person believes that engaging in online transactions via mobile commerce would be free of effort.

- I think learning to use mobile commerce is easy.
- I think finding what I want via mobile commerce is easy.
- I think becoming skillful at using mobile commerce is easy.
- I think using mobile commerce is easy.

Anhang D

Anweisungen für die Sortieraufgaben

Anweisung A

Aufgabe

Vor sich haben Sie einen Stapel Karten, auf denen Aussagen bezüglich einer fiktiven SOFTWARE aufgedruckt sind. SOFTWARE könnte hierbei z. B. ein Office-Paket wie MS Office, ein Statistikpaket wie SPSS, ein Webbrowser wie der Internet Explorer, aber auch eine Internetseite wie www.google.com sein.

Ihre Aufgabe besteht darin, die Karten auf mehrere Stapel zu sortieren. Legen Sie alle Karten, die Ihrer Meinung nach inhaltlich zusammen gehören auf einen Stapel. Sie können so viele Stapel wie nötig bilden. Die Anzahl Karten, die Sie zu einem Stapel zusammenfassen, ist beliebig. Sie können also auch einen Stapel mit nur einer Karte bilden. Es gibt keine richtigen oder falschen Lösungen!

Vielen Dank!

Anweisung B

Aufgabe

Vor sich haben Sie einen Stapel Karten, auf denen Aussagen bezüglich einer fiktiven SOFTWARE aufgedruckt sind. SOFTWARE könnte hierbei z. B. ein Office-Paket wie MS Office, ein Statistikpaket wie SPSS, ein Webbrowser wie der Internet Explorer, aber auch eine Internetseite wie www.google.com sein.

Ihre Aufgabe besteht darin, die Karten zu sortieren nach:

dem Grad zu dem Sie glauben, dass die Benutzung einer bestimmten Software leicht zu erlernen ist.

Legen Sie die Karte, die am besten passt ganz nach links und reihen Sie die anderen Karten nach rechts auf. Es gibt keine richtigen oder falschen Lösungen!

Vielen Dank!

Anweisung C

Aufgabe

Vor sich haben Sie einen Stapel Karten, auf denen Aussagen bezüglich einer fiktiven SOFTWARE aufgedruckt sind. SOFTWARE könnte hierbei z. B. ein Office-Paket wie MS Office, ein Statistikpaket wie SPSS, ein Webbrowser wie der Internet Explorer, aber auch eine Internetseite wie www.google.com sein.

Ihre Aufgabe besteht darin, die Karten zu sortieren nach:

dem Grad zu dem Sie glauben, dass die Nutzung einer bestimmten Software für den darin geübten Benutzer nicht aufwändig ist.

Legen Sie die Karte, die am besten passt ganz nach links und reihen Sie die anderen Karten nach rechts auf. Es gibt keine richtigen oder falschen Lösungen!

Vielen Dank!

Anhang E

Ähnlichkeitsmatrix der Rohitems

	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	E12	E13	E14	E15	E16	E17	L01	L02	L03	L04	L05	L06	L07	L08	L09	L10	L11	L12	L13	L14	L15	L16	L17		
E01	6	9	2	6	6	1	6	6	1	9	1	9	1	8	2	7	3	4	2	1	0	0	1	0	1	0	0	1	1	1	2	2	1	1	1	
E02	3	6	5	6	0	2	7	0	9	9	4	9	4	4	6	4	6	2	0	1	5	8	4	7	2	2	2	2	7	2	3	5	4	4	2	
E03	3	4	8	3	1	10	4	3	6	2	12	2	5	1	2	1	2	1	2	1	2	2	2	0	1	1	2	0	1	1	1	1	1	1	1	
E04	2	6	4	8	1	1	6	3	0	4	4	5	5	2	1	1	1	1	1	1	4	6	2	7	2	2	2	2	7	4	3	3	6	4	0	
E05	6	5	8	8	2	1	9	3	0	4	2	8	1	2	3	1	1	0	0	2	2	1	3	0	1	0	1	0	3	1	2	3	4	2	1	
E06	6	6	3	1	2	0	1	8	0	9	3	3	2	9	3	8	2	0	0	2	3	2	3	1	1	2	2	0	2	2	1	4	3	0	1	
E07	1	0	1	1	1	0	2	1	0	1	1	1	1	9	0	7	0	0	0	0	1	1	1	0	0	1	0	1	0	1	2	0	1	2	0	1
E08	6	2	10	6	9	1	2	2	4	0	9	1	3	0	2	0	2	0	0	0	1	1	1	1	0	1	0	1	1	0	0	3	1	1	1	
E09	6	7	4	3	3	8	1	2	2	9	7	5	5	13	6	4	0	2	2	4	5	3	5	3	3	2	4	1	1	1	1	4	3	1	1	
E10	1	0	3	0	0	2	2	2	0	3	2	1	1	2	3	10	12	2	3	1	4	2	2	2	2	6	9	1	2	4	2	3	4	10	0	10
E11	9	9	6	4	4	9	1	4	9	0	3	7	3	11	2	6	2	1	1	3	8	7	5	7	4	2	3	3	2	2	2	3	3	2	2	
E12	1	9	2	4	2	3	1	0	7	3	3	1	5	5	4	1	4	3	8	7	5	7	5	7	4	2	3	7	4	2	5	6	4	2	2	
E13	8	4	12	5	8	3	1	9	5	2	7	1	1	6	1	1	0	1	1	2	4	3	1	1	1	2	1	2	2	0	1	2	0	1	2	0
E14	2	4	2	5	1	2	9	1	5	1	3	5	1	2	6	1	0	1	1	2	5	1	5	2	0	1	5	2	0	1	2	5	2	0	2	0
E15	7	6	5	2	2	9	0	3	13	1	11	5	6	2	3	5	0	1	1	3	4	2	3	2	4	2	4	2	2	0	0	0	2	2	2	2
E16	3	4	1	1	3	3	7	0	6	2	2	5	1	6	3	4	2	2	1	3	2	3	3	1	0	1	0	1	4	3	3	4	3	4	2	2
E17	4	6	2	1	1	8	0	2	4	3	6	4	1	1	5	4	3	1	0	3	2	4	3	1	1	3	2	1	3	2	1	3	1	5	6	2
L01	2	2	1	1	2	0	0	0	10	2	1	0	0	0	2	3	13	2	3	3	5	4	2	7	12	4	6	9	6	2	5	11	0	11	5	11
L02	1	0	2	1	0	0	0	2	12	1	4	1	1	1	1	2	1	13	4	4	4	2	4	4	4	8	13	3	5	8	4	3	4	10	0	10
L03	0	1	1	1	0	0	0	2	2	1	3	1	1	1	1	0	2	4	3	3	3	3	4	15	4	4	4	4	3	4	2	1	2	1	2	1
L04	0	5	2	4	2	2	1	0	4	3	3	8	2	2	3	3	3	4	3	8	11	10	4	6	5	7	8	7	10	6	11	4	0	11	4	4
L05	1	8	2	6	2	3	1	1	5	1	5	7	4	5	4	2	2	3	2	3	8	10	10	3	5	3	13	8	5	7	6	6	2	6	2	2
L06	0	4	2	2	1	2	1	1	3	4	3	5	3	1	2	3	4	5	4	3	11	10	7	3	6	5	8	11	8	10	4	11	5	0	5	6
L07	1	7	0	7	3	3	0	1	5	2	4	7	1	5	3	3	3	4	4	4	10	10	7	5	6	4	13	8	7	7	9	9	4	0	9	4
L08	0	2	1	2	0	1	0	0	3	2	2	4	1	2	2	1	1	2	4	15	4	3	3	5	3	5	4	3	2	1	3	1	3	1	3	1
L09	0	2	1	2	1	1	1	1	3	6	3	2	2	0	4	0	1	7	8	4	6	5	6	6	3	7	5	4	6	3	3	6	8	0	8	8
L10	1	2	2	2	0	2	0	0	2	9	3	3	1	1	2	1	3	12	13	4	5	3	5	4	5	7	2	5	10	4	1	4	8	0	4	8
L11	1	7	0	7	3	2	1	1	4	1	3	7	2	5	2	4	2	4	3	4	7	13	8	13	4	5	2	9	6	7	7	7	3	0	3	3
L12	1	2	1	4	1	0	2	1	1	2	2	4	2	3	0	3	1	6	5	3	8	8	11	8	3	4	5	9	9	5	8	3	0	5	8	3
L13	2	3	1	3	2	2	0	0	1	4	2	2	0	1	0	3	3	9	8	4	7	5	8	7	3	6	10	6	9	9	3	8	6	0	6	6
L14	2	5	1	3	3	2	1	0	1	2	2	5	1	2	0	4	3	6	4	2	10	7	10	7	2	3	4	7	9	9	5	9	4	0	9	4
L15	1	4	1	6	4	1	2	3	4	3	3	6	2	5	2	3	1	2	3	1	6	6	4	9	1	3	1	7	5	3	5	4	3	0	4	3
L16	1	4	1	4	2	4	0	1	3	4	3	4	0	2	2	4	5	5	4	2	11	6	11	9	3	6	4	7	8	8	9	4	7	0	7	7
L17	1	2	1	0	1	3	1	1	1	10	2	2	0	0	2	2	6	11	10	1	4	2	5	4	1	8	8	3	3	6	4	3	7	0	4	7

Tab 14: Ähnlichkeitsmatrix der Rohitems

Anhang F

Liste der Rohitems mit Kodierung

Skala für Perceived Learnability

- L01: Die Bedienung von SOFTWARE ist für mich leicht zu lernen.
- L02: Es ist leicht für mich ein geschickter Nutzer von SOFTWARE zu werden.
- L03: SOFTWARE verhält sich so, wie ich es von anderen Programmen kenne.
- L04: Es ist leicht für mich zu erkennen, welche Funktionen SOFTWARE bereitstellt.
- L05: SOFTWARE wirkt für mich sehr übersichtlich.
- L06: Wie SOFTWARE funktioniert ist für mich leicht zu erkennen.
- L07: Die Funktion von allen Bedienelementen ist offensichtlich.
- L08: Die Bedienung von SOFTWARE funktioniert wie in mir bekannten Programmen.
- L09: Ich brauche keine besonderen Kenntnisse für die Nutzung von SOFTWARE.
- L10: Ich kann mich leicht in die Bedienung von SOFTWARE einarbeiten.
- L11: Die Oberfläche von SOFTWARE wirkt auf mich gut strukturiert.
- L12: Die Struktur von SOFTWARE ist für mich leicht zu erkennen.
- L13: Ich komme mit SOFTWARE auf Anhieb gut zurecht.
- L14: Die Funktionen von SOFTWARE sind für mich leicht zu erkennen.
- L15: Ich finde alle für mich zur Benutzung wichtigen Informationen in SOFTWARE.
- L16: Die Bedienung von SOFTWARE ist für mich leicht verständlich.
- L17: Ich kann die Bedienung von SOFTWARE gut durch Ausprobieren erlernen.

Skala für Perceived Ease of Use for Experienced Users

- E01: Mit SOFTWARE kann ich effizient arbeiten.
- E02: Wichtige Funktionen sind für mich in SOFTWARE schnell erreichbar.
- E03: SOFTWARE passt zu meinem Arbeitsstil.
- E04: Die Bedienoberfläche von SOFTWARE ist für mich optimal gestaltet.
- E05: Die Ausgaben von SOFTWARE sind genau so wie ich es mir wünsche.
- E06: Die Eingabe von Informationen in SOFTWARE ist nicht aufwendig.
- E07: Ich fühle mich durch SOFTWARE nicht bevormundet.
- E08: Die Bedienung von SOFTWARE entspricht genau meinen Anforderungen.
- E09: Ich muss keine unnötigen Arbeitsschritte machen, wenn ich SOFTWARE bediene.
- E10: Mit Übung kann ich SOFTWARE fehlerfrei nutzen.
- E11: Mit SOFTWARE kann ich meine Ziele über wenige Arbeitsschritte erreichen.
- E12: Auch selten genutzte Funktionen sind für mich in SOFTWARE gut erreichbar.
- E13: Die Struktur von SOFTWARE passt für mich gut zu meinen Aufgaben.
- E14: Ich werde durch die Oberfläche von SOFTWARE nicht eingeschränkt.
- E15: Ich muss in SOFTWARE keine für mich unnötigen Arbeitsschritte machen.
- E16: Ich bekomme von SOFTWARE keine für mich überflüssigen Hilfestellungen.
- E17: Ich kann meine eigenen Fehler in SOFTWARE problemlos korrigieren.

Anhang G

Planungsaufgabe der zweiten Itemselektion

Aufgabe 1

Sie wollen für ein Wochenende in die Nähe von Dresden fahren. Planen Sie bitte die Strecke, die Sie Freitag nach der Arbeit nach Rathen fahren werden. Überprüfen Sie bitte die Zieladresse.

Startort:

Severinsmühlengasse 3

50678 Köln

Zielort:

Rathener Straße 15

01824 Rathen

Frage: Aus welcher Strasse biegen Sie laut Routenplaner am Ende ihrer Fahrt in die Rathener Strasse ein?

Aufgabe 2

Sie wollen direkt im Anschluss ins Kino. Dafür wollen Sie noch Bekannte abholen, die in der Stadt wohnen. Bitte planen Sie eine entsprechende Route (von hier aus).

Startort:

aktueller Aufenthaltsort

Zwischenstopp:

Auf der Rübekuhle 4

21335 Lüneburg

Zielort (CineStar):

Fährsteg 1

21339 Lüneburg

Frage: Wie häufig führt der Routenplaner Sie über die Lindenstrasse?

Aufgabe 3

Sie planen die Hinreise zu Ihrem Urlaubsort. Sie fahren in Hamburg los und wollen abends in Prag ankommen.

Startort:

Überseering 25
22297 Hamburg

Zielort:

Express Hotel (by Holiday Inn)
Opletalova 960/33
11000 Prag
Tschechien

Frage: An welchem Grenzübergang fahren Sie aus der BRD in die Tschechische Republik?

Aufgabe 4

Planen Sie bitte für Ihren letzten Urlaubstag die Rückreise aus Prag. Dabei wollen Sie einen Zwischenstop in Leipzig einlegen, um die Stadt zu besichtigen. Schauen Sie sich bitte auf der Karte das Zentrum von Leipzig an.

Zwischenstop:

Leipzig (Zentrum)

Frage 1: In welche Strasse in Leipzig führt Sie der Routenplaner?

Frage 2: Wie weit ist es von dort bis zu der Adresse in Hamburg?

Aufgabe 5

Um eine Ausstellung zu besuchen wollen Sie nach Stockholm fahren. Ihr Hotel ist in der Lill-Jans Plan.

Startort:

Am Maschhop 15
29649 Wietzendorf

Zielort:

Lill-Jans Plan
11425 Stockholm
Schweden

Frage: Führt die Route über Flensburg?

Aufgabe 6

Planen Sie bitte die Rückreise aus Stockholm.

Frage: Um wieviele Kilometer unterscheiden sich Hin- und Rückreise?

Anhang H

Häufigkeitsverteilungen der Items (zweite Itemselektion)

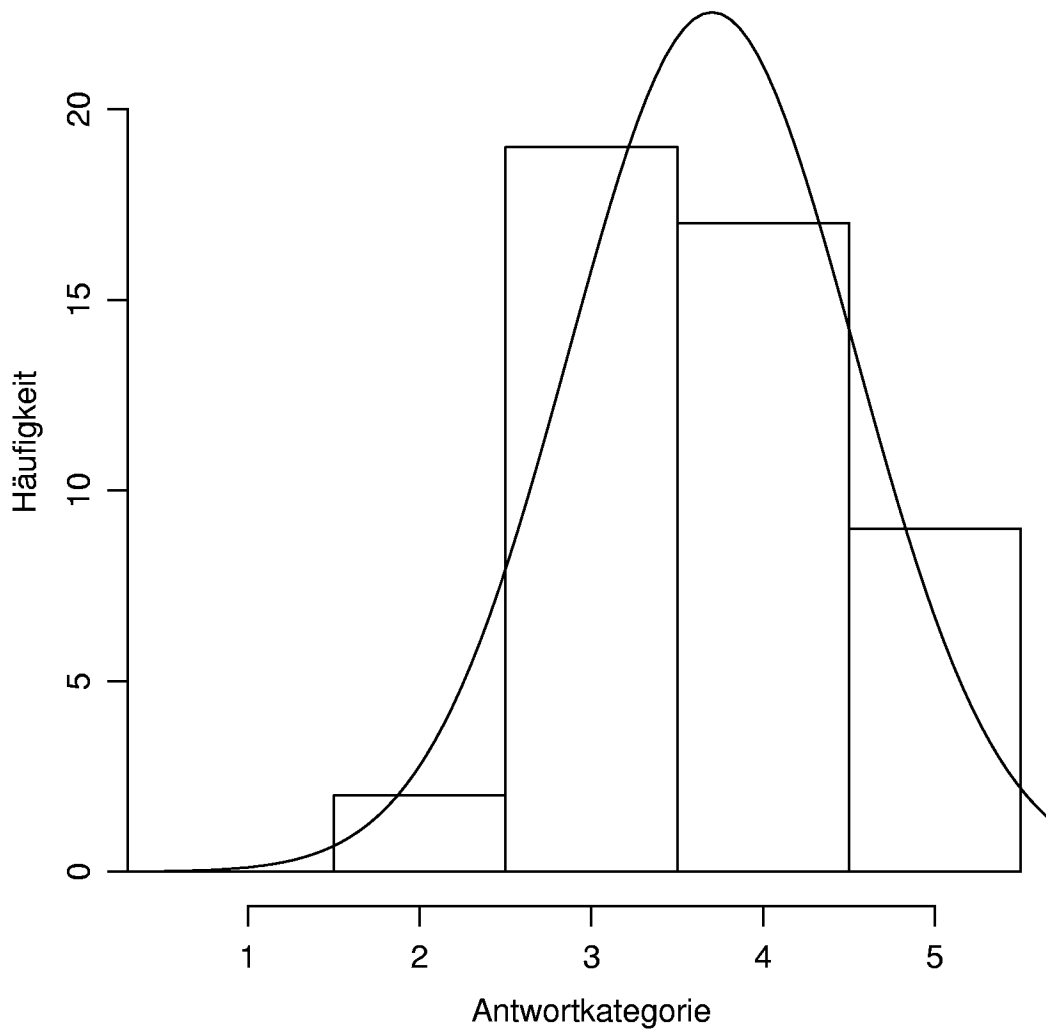


Abb. 54: Verteilung der Antworten für Item L01

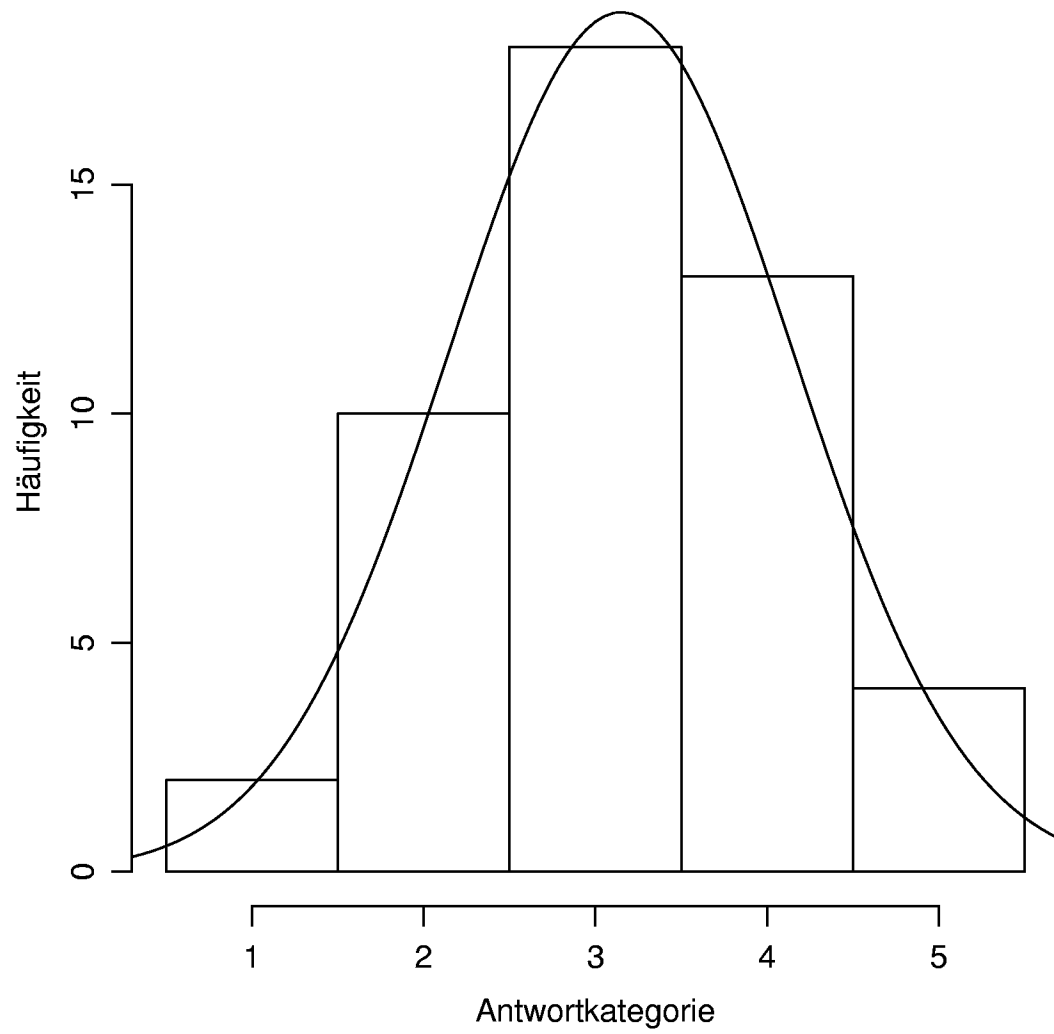


Abb. 55: Verteilung der Antworten für Item L02

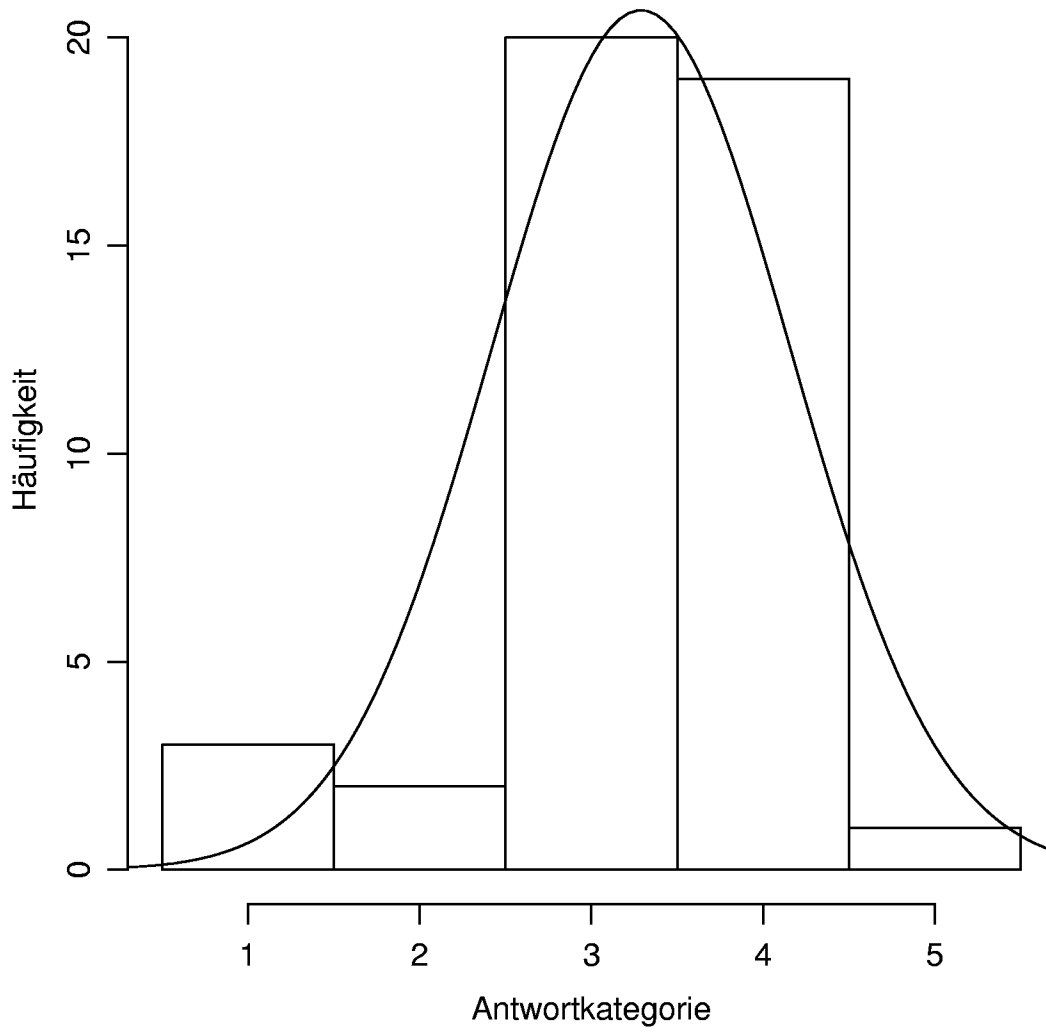


Abb. 56: Verteilung der Antworten für Item L03

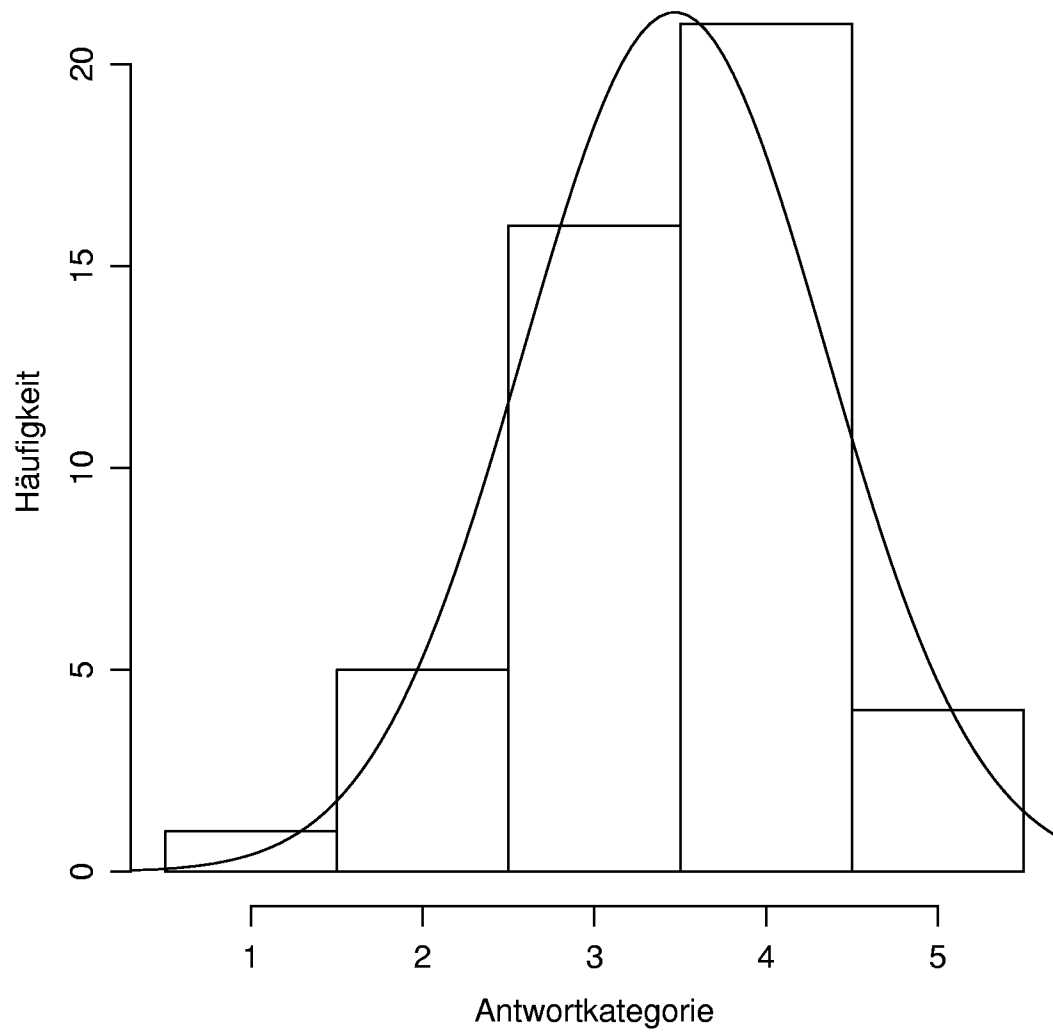


Abb. 57: Verteilung der Antworten für Item L06

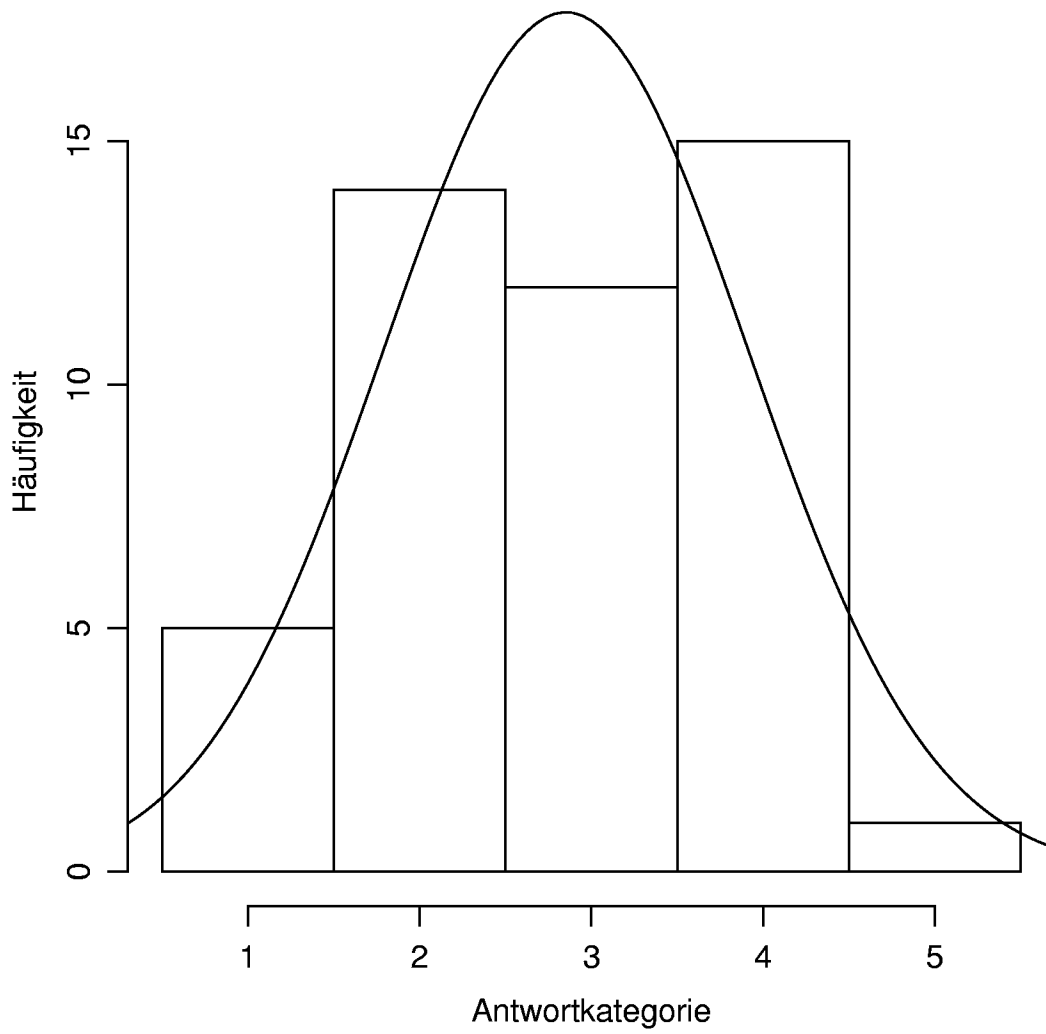


Abb. 58: Verteilung der Antworten für Item L07

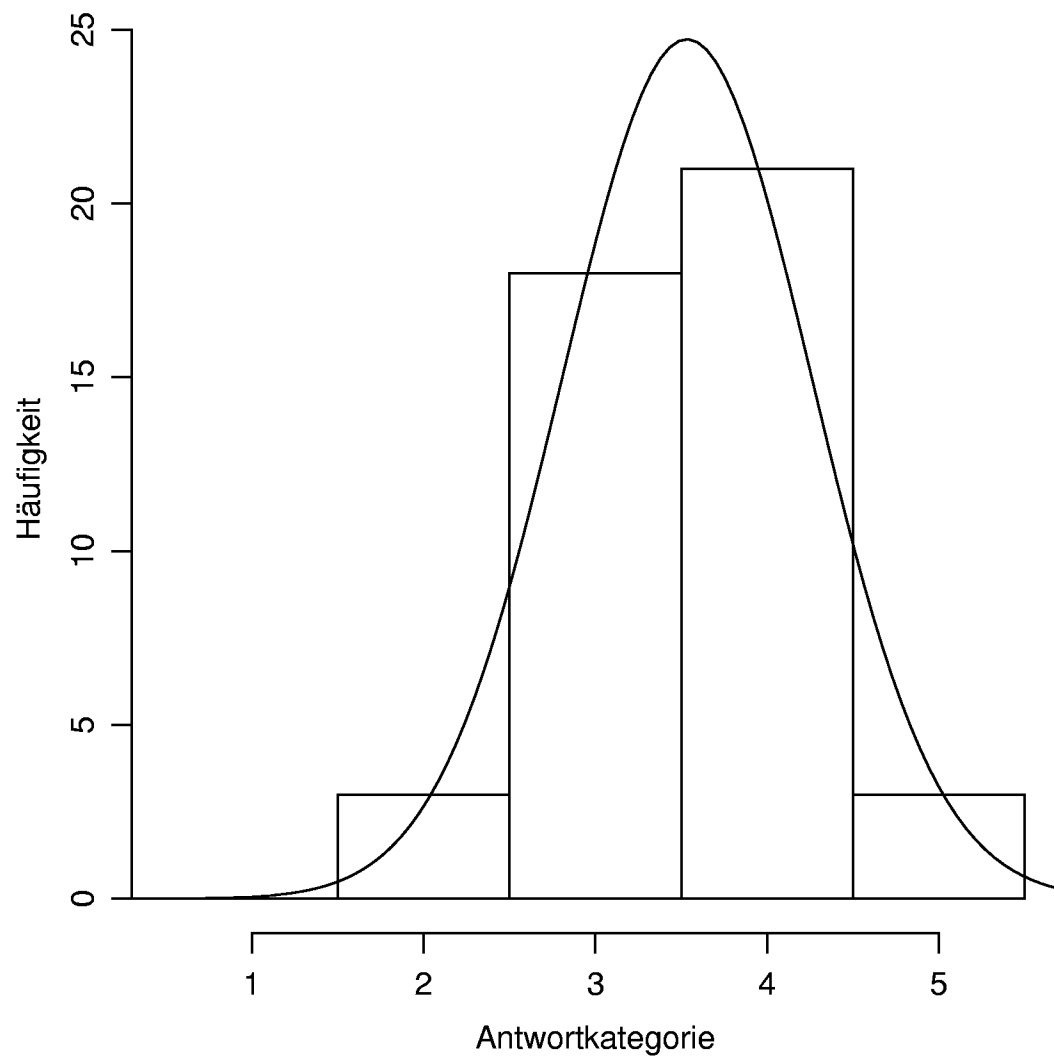


Abb. 59: Verteilung der Antworten für Item L08

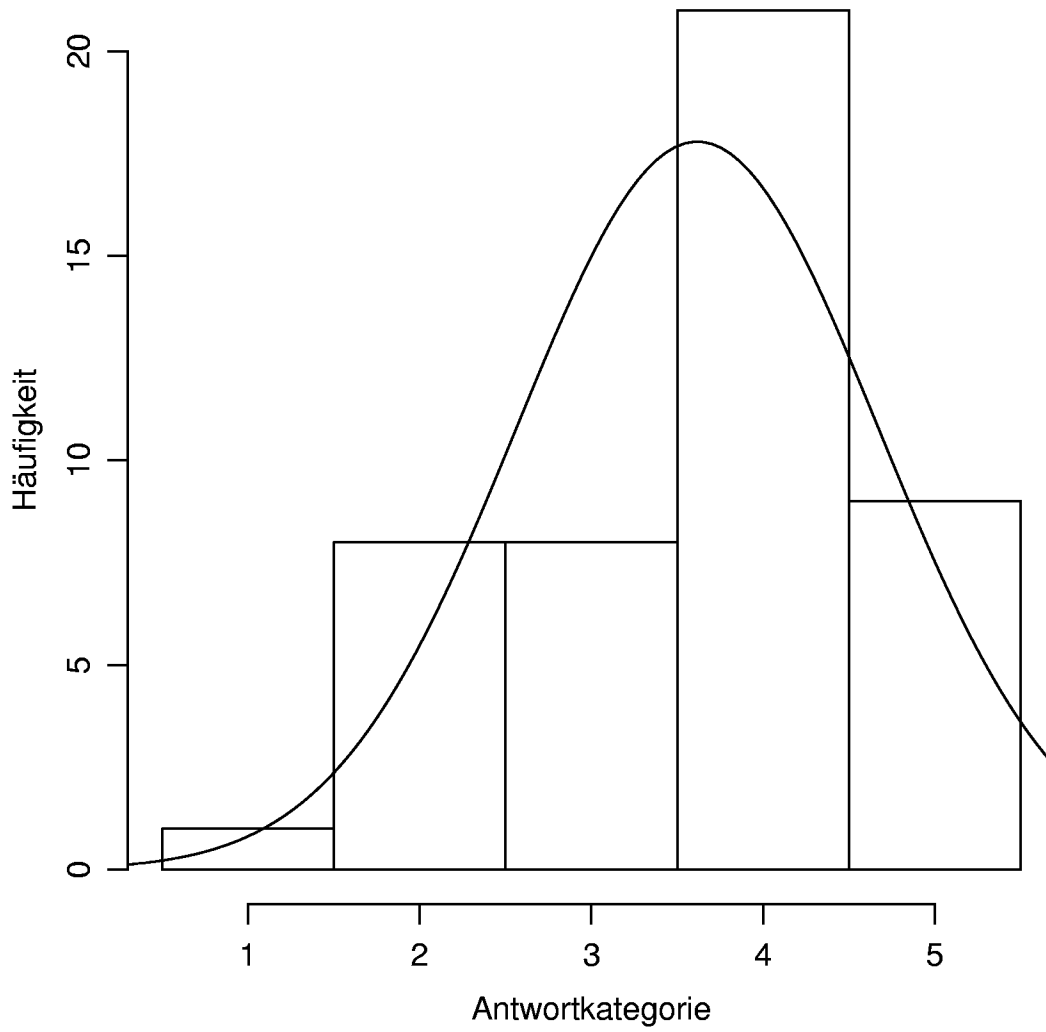


Abb. 60: Verteilung der Antworten für Item L09

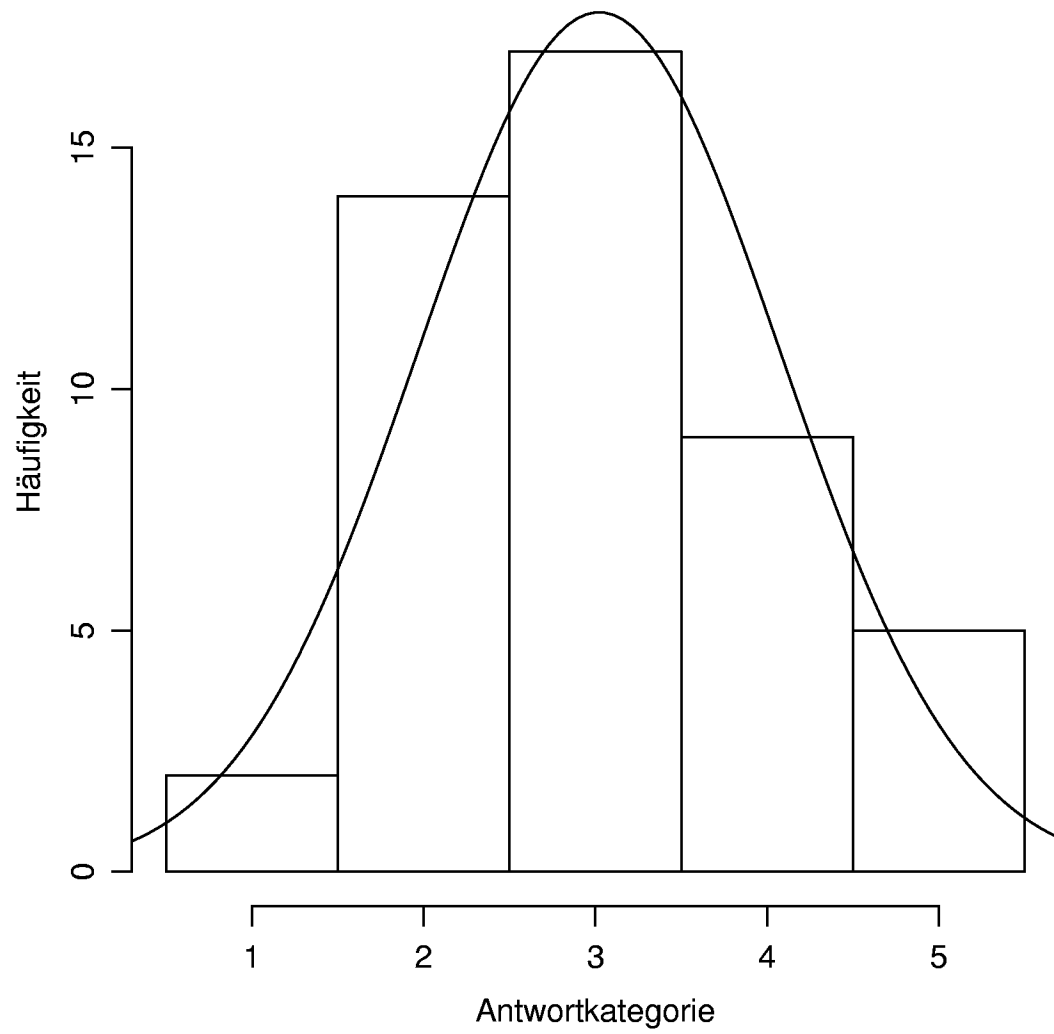


Abb. 61: Verteilung der Antworten für Item L11

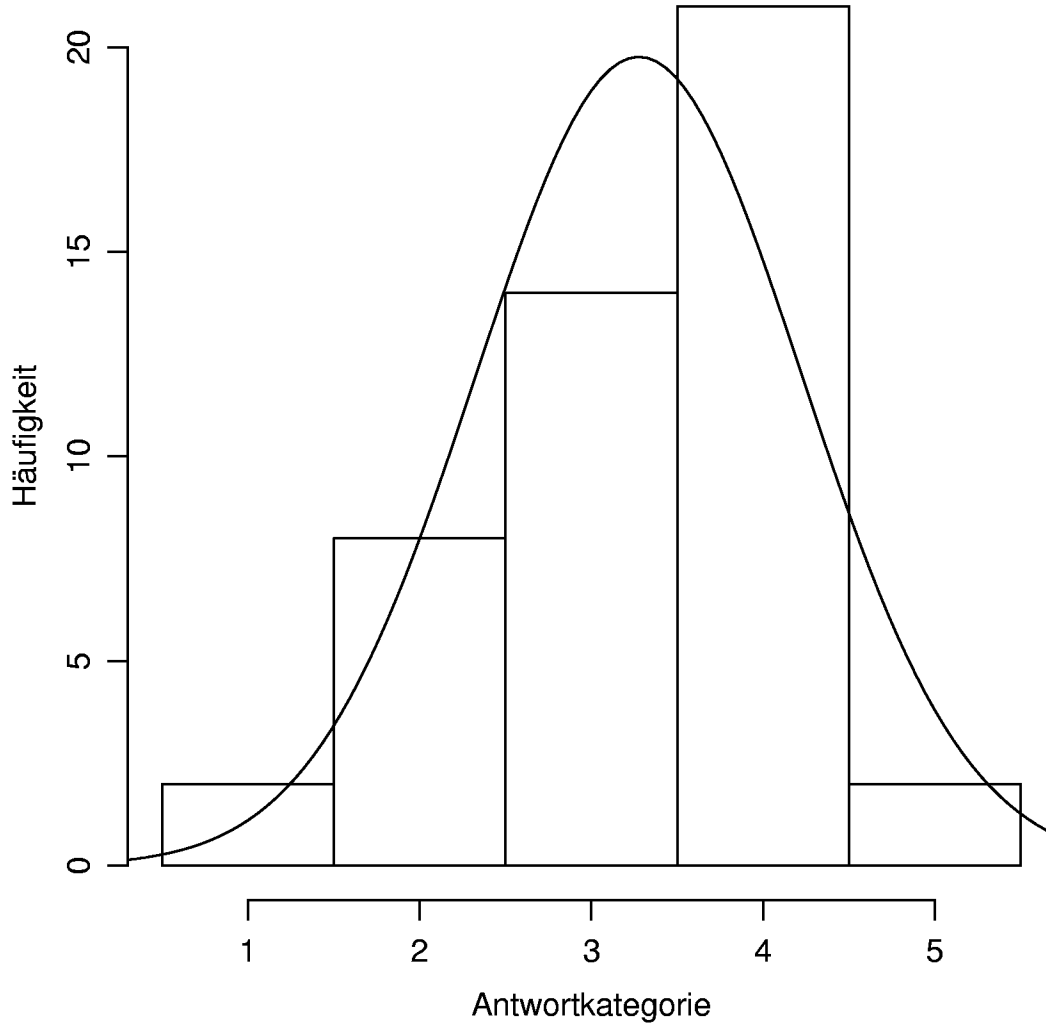


Abb. 62: Verteilung der Antworten für Item L12

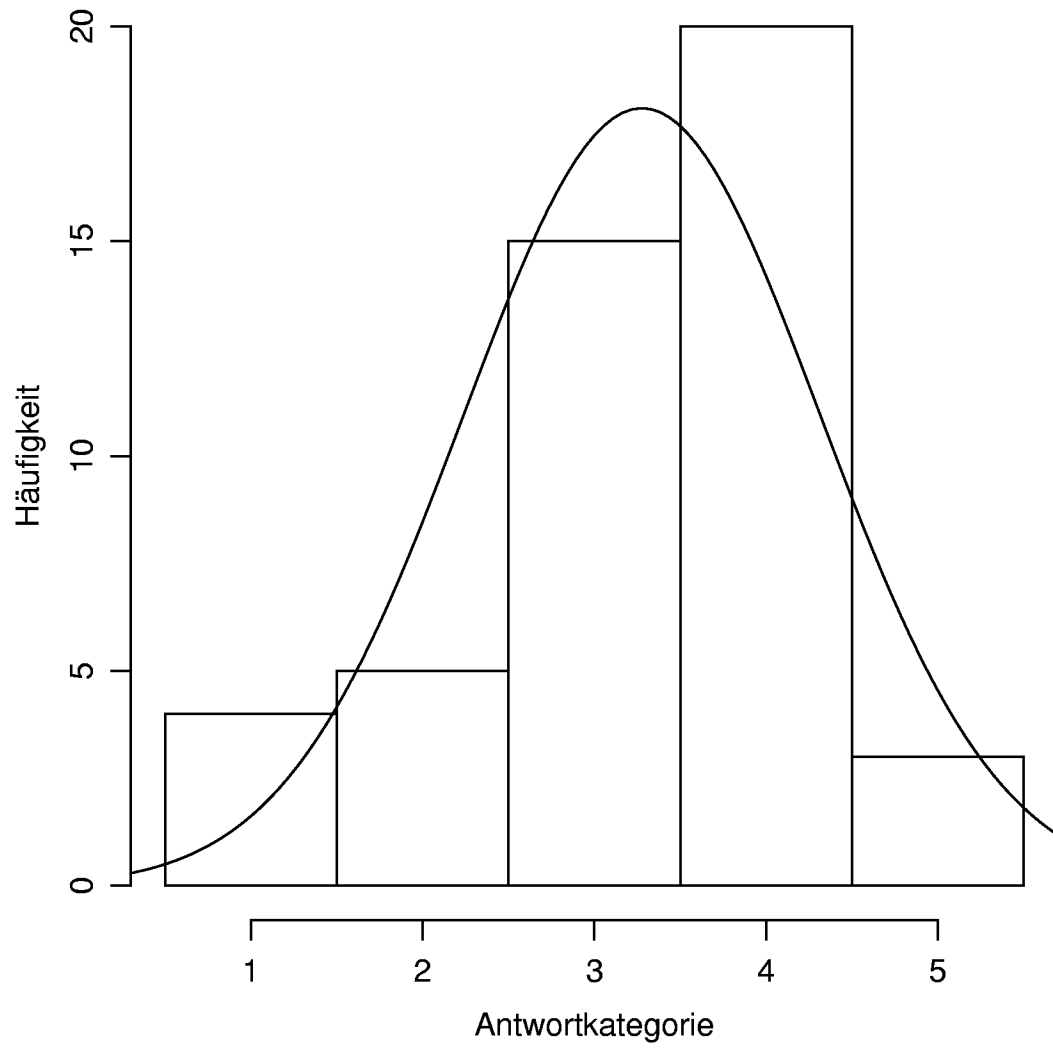


Abb. 63: Verteilung der Antworten für Item L13

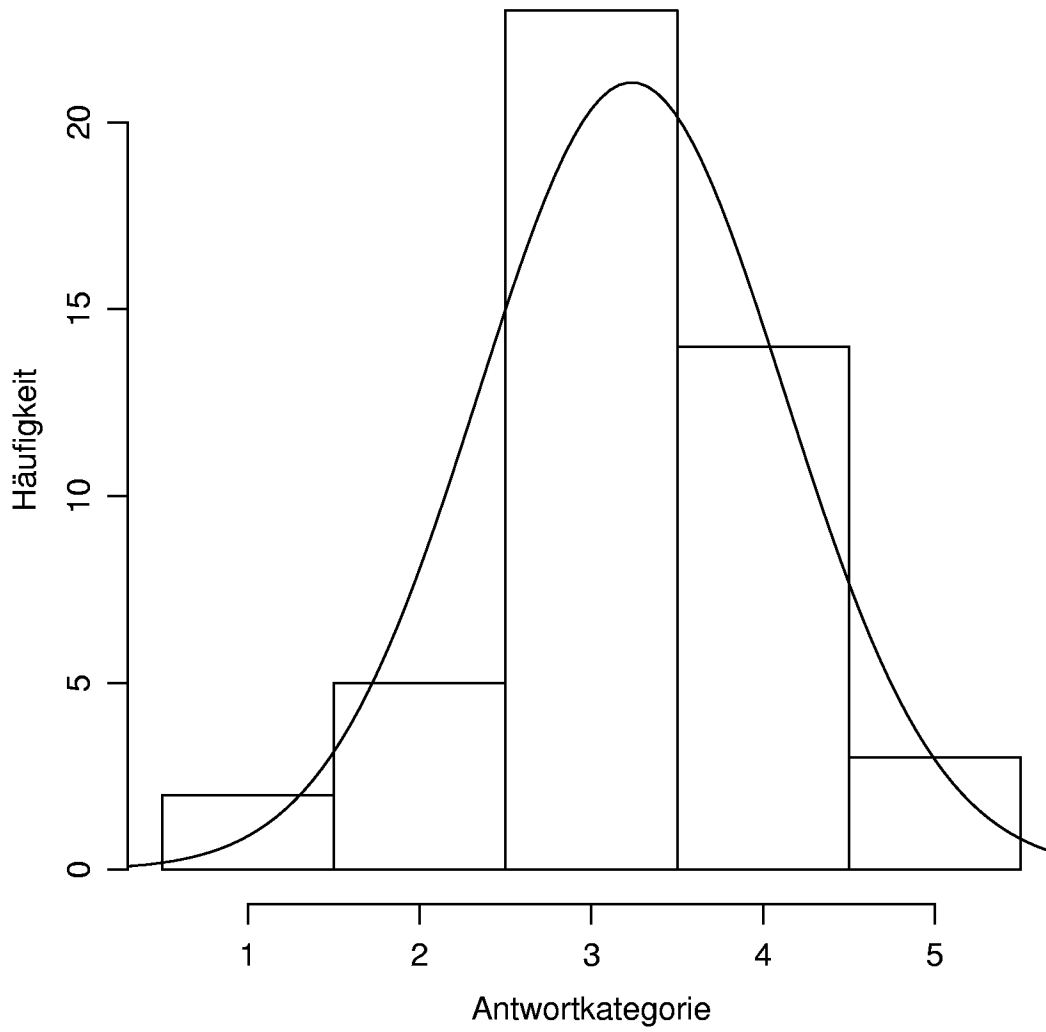


Abb. 64: Verteilung der Antworten für Item E01

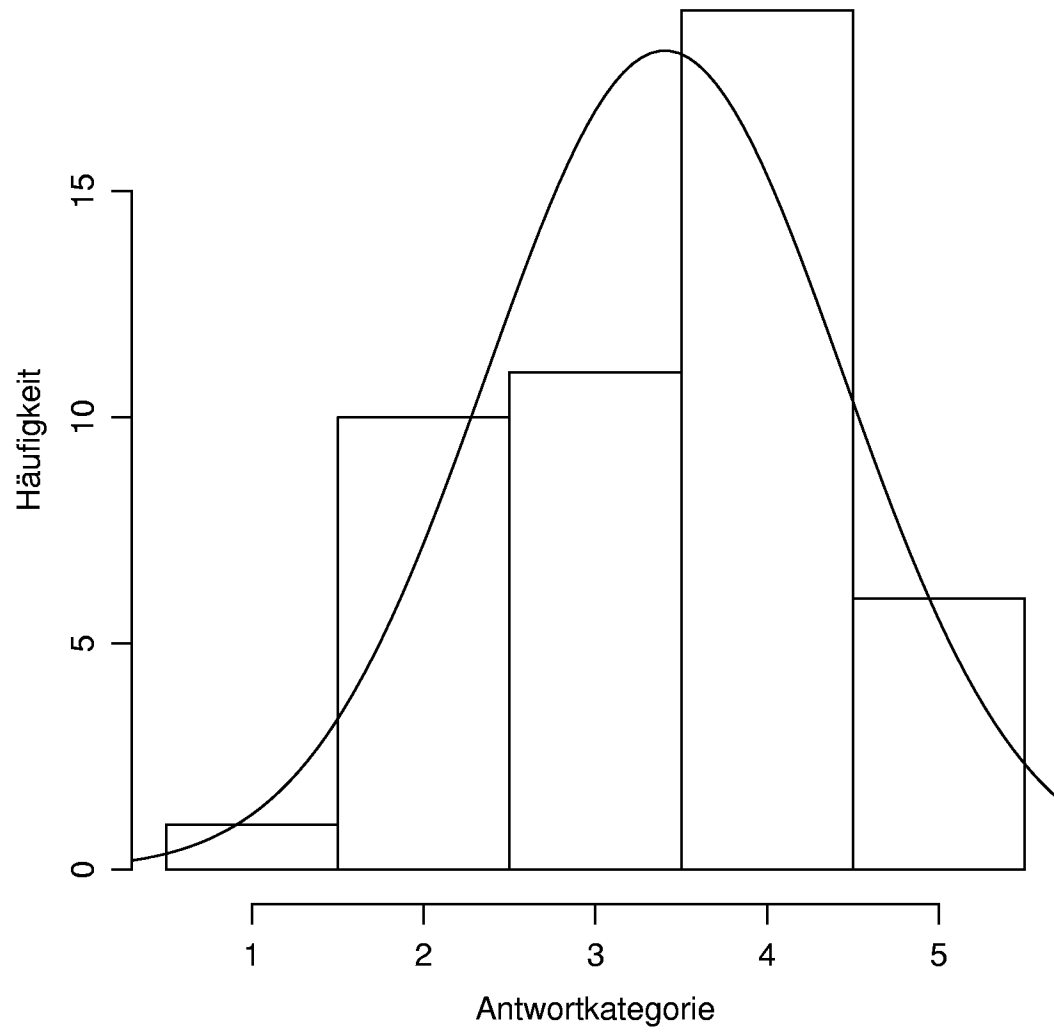


Abb. 65: Verteilung der Antworten für Item E02

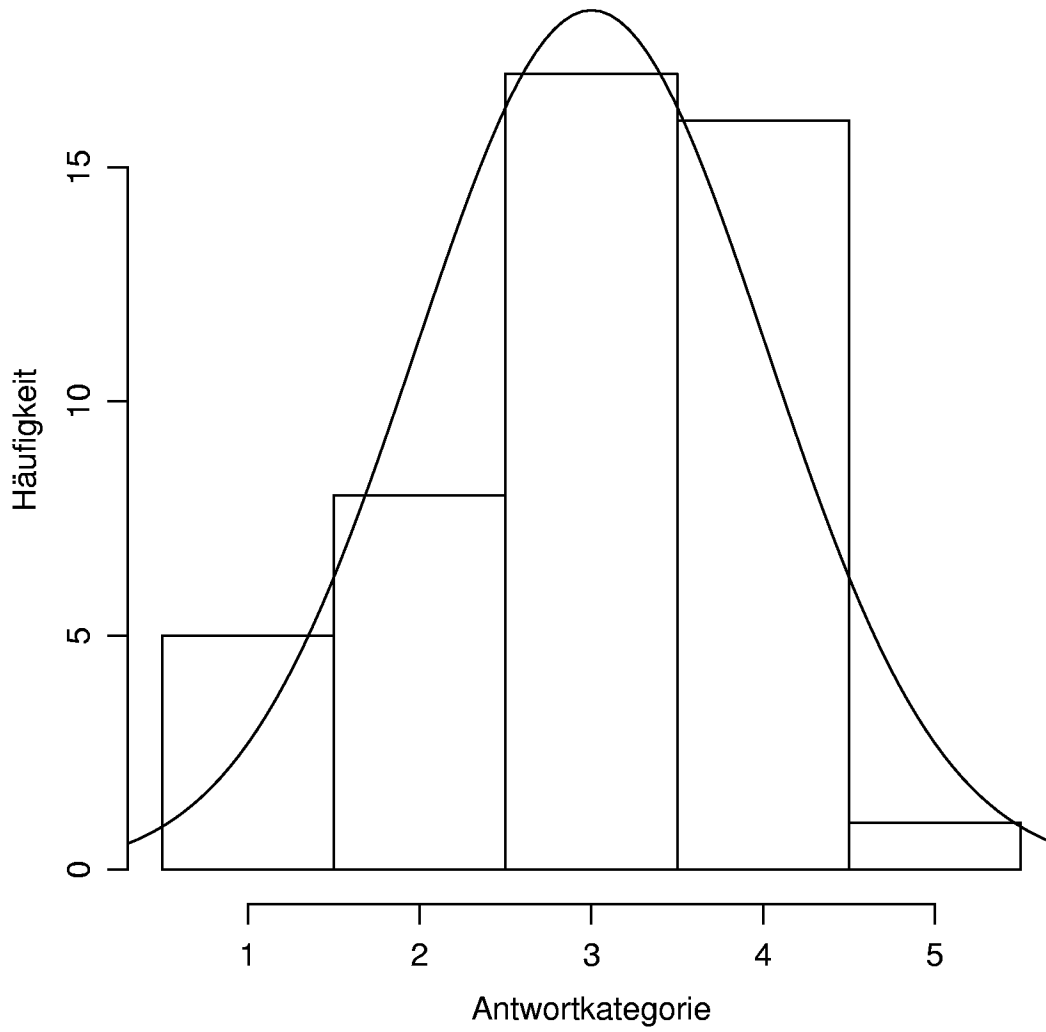


Abb. 66: Verteilung der Antworten für Item E03

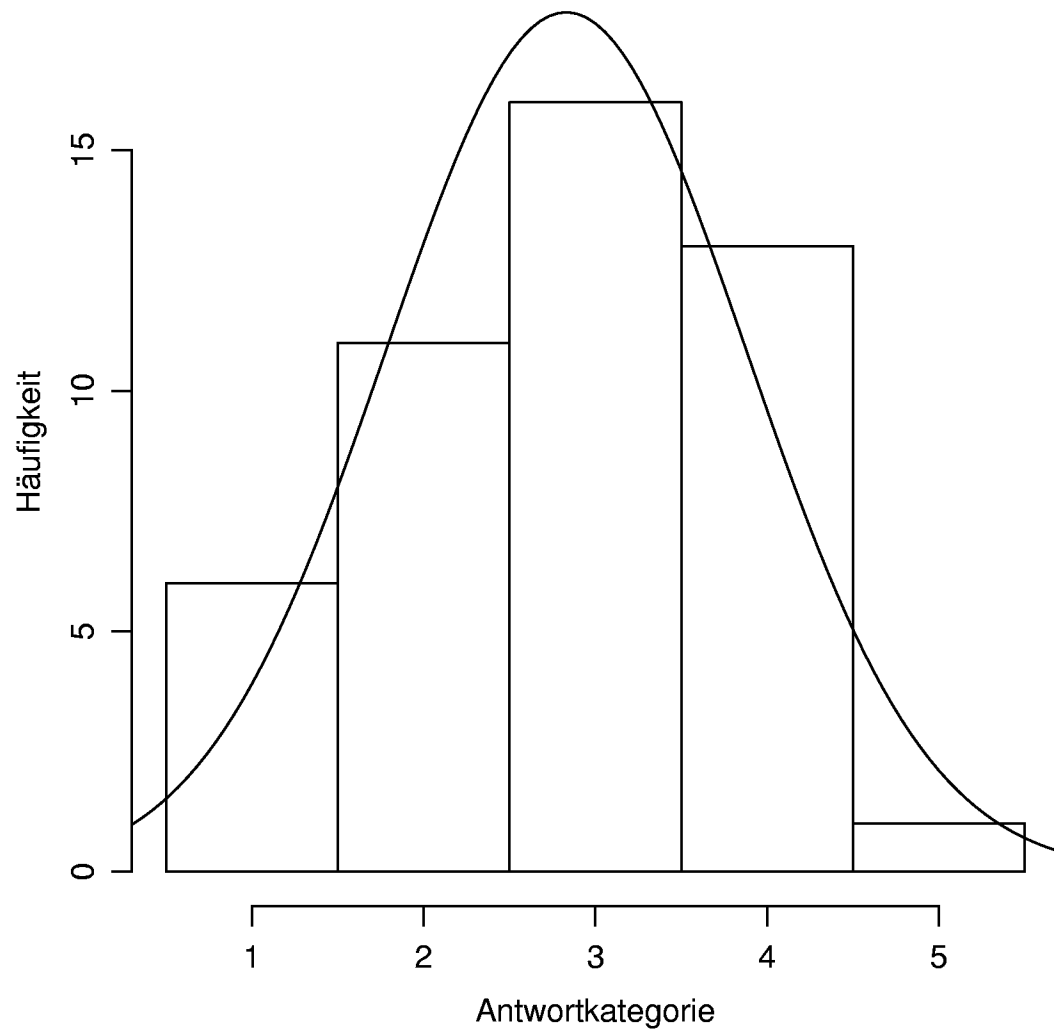


Abb. 67: Verteilung der Antworten für Item E04

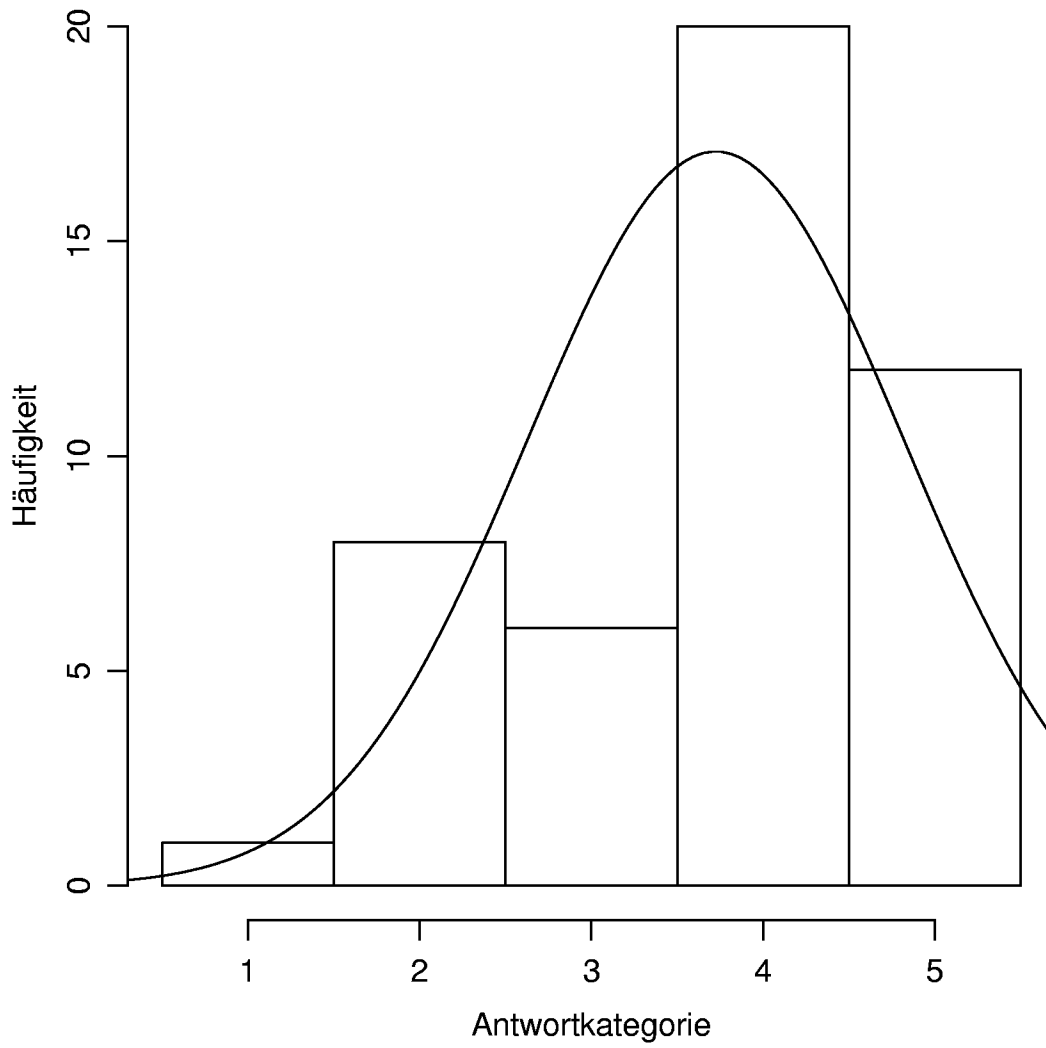


Abb. 68: Verteilung der Antworten für Item E07

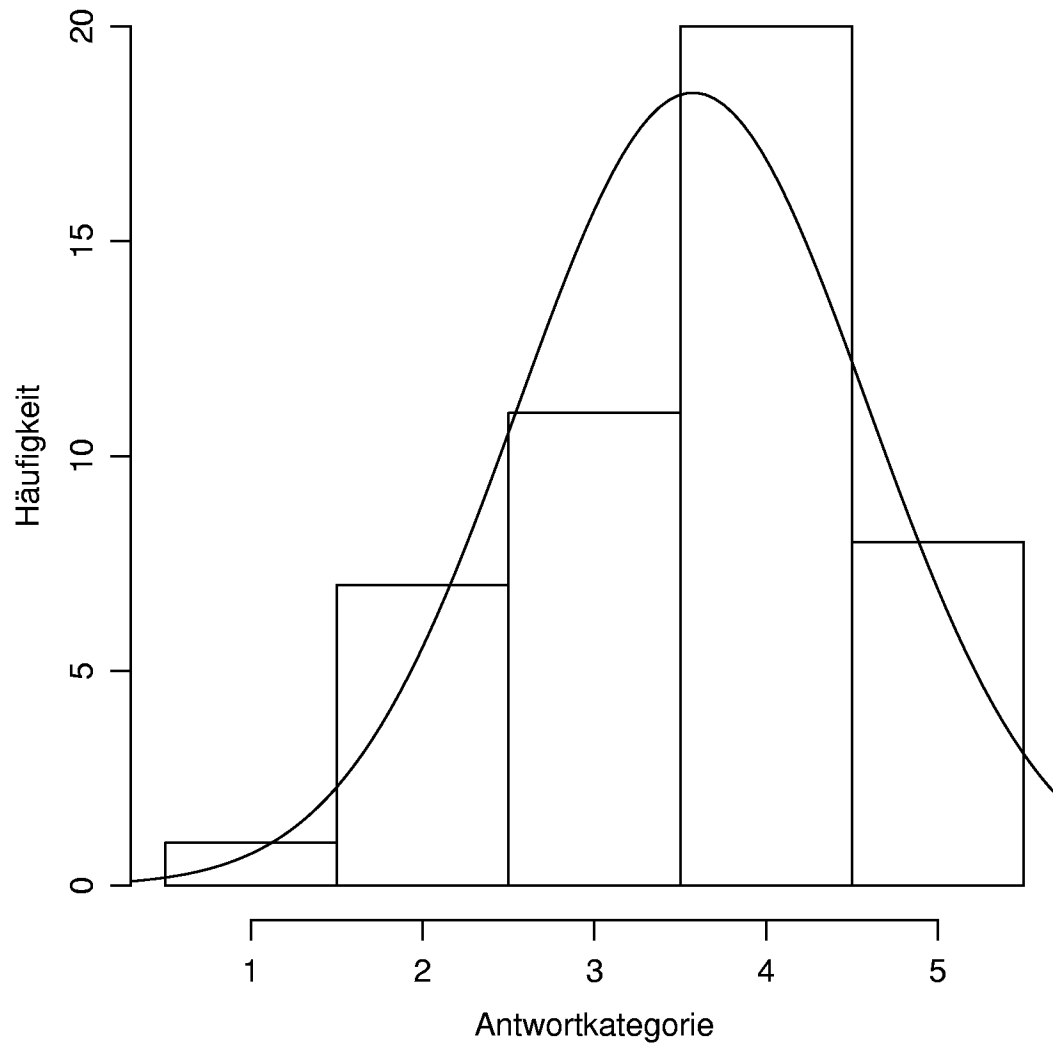


Abb. 69: Verteilung der Antworten für Item E11

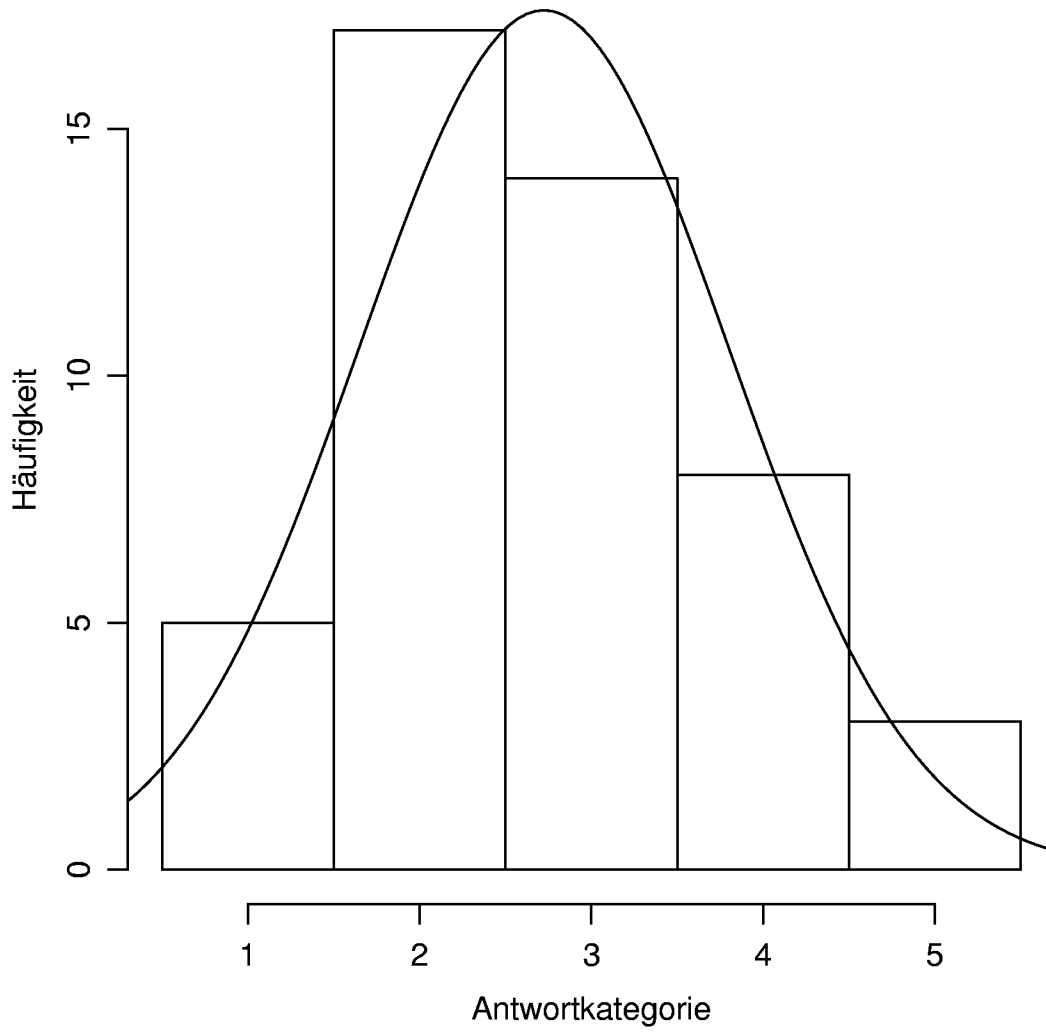


Abb. 70: Verteilung der Antworten für Item E12

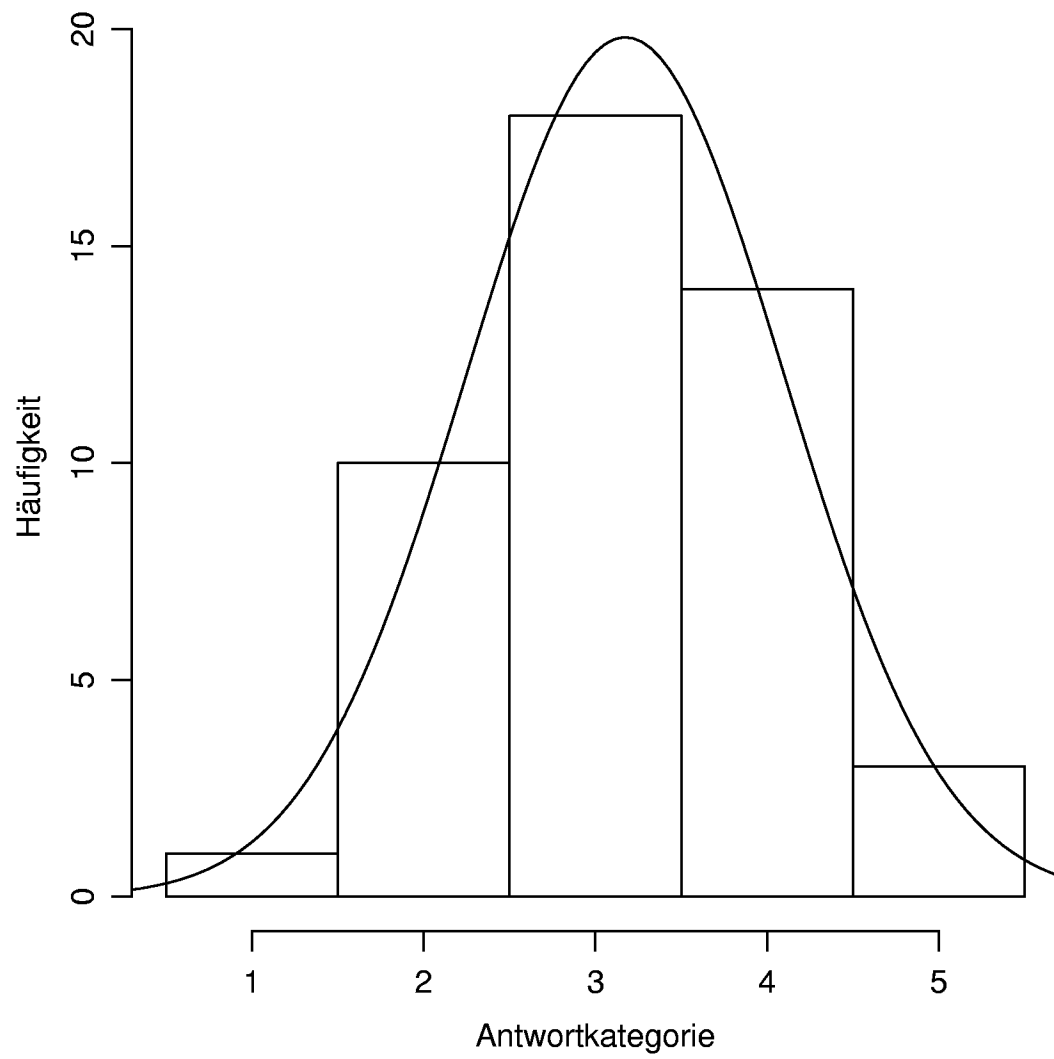


Abb. 71: Verteilung der Antworten für Item E13

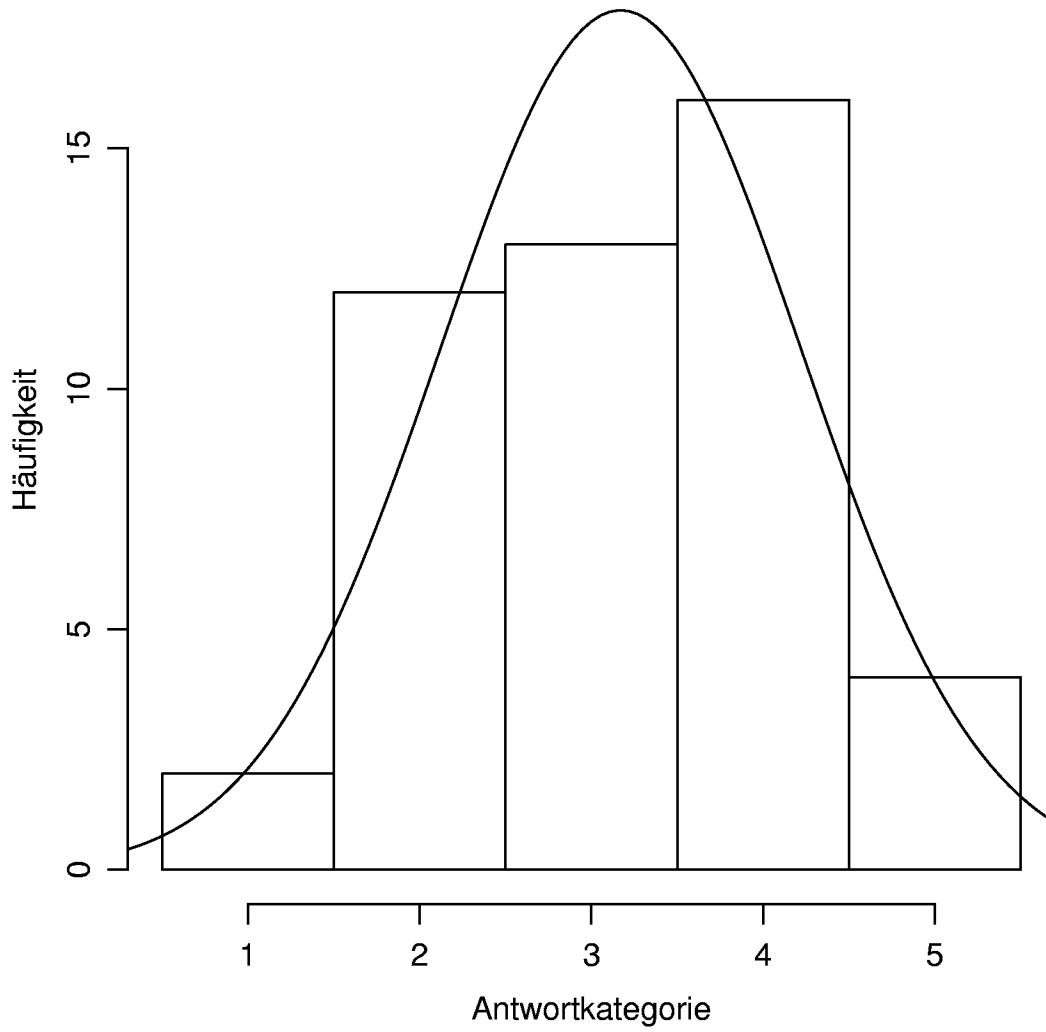


Abb. 72: Verteilung der Antworten für Item E14

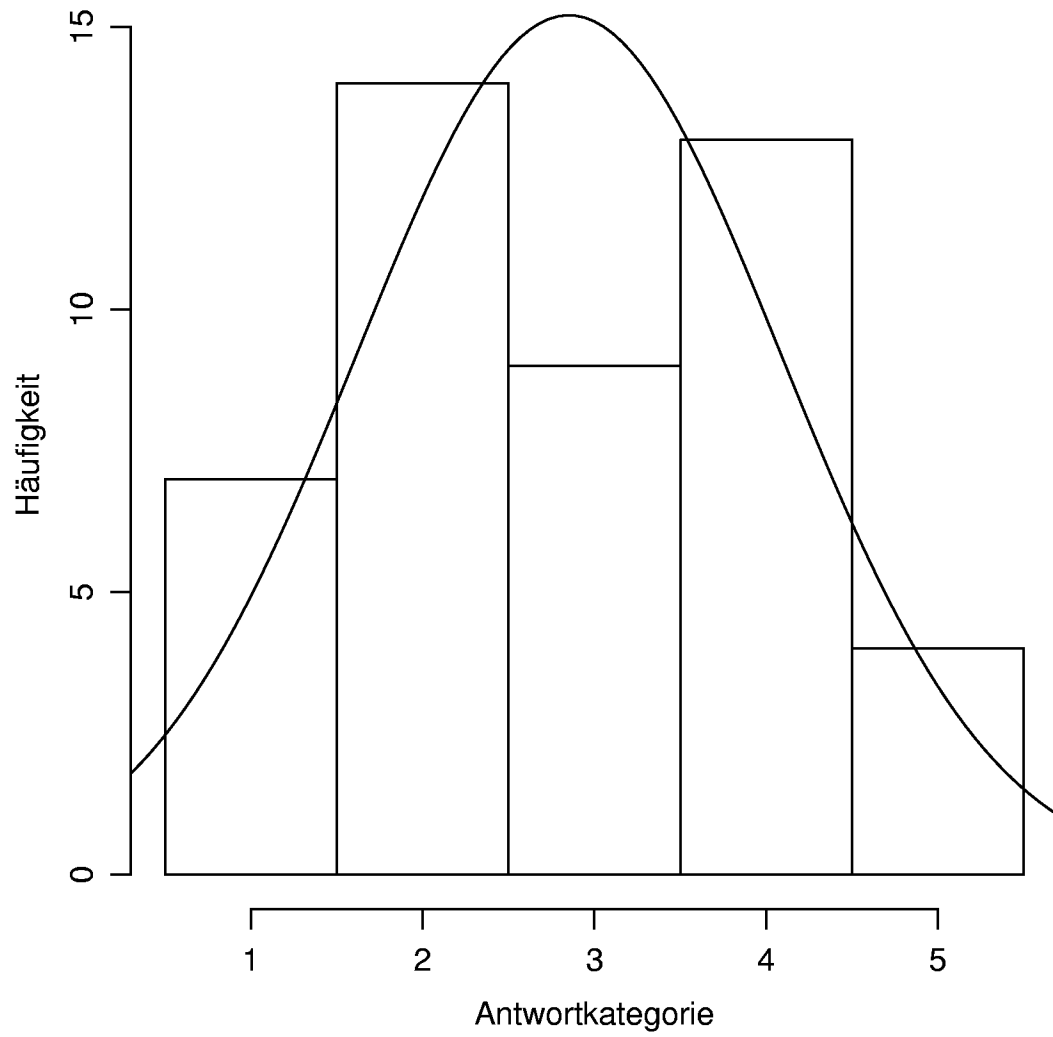


Abb. 73: Verteilung der Antworten für Item E15

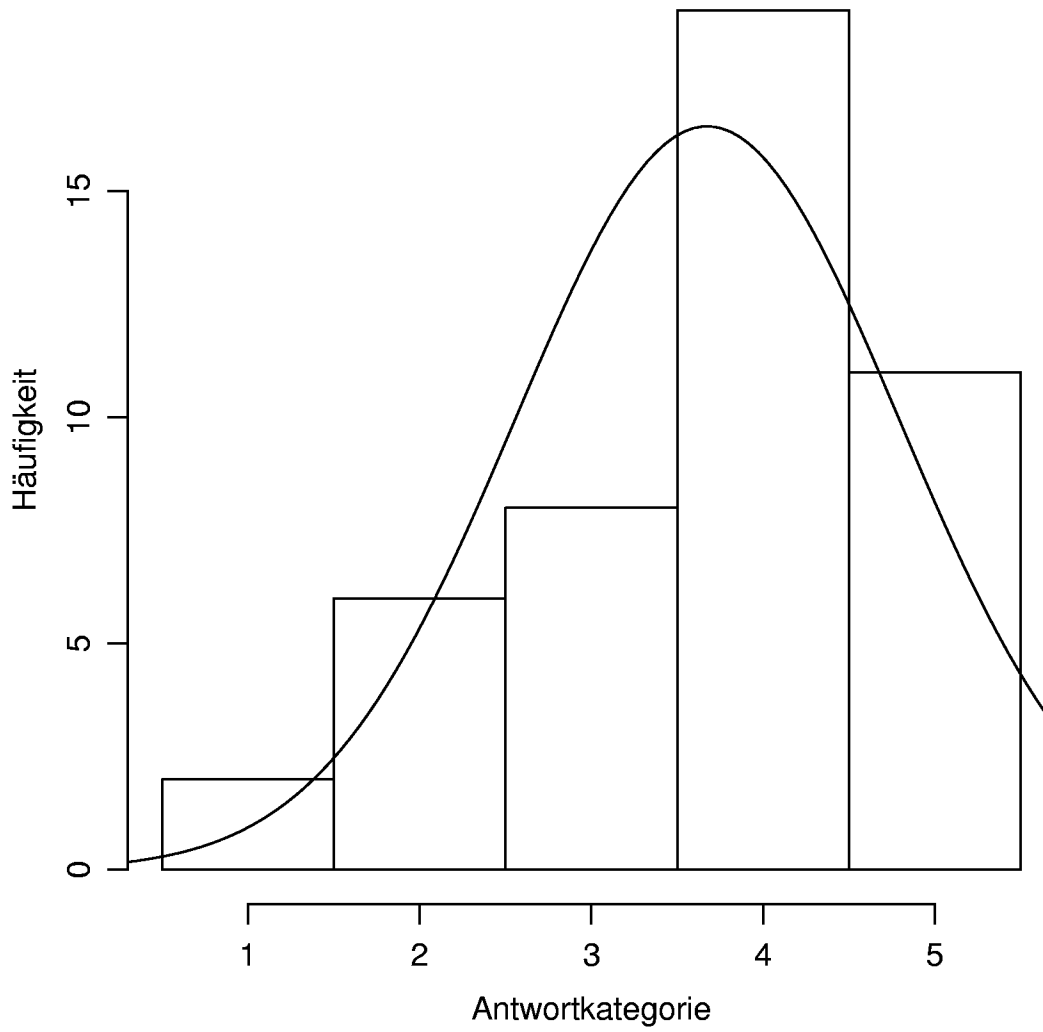


Abb. 74: Verteilung der Antworten für Item E16

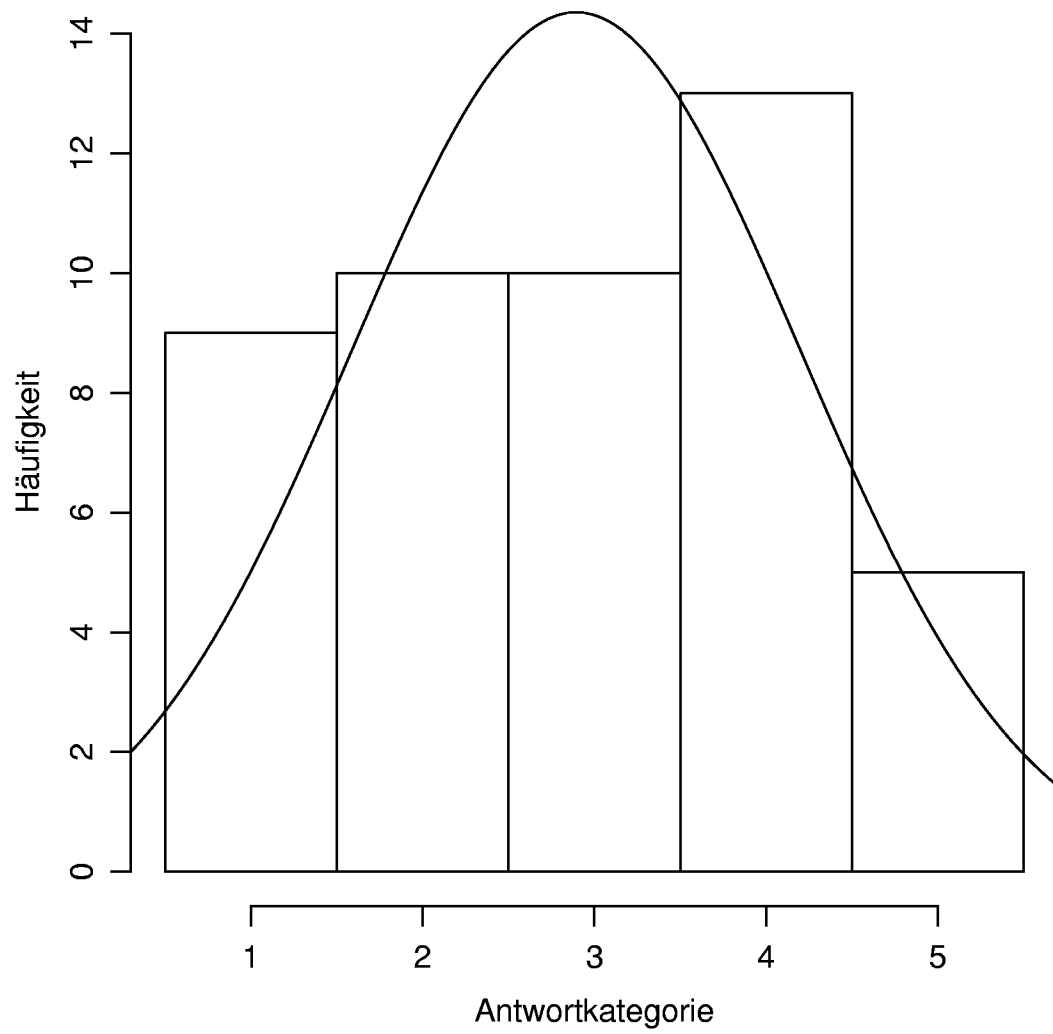


Abb. 75: Verteilung der Antworten für Item E17

Anhang I

Routen des Prototypen

Route 1

Start

Neuwerker Weg

90547 Stein

Deutschland

Ziel

Raby Street

M16 Manchester

Großbritannien

Route 2

Start

Raby Street

M16 Manchester

Großbritannien

Ziel

Neuwerker Weg

90547 Stein

Deutschland

Route 3

Start

Varnhagenstraße

10439 Berlin

Deutschland

Ziel

Emile Wittmannstraat
1030 Brüssel
Belgien

Route 4
Start
Emile Wittmannstraat
1030 Brüssel
Belgien

Ziel
Varnhagenstraße
10439 Berlin
Deutschland

Route 5
Start
Varnhagenstraße
10439 Berlin
Deutschland

Zwischenstop
Markt
52062 Aachen
Deutschland

Ziel
Emile Wittmannstraat
1030 Brüssel
Belgien

Route 6

Start

Emile Wittmannstraat

1030 Brüssel

Belgien

Zwischenstop

Markt

52062 Aachen

Deutschland

Ziel

Varnhagenstraße

10439 Berlin

Deutschland

Route 7

Start

Varnhagenstraße

10439 Berlin

Deutschland

Ziel

Markt

52062 Aachen

Deutschland

Route 8

Start

Markt

52062 Aachen

Deutschland

Ziel

Emile Wittmannstraat

1030 Brüssel

Belgien

Route 9

Start

Emile Wittmannstraat

1030 Brüssel

Belgien

Ziel

Markt

52062 Aachen

Deutschland

Route 10

Start

Markt

52062 Aachen

Deutschland

Ziel

Varnhagenstraße

10439 Berlin

Deutschland

Route 11

Start

Haußmannstraße

70188 Stuttgart (Ost)
Deutschland

Ziel

Max-Reimann-Straße
18556 Altenkirchen
Deutschland

Route 12

Start

Max-Reimann-Straße
18556 Altenkirchen
Deutschland

Ziel

Haußmannstraße
70188 Stuttgart (Ost)
Deutschland

Route 13

Start

Haußmannstraße
70188 Stuttgart (Ost)
Deutschland

Zwischenstop

Hugenottenstraße
68229 Neu-Edingen
Deutschland

Ziel

Max-Reimann-Straße
18556 Altenkirchen
Deutschland

Route 14

Start

Max-Reimann-Straße
18556 Altenkirchen
Deutschland

Zwischenstop

Hugenottenstraße
68229 Neu-Edingen
Deutschland

Ziel

Haußmannstraße
70188 Stuttgart (Ost)
Deutschland

Route 15

Start

Haußmannstraße
70188 Stuttgart (Ost)
Deutschland

Ziel

Hugenottenstraße
68229 Neu-Edingen
Deutschland

Route 16

Start

Hugenottenstraße
68229 Neu-Edingen
Deutschland

Ziel

Max-Reimann-Straße
18556 Altenkirchen
Deutschland

Route 17

Start

Max-Reimann-Straße
18556 Altenkirchen
Deutschland

Ziel

Hugenottenstraße
68229 Neu-Edingen
Deutschland

Route 18

Start

Hugenottenstraße
68229 Neu-Edingen
Deutschland

Ziel

Haußmannstraße
70188 Stuttgart (Ost)
Deutschland

Route 19

Start

Scharnhorststraße

26131 Oldenburg

Deutschland

Ziel

Calata Nord

17047 Savona

Italien

Route 20

Start

Calata Nord

17047 Savona

Italien

Ziel

Scharnhorststraße

26131 Oldenburg

Deutschland

Route 21

Start

Linder Höhe

51447 Köln

Deutschland

Ziel

Kirchlyok & tja 205

1039 Budapest
Ungarn

Route 22

Start

Királyok tja 205

1039 Budapest

Ungarn

Ziel

Linder Höhe

51447 Köln

Deutschland

Route 23

Start

Severinsmühlengasse

50678 Köln

Deutschland

Ziel

Rue du Grouanoch

29570 Camaret-sur-Mer

Frankreich

Route 24

Start

Rue du Grouanoch

29570 Camaret-sur-Mer

Frankreich

Ziel
Severinsmühlengasse
50678 Köln
Deutschland

Route 25
Start
Auf der Rübekuhle
21335 Lüneburg
Deutschland

Ziel
Wiesengasse
8020 Graz
Österreich

Route 26
Start
Wiesengasse
8020 Graz
Österreich

Ziel
Auf der Rübekuhle
21335 Lüneburg
Deutschland

Route 27
Start
Auf der Rübekuhle
21335 Lüneburg

Deutschland

Zwischenstop

Heinrich-Beck-Straße

09112 Chemnitz

Deutschland

Ziel

Wiesengasse

8020 Graz

Österreich

Route 28

Start

Wiesengasse

8020 Graz

Österreich

Zwischenstop

Heinrich-Beck-Straße

09112 Chemnitz

Deutschland

Ziel

Auf der Rübekuhle

21335 Lüneburg

Deutschland

Route 29

Start

Auf der Rübekuhle

21335 Lüneburg

Deutschland

Ziel

Heinrich-Beck-Straße

09112 Chemnitz

Deutschland

Route 30

Start

Heinrich-Beck-Straße

09112 Chemnitz

Deutschland

Ziel

Wiesengasse

8020 Graz

Österreich

Route 31

Start

Wiesengasse

8020 Graz

Österreich

Ziel

Heinrich-Beck-Straße

09112 Chemnitz

Deutschland

Route 32

Start

Heinrich-Beck-Straße
09112 Chemnitz
Deutschland

Ziel
Auf der Rübekuhle
21335 Lüneburg
Deutschland

Route 33
Start
Dechbettener Straße
93049 Regensburg
Deutschland

Ziel
Alter Postweg
21745 Hemmoor
Deutschland

Route 34
Start
Alter Postweg
21745 Hemmoor
Deutschland

Ziel
Dechbettener Straße
93049 Regensburg
Deutschland

Route 35

Start

Dechbettener Straße

93049 Regensburg

Deutschland

Zwischenstop

Albert-Magnus-Straße

93053 Regensburg

Deutschland

Ziel

Alter Postweg

21745 Hemmoor

Deutschland

Route 36

Start

Alter Postweg

21745 Hemmoor

Deutschland

Zwischenstop

Albert-Magnus-Straße

93053 Regensburg

Deutschland

Ziel

Dechbettener Straße

93049 Regensburg

Deutschland

Route 37

Start

Dechbettener Straße

93049 Regensburg

Deutschland

Ziel

Albert-Magnus-Straße

93053 Regensburg

Deutschland

Route 38

Start

Albert-Magnus-Straße

93053 Regensburg

Deutschland

Ziel

Alter Postweg

21745 Hemmoor

Deutschland

Route 39

Start

Alter Postweg

21745 Hemmoor

Deutschland

Ziel

Albert-Magnus-Straße

93053 Regensburg

Deutschland

Route 40

Start

Albert-Magnus-Straße

93053 Regensburg

Deutschland

Ziel

Dechbettener Straße

93049 Regensburg

Deutschland

Route 41

Start

Glauburgstraße

60318 Frankfurt

Deutschland

Ziel

Viale delle Provincie

00162 Rom

Italien

Route 42

Start

Viale delle Provincie

00162 Rom

Italien

Ziel

Glauburgstraße

60318 Frankfurt
Deutschland

Route 43
Start
Wilhelmshöher Allee
34131 Kassel
Deutschland

Ziel
Lill-Jans Plan
11425 Stockholm
Schweden

Route 44
Start
Lill-Jans Plan
11425 Stockholm
Schweden

Ziel
Wilhelmshöher Allee
34131 Kassel
Deutschland

Route 45
Start
Am Maschhop
29649 Wietendorf
Deutschland

Ziel

Klostermauer

29413 Diesdorf

Deutschland

Route 46

Start

Klostermauer

29413 Diesdorf

Deutschland

Ziel

Am Maschhop

29649 Wietzendorf

Deutschland

Route 47

Start

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Ziel

Fährsteg

21339 Lüneburg

Deutschland

Route 48

Start

Fährsteg

21339 Lüneburg

Deutschland

Ziel

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Route 49

Start

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Zwischenstop

Meinekenhop

21339 Lüneburg

Deutschland

Ziel

Fährsteg

21339 Lüneburg

Deutschland

Route 50

Start

Fährsteg

21339 Lüneburg

Deutschland

Zwischenstop

Meinekenhop

21339 Lüneburg

Deutschland

Ziel

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Route 51

Start

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Ziel

Meinekenhop

21339 Lüneburg

Deutschland

Route 52

Start

Meinekenhop

21339 Lüneburg

Deutschland

Ziel

Fährsteg

21339 Lüneburg

Deutschland

Route 53

Start

Fährsteg

21339 Lüneburg

Deutschland

Ziel

Meinekenhop

21339 Lüneburg

Deutschland

Route 54

Start

Meinekenhop

21339 Lüneburg

Deutschland

Ziel

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Route 55

Start

Flughafenstraße

40474 Düsseldorf

Deutschland

Ziel

Gelsenkirchener Straße

45309 Essen

Deutschland

Route 56

Start

Gelsenkirchener Straße

45309 Essen

Deutschland

Ziel

Flughafenstraße

40474 Düsseldorf

Deutschland

Route 57

Start

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Ziel

Hamburger Straße

21339 Lüneburg

Deutschland

Route 58

Start

Hamburger Straße

21339 Lüneburg

Deutschland

Ziel

Wilschenbrucher Weg

21335 Lüneburg

Deutschland

Route 59

Start

Auwiesenstraße

72770 Reutlingen

Deutschland

Ziel

Porscheplatz

70435 Stuttgart

Deutschland

Route 60

Start

Porscheplatz

70435 Stuttgart

Deutschland

Ziel

Auwiesenstraße

72770 Reutlingen

Deutschland

Route 61

Start

Auwiesenstraße

72770 Reutlingen

Deutschland

Zwischenstop

Herrenberger Straße
72070 Tübingen
Deutschland

Ziel
Porscheplatz
70435 Stuttgart
Deutschland

Route 62
Start
Porscheplatz
70435 Stuttgart
Deutschland

Zwischenstop
Herrenberger Straße
72070 Tübingen
Deutschland

Ziel
Auwiesenstraße
72770 Reutlingen
Deutschland

Route 63
Start
Auwiesenstraße
72770 Reutlingen
Deutschland

Ziel

Herrnberger Straße
72070 Tübingen
Deutschland

Route 64
Start
Herrnberger Straße
72070 Tübingen
Deutschland

Ziel
Porscheplatz
70435 Stuttgart
Deutschland

Route 65
Start
Porscheplatz
70435 Stuttgart
Deutschland

Ziel
Herrnberger Straße
72070 Tübingen
Deutschland

Route 66
Start
Herrnberger Straße
72070 Tübingen
Deutschland

Ziel

Auwiesenstraße

72770 Reutlingen

Deutschland

Route 67

Start

Baukampstraße

47059 Duisburg

Deutschland

Zwischenstop

Flockenweg

45478 Mülheim a. d. Ruhr

Deutschland

Ziel

Königsallee

44789 Bochum

Deutschland

Route 68

Start

Baukampstraße

47059 Duisburg

Deutschland

Ziel

Flockenweg

45478 Mülheim a. d. Ruhr

Deutschland

Route 69

Start

Flockenweg

45478 Mülheim a. d. Ruhr

Deutschland

Ziel

Königsallee

44789 Bochum

Deutschland

Route 70

Start

Zur Schmiede

33098 Paderborn

Deutschland

Ziel

Schiphol Boulevard

1118 Schiphol

Niederlande

Route 71

Start

Schiphol Boulevard

1118 Schiphol

Niederlande

Ziel

Zur Schmiede

33098 Paderborn
Deutschland

Route 72
Start
Auf der Bunte
29683 Bad Fallingbostel
Deutschland

Ziel
Ossenhoek
3253 Ouddorp
Niederlande

Route 73
Start
Ossenhoek
3253 Ouddorp
Niederlande

Ziel
Auf der Bunte
29683 Bad Fallingbostel
Deutschland

Route 74
Start
Schönwalder Landstraße
17491 Greifswald
Deutschland

Ziel

Calle de Pablo Picasso

36618 Vilagarcia de Arousa

Spanien

Route 75

Start

Calle de Pablo Picasso

36618 Vilagarcia de Arousa

Spanien

Ziel

Schönwalder Landstraße

17491 Greifswald

Deutschland

Anhang J

Fragebogen

Ich brauche keine besonderen Kenntnisse für die Nutzung von ROUTENPLANER.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Indem ich ROUTENPLANER nutze kann ich Routen schneller planen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Ich fühle mich durch ROUTENPLANER nicht bevormundet.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Ich muss in ROUTENPLANER keine für mich unnötigen Arbeitsschritte machen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Falls ROUTENPLANER für mich verfügbar wäre, würde ich es benutzen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Die Bedienung von ROUTENPLANER funktioniert wie in mir bekannten Programmen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Wie ROUTENPLANER funktioniert ist für mich leicht zu erkennen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Die Funktion von allen Bedienelementen ist offensichtlich.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Wichtige Funktionen sind für mich in ROUTENPLANER schnell erreichbar.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Indem ich ROUTENPLANER nutze kann ich Routen leichter planen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Ich kann meine eigenen Fehler in ROUTENPLANER problemlos korrigieren.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Indem ich ROUTENPLANER nutze habe ich bessere Möglichkeiten, Routen zu planen.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Die Oberfläche von ROUTENPLANER wirkt auf mich gut strukturiert.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Wenn möglich würde ich ROUTENPLANER häufig nutzen.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Ich komme mit ROUTENPLANER auf Anhieb gut zurecht.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Mit ROUTENPLANER kann ich meine Ziele über wenige Arbeitsschritte erreichen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Die Struktur von ROUTENPLANER ist für mich leicht zu erkennen.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Auch selten genutzte Funktionen sind für mich in ROUTENPLANER gut erreichbar.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Ich finde ROUTENPLANER nützlich.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Die Struktur von ROUTENPLANER passt für mich gut zu meinen Aufgaben.

--	-	•	+	++
stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr
1	2	3	4	5

Indem ich ROUTENPLANER nutze kann ich Routen effizienter planen.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Ich werde durch die Oberfläche von ROUTENPLANER nicht eingeschränkt.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Es ist leicht für mich ein geschickter Nutzer von ROUTENPLANER zu werden.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Indem ich ROUTENPLANER nutze kann ich bessere Routen planen.

-- stimmt nicht	- stimmt wenig	• stimmt mittelmäßig	+ stimmt ziemlich	++ stimmt sehr
1	2	3	4	5

Anhang K

Planungsaufgaben

Aufgabe 1

Sie wollen nach der Uni Ihr Auto in die Autowerkstatt bringen. Bitte planen Sie, wie Sie zum Autohaus Krull fahren.

Startort:

Wilschenbrucher Weg 84
21335 Lüneburg

Zielort:

Hamburger Str. 37
21339 Lüneburg

Frage: Wie lange dauert die Fahrt laut Routenplaner?

Aufgabe 2

Sie planen eine Dienstreise nach Manchester. Sie fahren in Stein bei Nürnberg los und wollen abends in Manchester ankommen.

Startort:

Neuwerker Weg 3
90547 Stein

Zielort:

Raby Street 12
M16 Manchester
Großbritannien

Frage: Nutzen Sie den Eurotunnel oder setzen Sie per Fähre über?

Aufgabe 3

Sie wollen später ins Kino. Dafür wollen Sie noch Bekannte abholen, die in der Stadt wohnen. Bitte planen Sie eine entsprechende Route (von hier aus).

Startort:

Wilschenbrucher Weg 84
21335 Lüneburg

Zwischenstopp:

Meinekenhop 4
21339 Lüneburg

Zielort (CineStar):

Fährsteg 1
21339 Lüneburg

Frage: Wie häufig führt der Routenplaner Sie über die Lindenstrasse?

Aufgabe 4

Sie wollen für eine Städtereise nach Rom fahren. Bitte planen Sie die entsprechende Route.

Startort:

Glauburgstr. 13
60318 Frankfurt

Zielort:

Viale delle Provincie
00161 Rom
Italien

Frage: Führt die Route über den Brenner (Österreich)?

Aufgabe 5

Im Sommer wollen Sie mit dem eigenen Wagen nach Korsika fahren. Sie wollen um 18:00Uhr am Hafen sein. Bitte planen Sie, wie sich nach Savona zur Fähre kommen.

Startort:

Scharnhorststr. 9
26131 Oldenburg

Zielort:

Calata Nord
17047 Savona
Italien

Frage: Welche Abfahrtszeit schlägt der Routenplaner vor?

Aufgabe 6

Sie wollen anlässlich einer Ausstellung einen Kurzurlaub in Stockholm verbringen. Bitte planen Sie die Route zu Ihrem Hotel in der Lill-Jans Plan.

Startort:

Wilhelmshöher Allee 34
34131 Kassel

Zielort:

Lill-Jans Plan 21
11425 Stockholm
Schweden

Frage: Wie weit ist es nach Schweden (Grenzübergang)?

Aufgabe 7

Planen Sie bitte die Rückreise aus Stockholm.

Frage: Weicht die Streckenführung der Rückreise von der Hinreise ab?

Aufgabe 8

Bitte planen Sie eine Fahrt von Wietzendorf nach Diesdorf.

Startort:

Am Maschhop 15
29649 Wietzendorf

Zielort:

Klostermauer 2
29413 Diesdorf

Frage: Führt die Route über die Autobahn?

Aufgabe 9

Sie sollen Bekannte vom Flughafen in Düsseldorf abholen. Direkt danach wollen Sie das Weltkulturerbe „Zeche Zollverein“ (Gelsenkirchener Str. 181) in Essen besichtigen. Planen Sie, wie Sie vom Flughafen zum Zechengelände fahren.

Startort:

Flughafenstraße
40474 Düsseldorf

Zielort:

Gelsenkirchener Str. 181
45309 Essen

Frage: An welcher Autobahnabfahrt sollen Sie die Autobahn verlassen?

Aufgabe 10

Sie wollen nach Brüssel fahren. Bitte planen Sie, wie Sie von Berlin nach Brüssel kommen.

Startort:

Varnhagenstr. 33
10439 Berlin

Zielort:

Rue Emile Wittmann
1030 Brüssel
Belgien

Frage: Führt die Strecke über die Niederlande?

Aufgabe 11

Sie wollen auf der Fahrt nach Brüssel (aus Aufgabe 10) noch einen Stadtbummel in Aachen einlegen. Bitte planen Sie einen entsprechenden Zwischenstop ein.

Zwischenstop:

Markt

52062 Aachen

Frage: Um wieviel Kilometer verlängert sich die Strecke?

Aufgabe 12

Sie wollen an einer Fortbildung in Graz teilnehmen. Da Sie zu dritt anreisen, wollen Sie mit dem Wagen fahren. Bitte planen Sie die Anreise so, daß Sie gegen 21:00Uhr in Graz sind.

Startort:

Auf der Rübekuhle 4

21335 Lüneburg

Zielort:

Wiesengasse 23

8020 Graz

Österreich

Frage: Wieviele Pausen empfiehlt der Routenplaner?

Aufgabe 13

Auf der Rückreise von Graz wollen Sie noch Bekannte in Chemnitz besuchen. Bitte planen Sie einen Zwischenstop ein.

Zwischenstop:

Heinrich-Beck-Str. 8

09112 Chemnitz

Frage: Um welche Uhrzeit sind Sie bei Ihren Bekannten in Chemnitz, wenn sie um 8:30Uhr losfahren?

Aufgabe 14

Sie wollen mit Bekannten aus Mülheim nach Bochum ins Theater. Bitte planen Sie die Route.

Startort:

Baukampstr. 13
47059 Duisburg

Zielort:

Königsallee 15
44789 Bochum

Zwischenstop:

Flockenweg 7
45478 Mülheim

Frage: Fahren Sie von Duisburg nach Mülheim über die Autobahn?

Aufgabe 15

Sie wollen an die Nordsee fahren. Bitte planen Sie, wie sie ins Surfzentrum „Browsersdam“ kommen.

Startort:

Auf der Bunte 3
29683 Bad Fallingbosten

Zielort:

Ossenhoek 1
3253 Ouddorp
Niederlande

Frage: Welches ist die späteste Uhrzeit abends, um die Sie unter Berücksichtigung der empfohlenen Pausen ankommen können?

Aufgabe 16

Sie wollen von Regensburg nach Hemmor fahren. Bitte planen Sie die Route

Startort:

Dechbettener Str. 34
93049 Regensburg

Zielort:

Alter Postweg 2
21745 Hemmoor

Frage: Wann müssen Sie laut Routenplaner losfahren, damit Sie um 18:00Uhr in Hemmoor sind?

Aufgabe 17

Auf der Rückfahrt von Hemmoor nach Regensburg wollen Sie noch einen Kommilitonen an der Uni absetzen. Planen Sie einen Zwischenstop ein.

Zwischenstop:

Albert-Magnus-Str. 5
93053 Regensburg

Frage: Müssen Sie die Autobahn an einer anderen Ausfahrt verlassen, als wenn Sie auf direktem Weg zurück fahren würden?

Wie heißen die entsprechenden beiden Autobahnabfahrten?

Aufgabe 18

Sie wollen zusammen mit Freunden aus Neu-Edingen (bei Mannheim) in eine Ferienwohnung nach Rügen fahren. Bitte planen Sie die entsprechende Route.

Startort:

Haußmannstr. 93
70188 Stuttgart

Zielort:

Max-Reimann-Str. 2
18556 Altenkirchen (Rügen)

Zwischenstop:

Hugenottenstr. 18
68229 Neu-Edingen

Frage: Stimmt die Aussage, daß Neu-Edingen ungefähr auf halber Strecke liegt?

Aufgabe 19

Auf dem Rückweg von Rügen wollen Sie direkt nach Stuttgart fahren, da Ihre Bekannten noch länger bleiben. Planen Sie die entsprechende Strecke.

Frage: Wie lange brauchen Sie nach Stuttgart, wenn Sie um 6:30Uhr losfahren?

Anhang L

Häufigkeitsverteilungen der Items (Empirische Überprüfung des Prozessmodells)

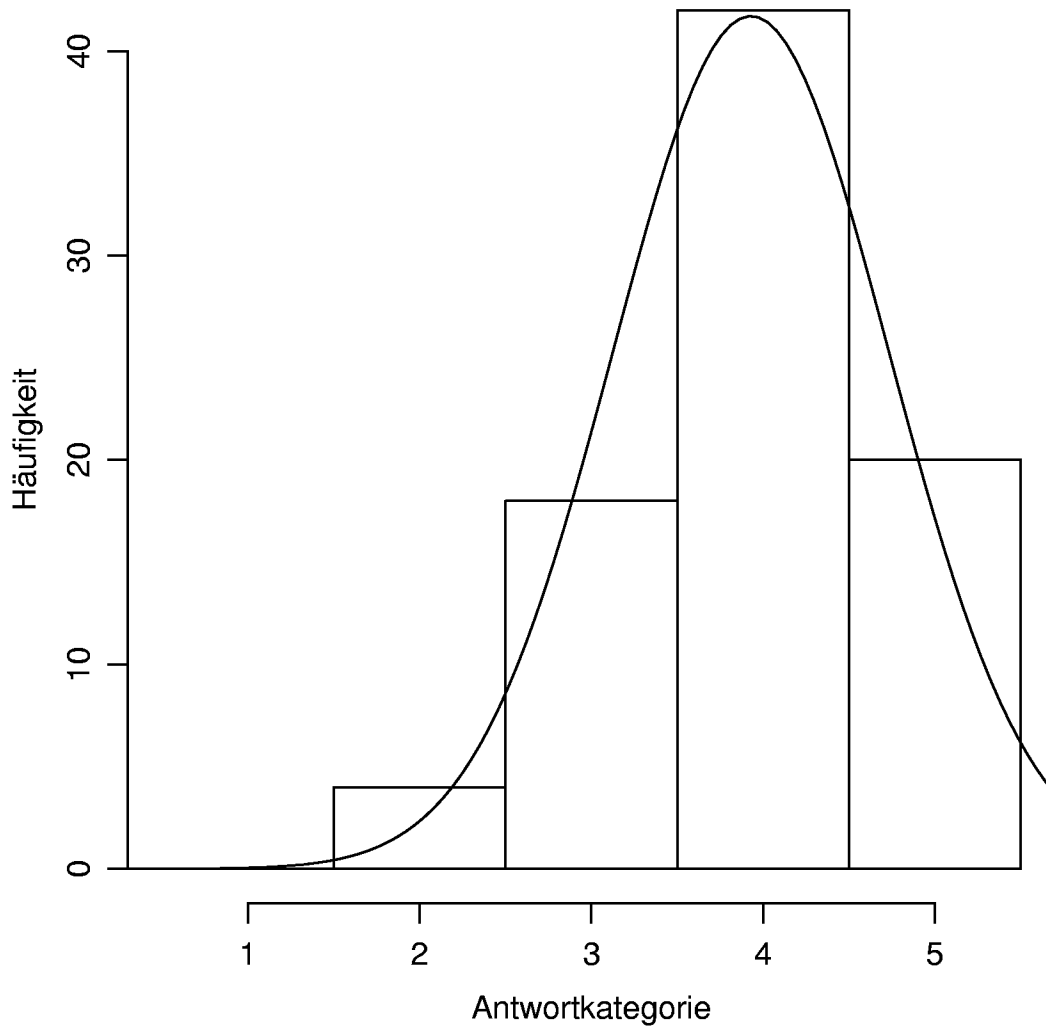


Abb. 76: Verteilung der Antworten für Item L02 zum ersten Messzeitpunkt

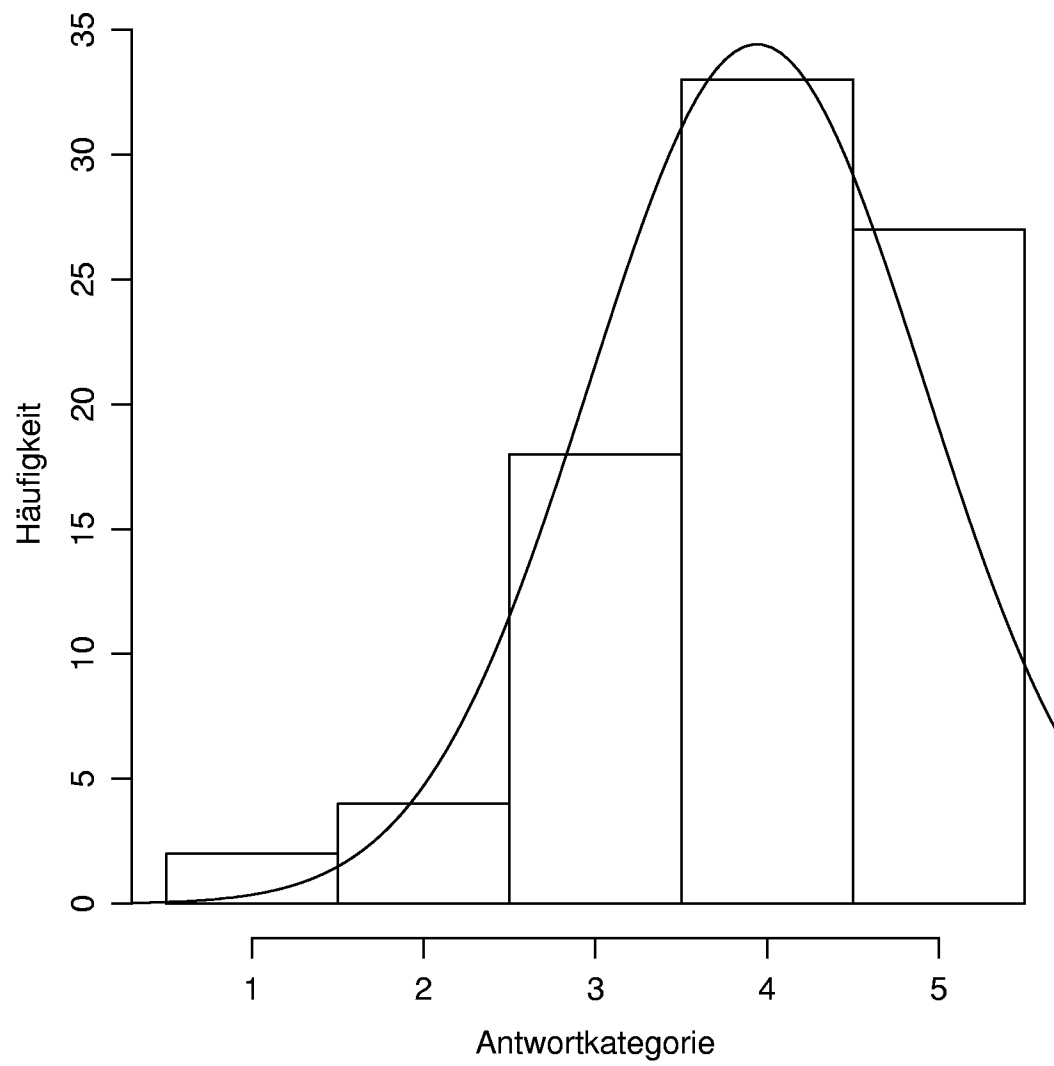


Abb. 77: Verteilung der Antworten für Item L02 zum zweiten Messzeitpunkt

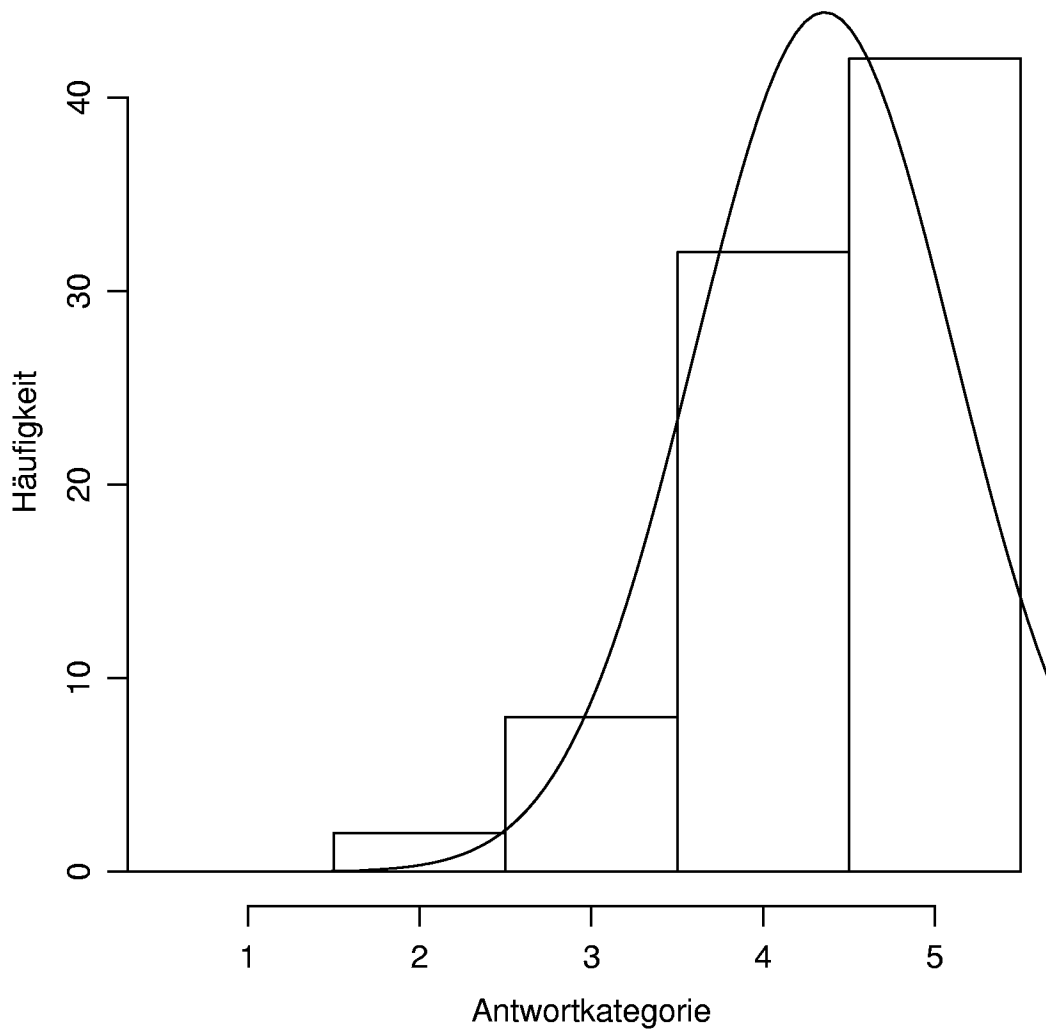


Abb. 78: Verteilung der Antworten für Item L06 zum ersten Messzeitpunkt

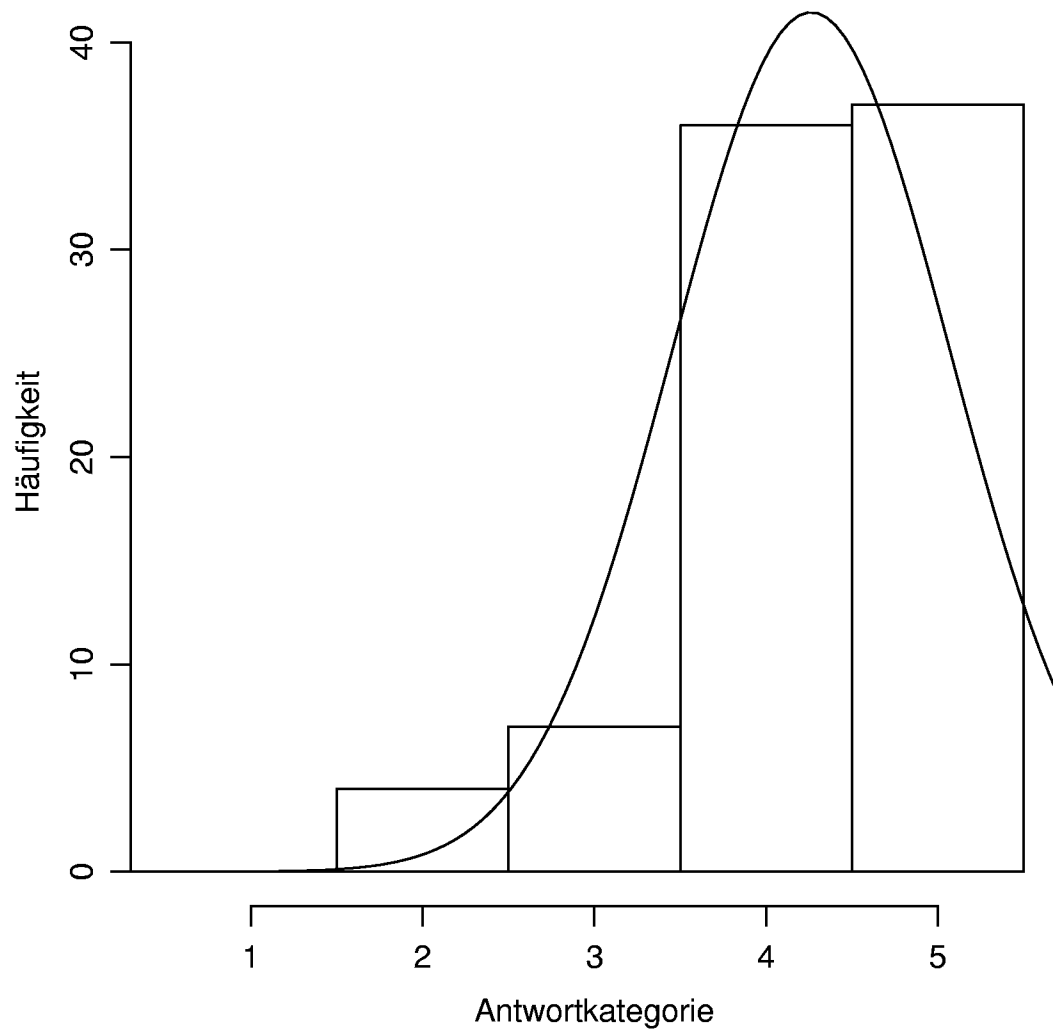


Abb. 79: Verteilung der Antworten für Item L06 zum zweiten Messzeitpunkt

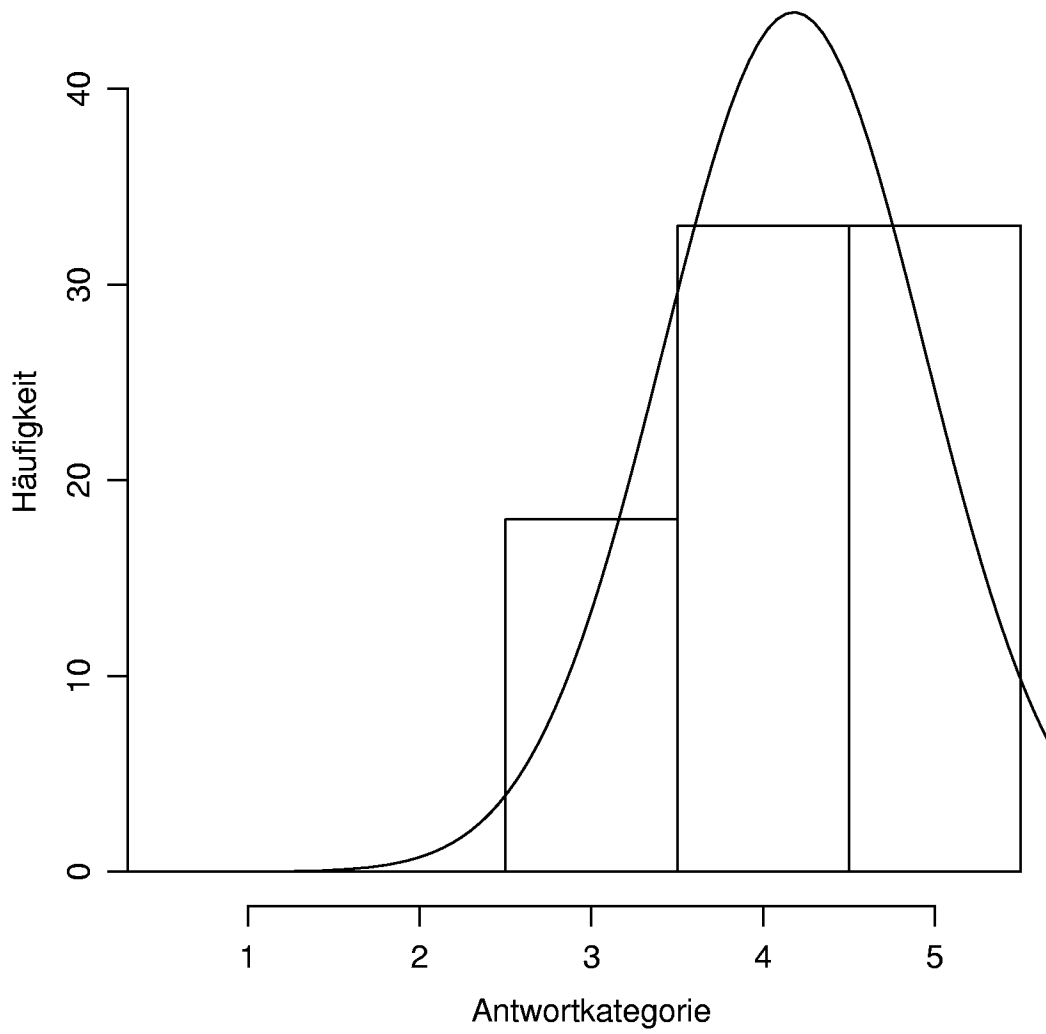


Abb. 80: Verteilung der Antworten für Item L07 zum ersten Messzeitpunkt

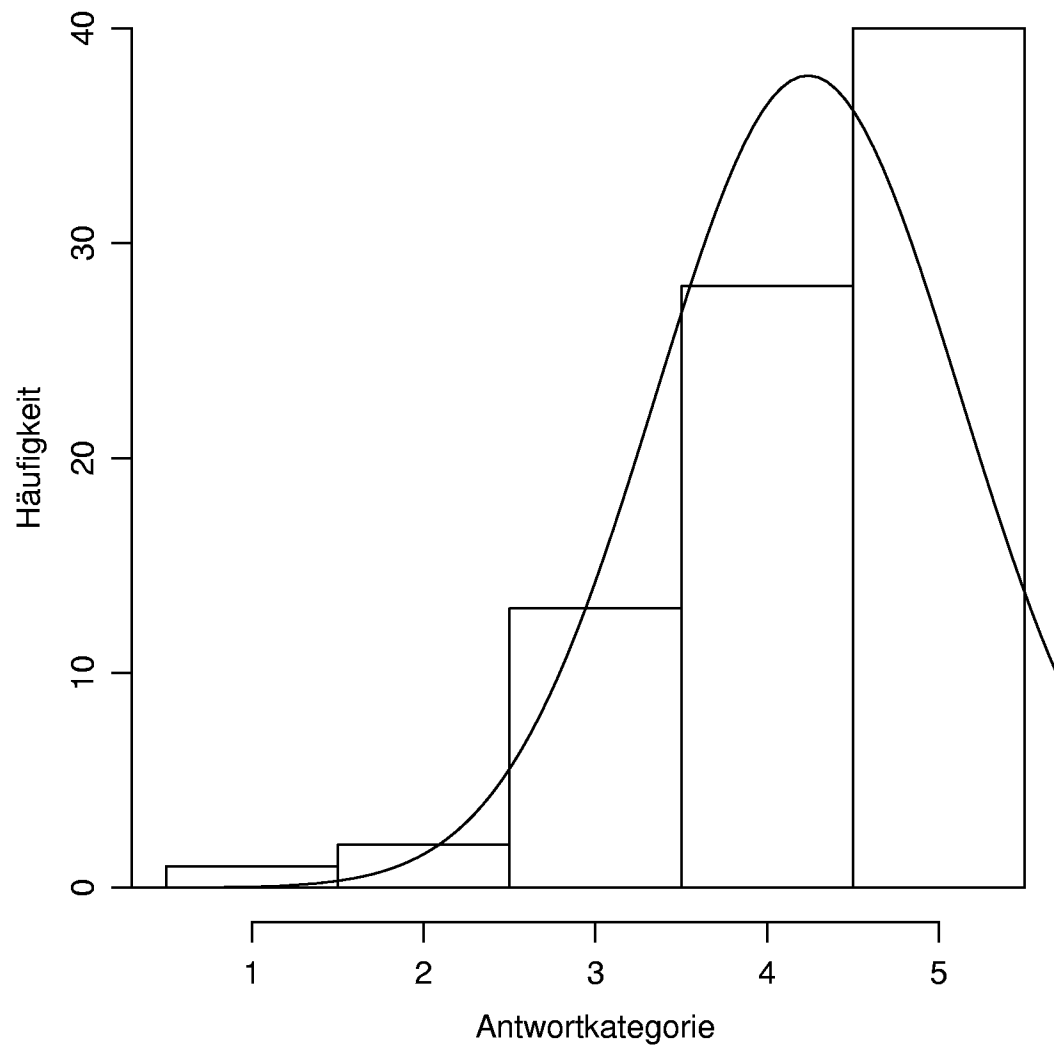


Abb. 81: Verteilung der Antworten für Item L07 zum zweiten Messzeitpunkt

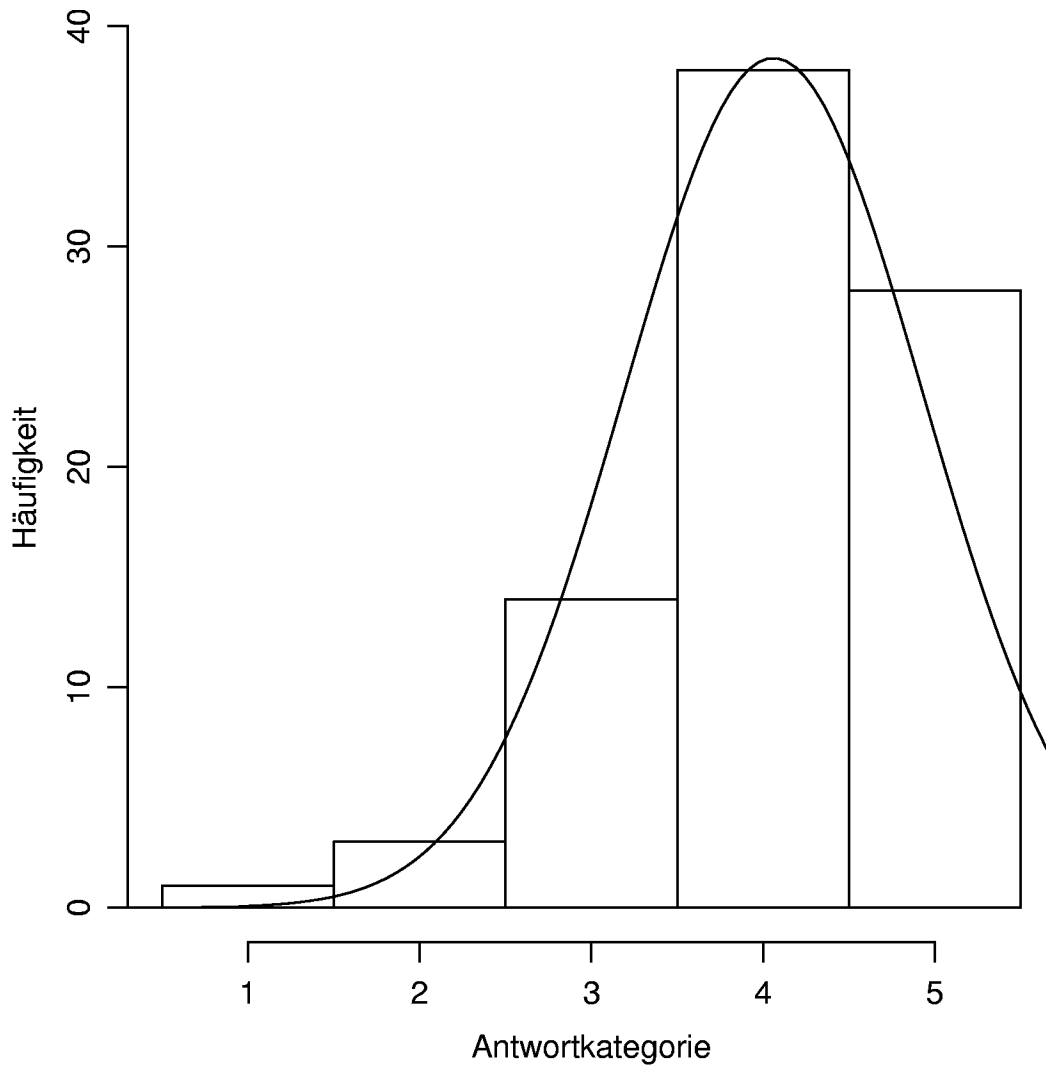


Abb. 82: Verteilung der Antworten für Item L08 zum ersten Messzeitpunkt

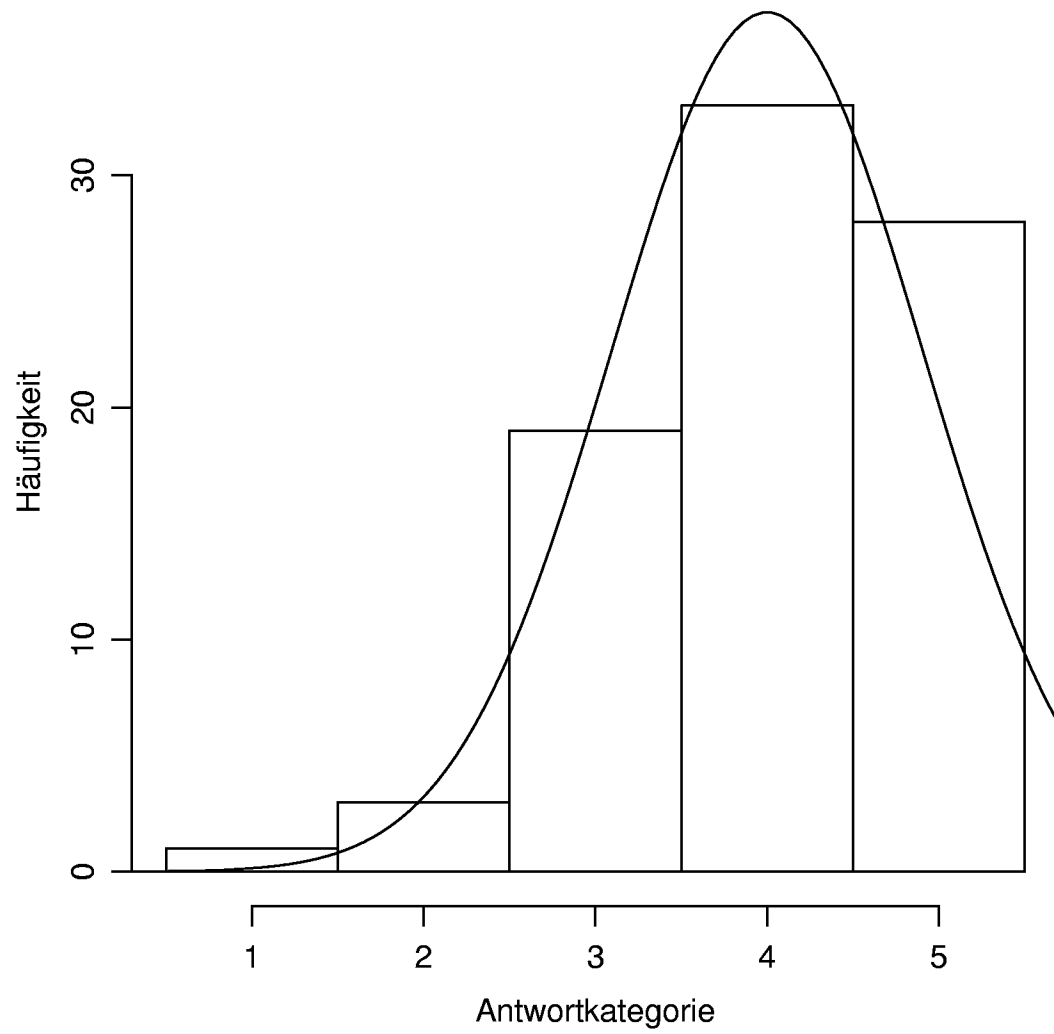


Abb. 83: Verteilung der Antworten für Item L08 zum zweiten Messzeitpunkt

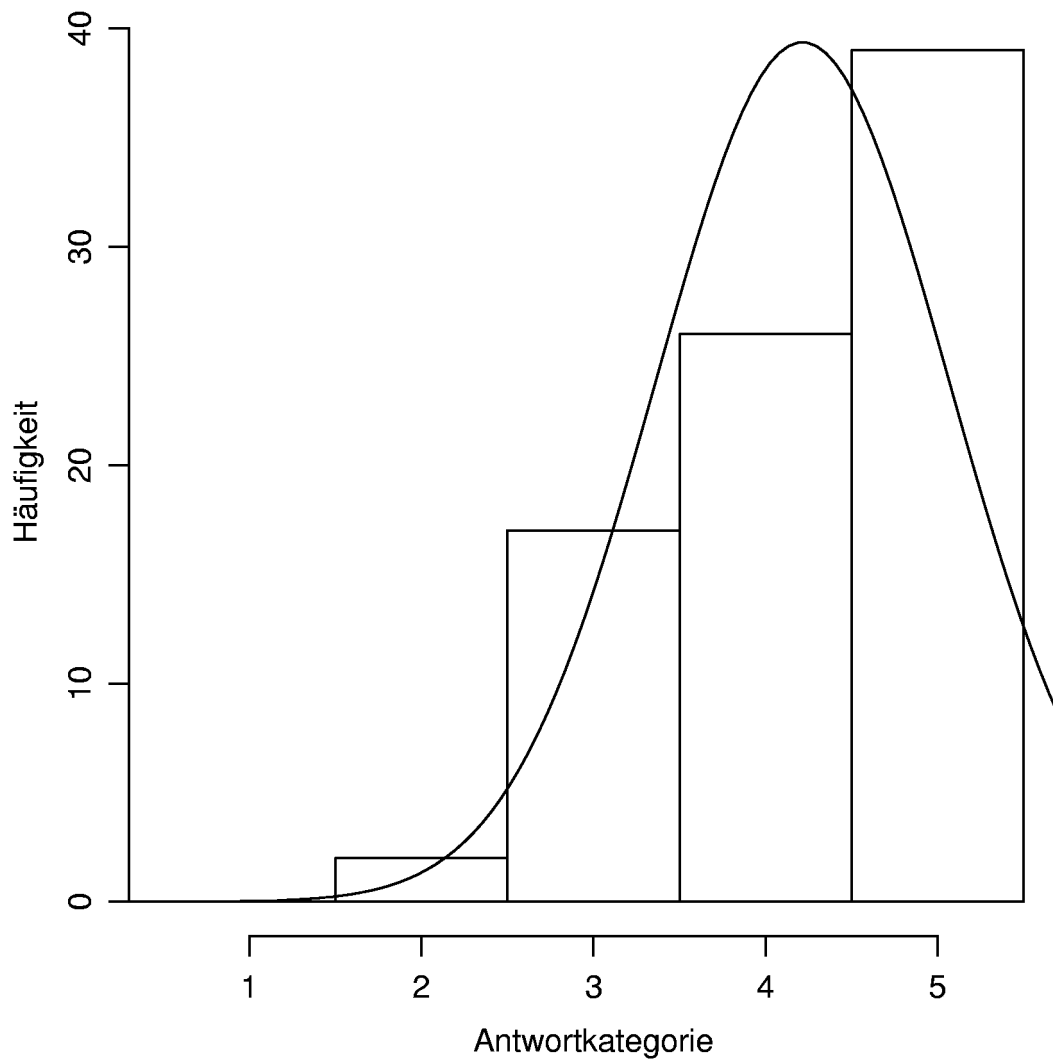


Abb. 84: Verteilung der Antworten für Item L09 zum ersten Messzeitpunkt

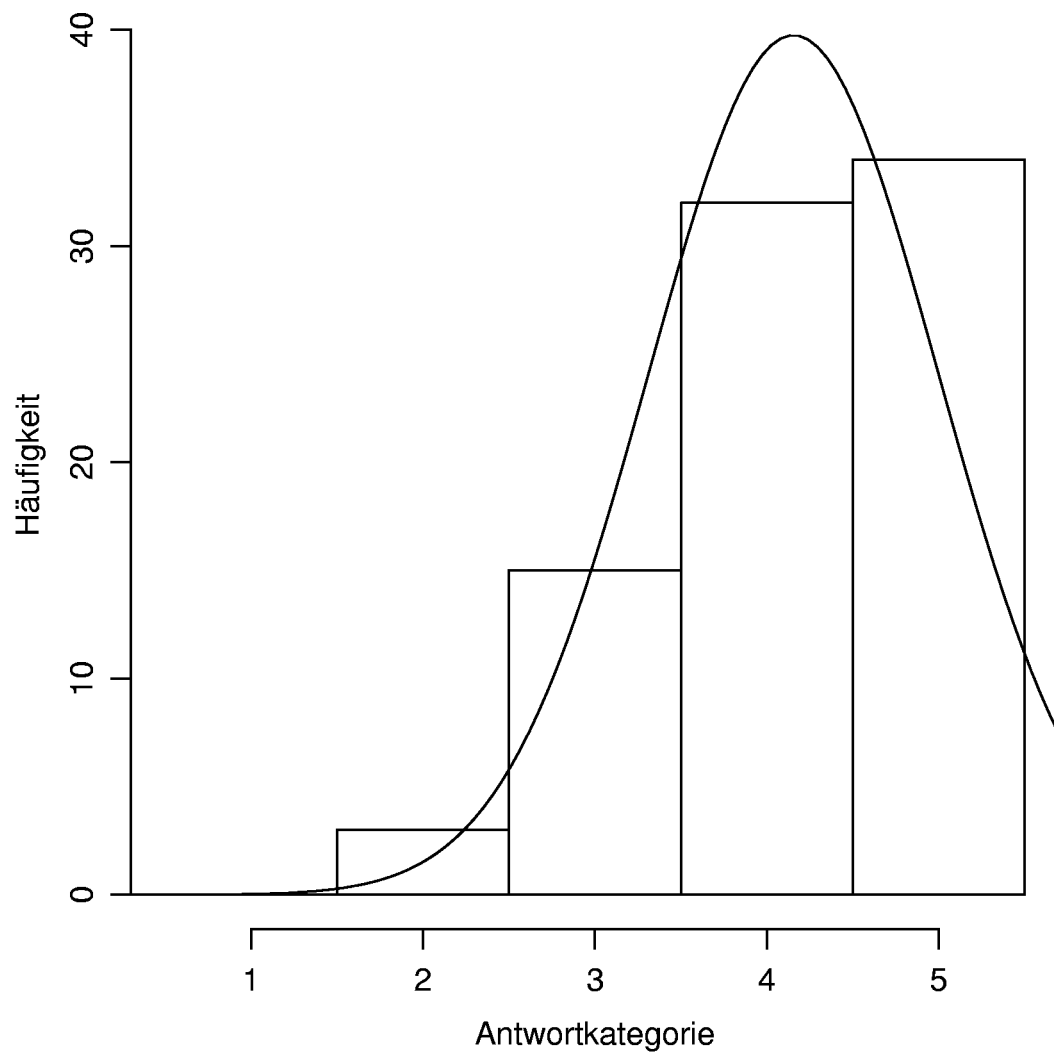


Abb. 85: Verteilung der Antworten für Item L09 zum zweiten Messzeitpunkt

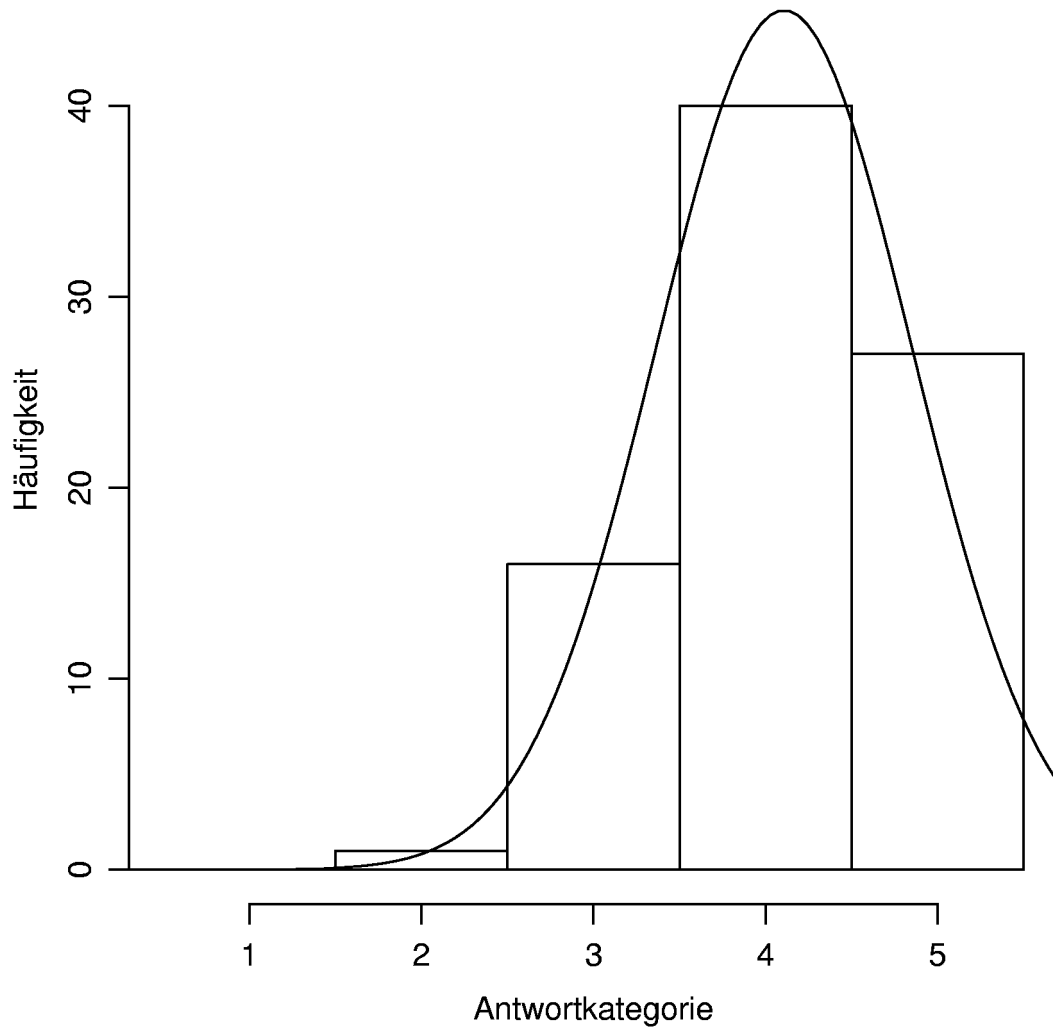


Abb. 86: Verteilung der Antworten für Item L11 zum ersten Messzeitpunkt

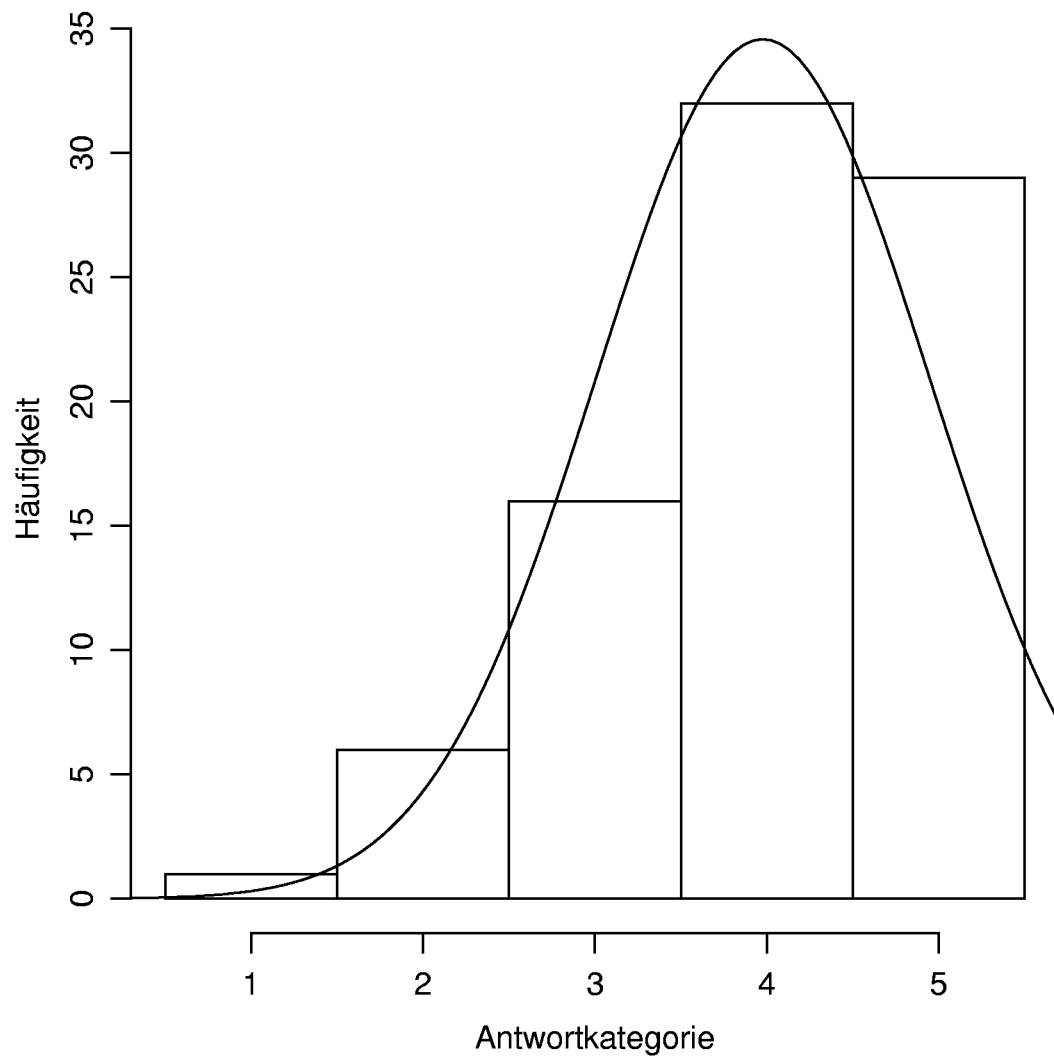


Abb. 87: Verteilung der Antworten für Item L11 zum zweiten Messzeitpunkt

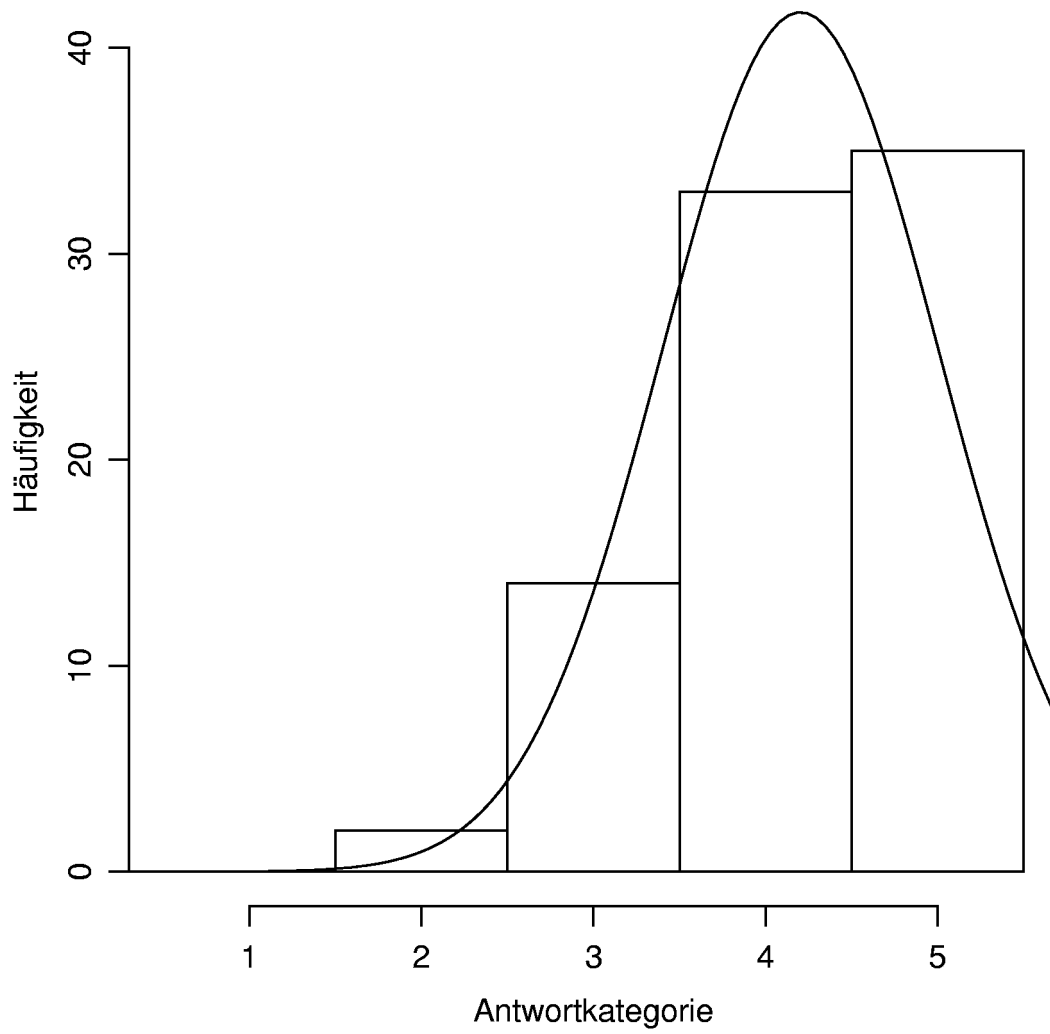


Abb. 88: Verteilung der Antworten für Item L12 zum ersten Messzeitpunkt

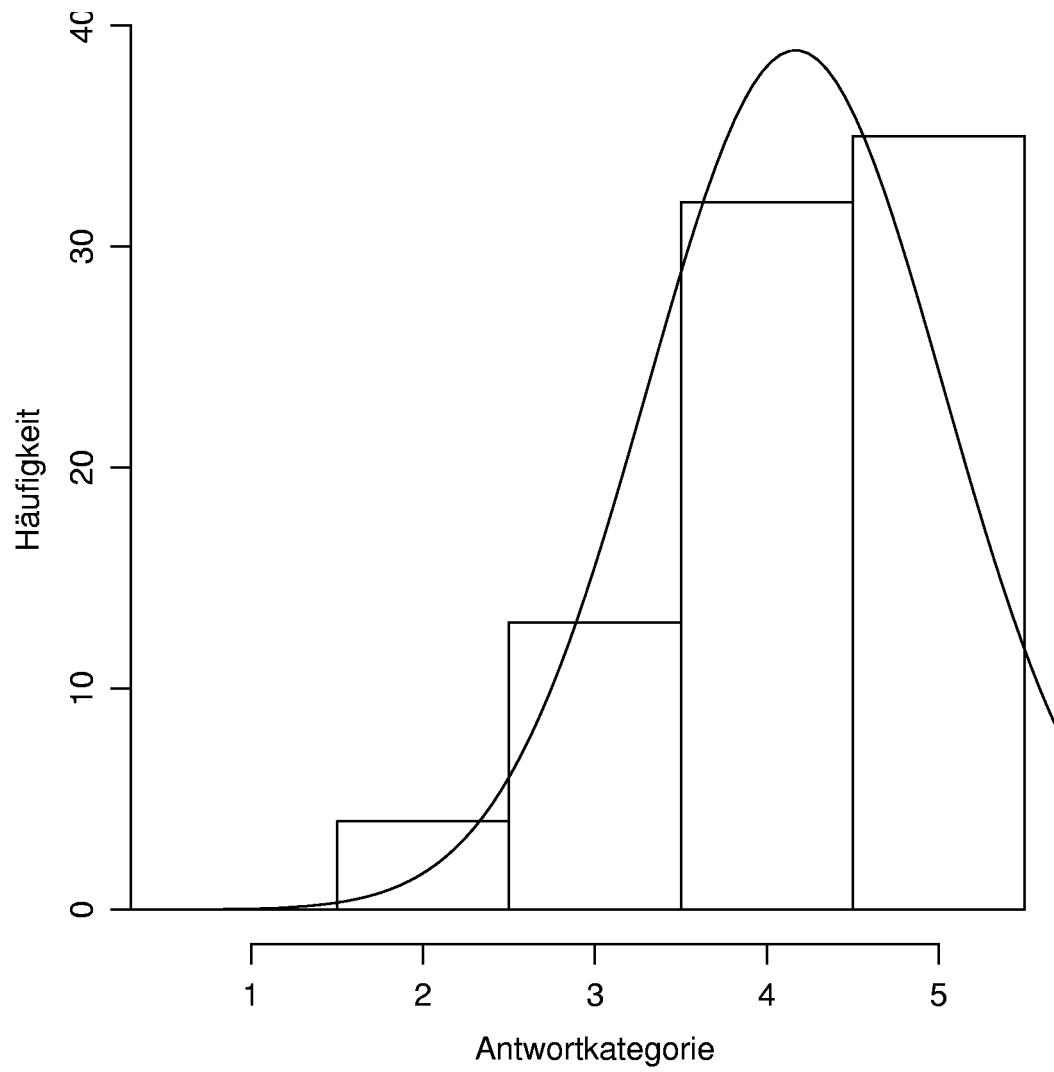


Abb. 89: Verteilung der Antworten für Item L12 zum zweiten Messzeitpunkt

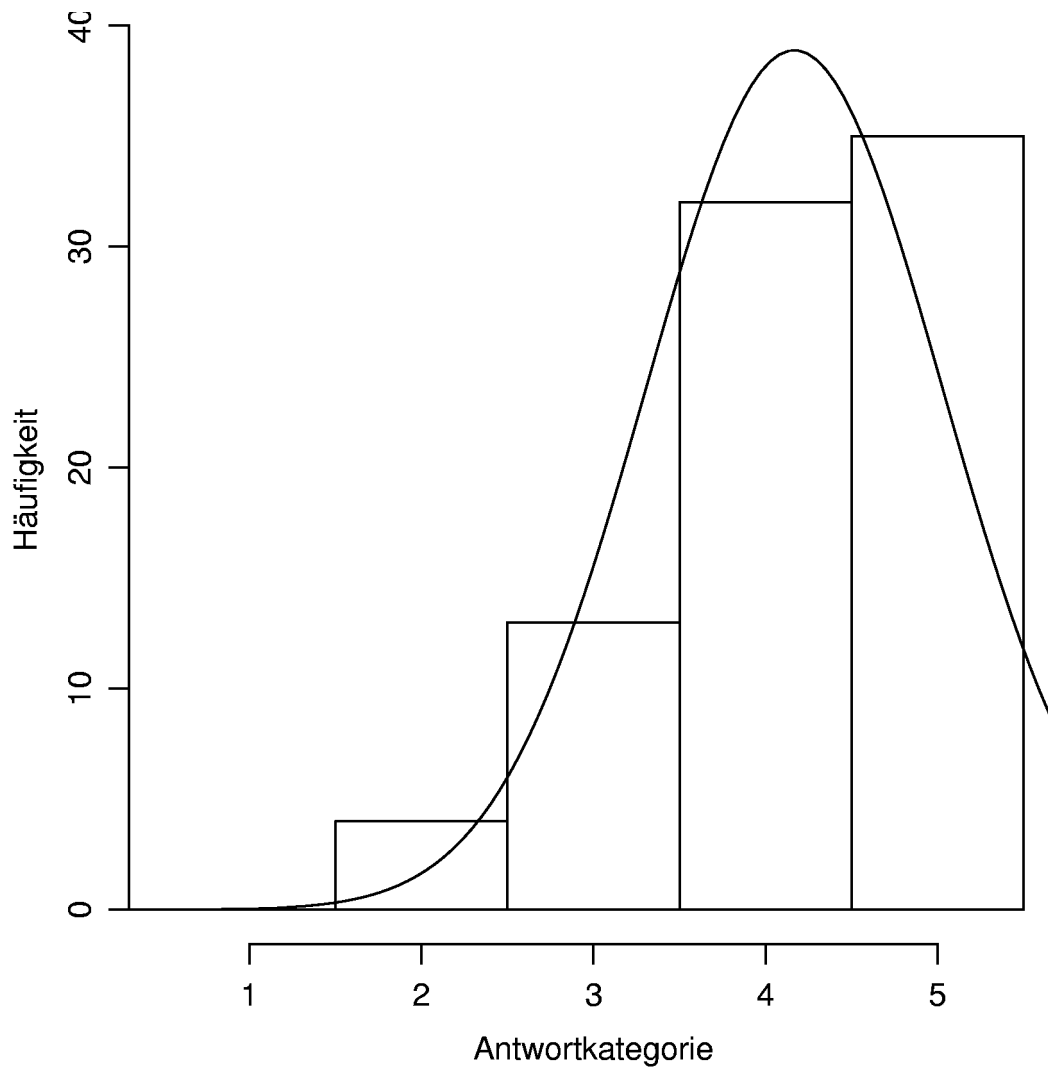


Abb. 90: Verteilung der Antworten für Item L13 zum ersten Messzeitpunkt

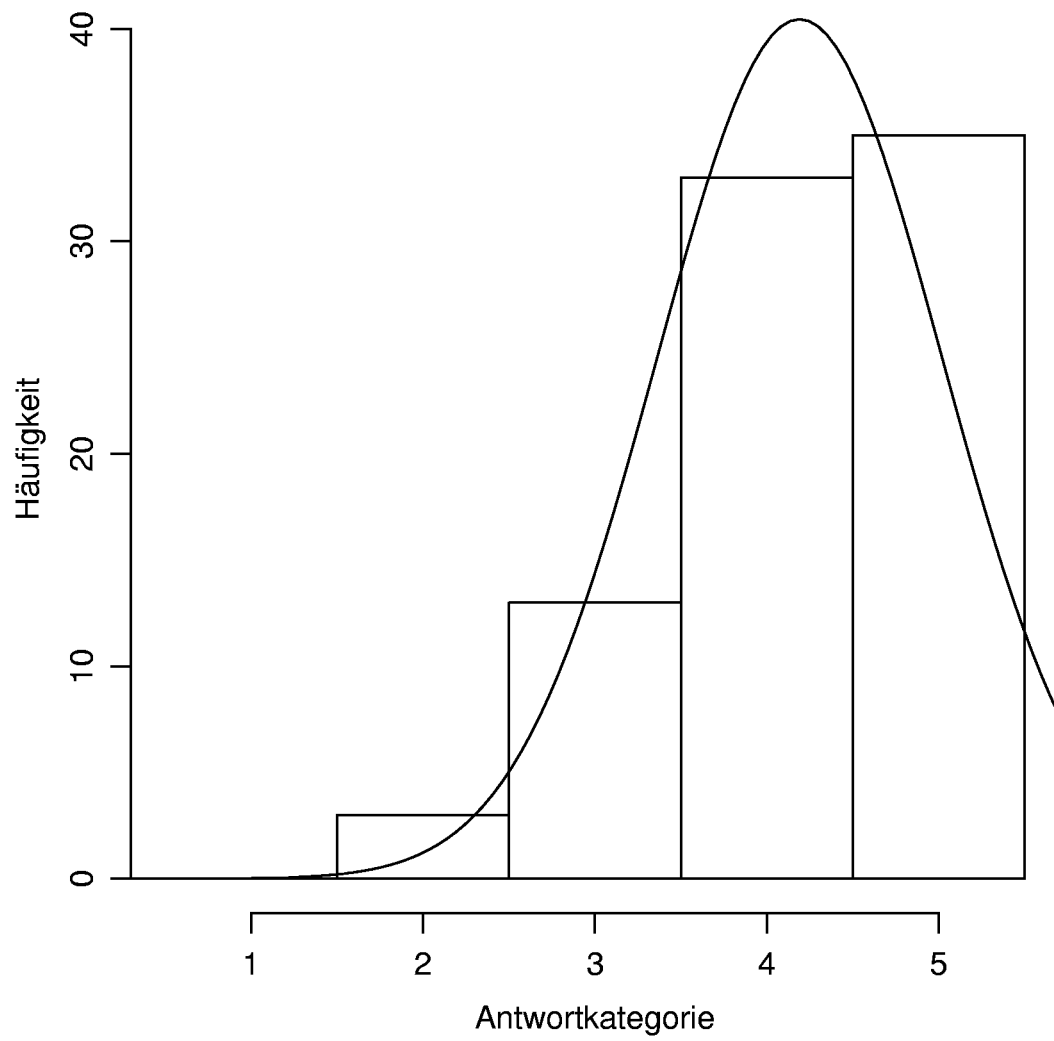


Abb. 91: Verteilung der Antworten für Item L13 zum zweiten Messzeitpunkt

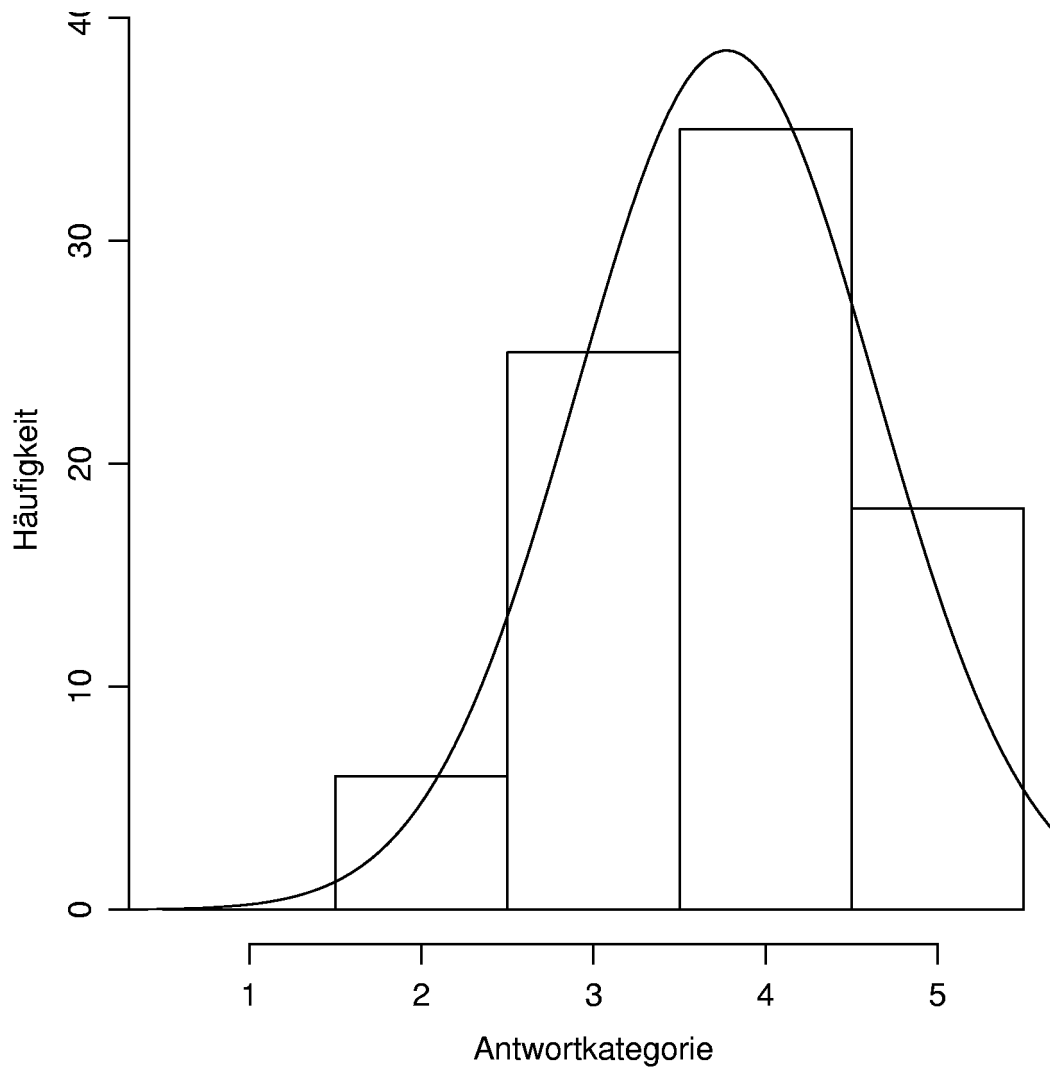


Abb. 92: Verteilung der Antworten für Item E02 zum ersten Messzeitpunkt

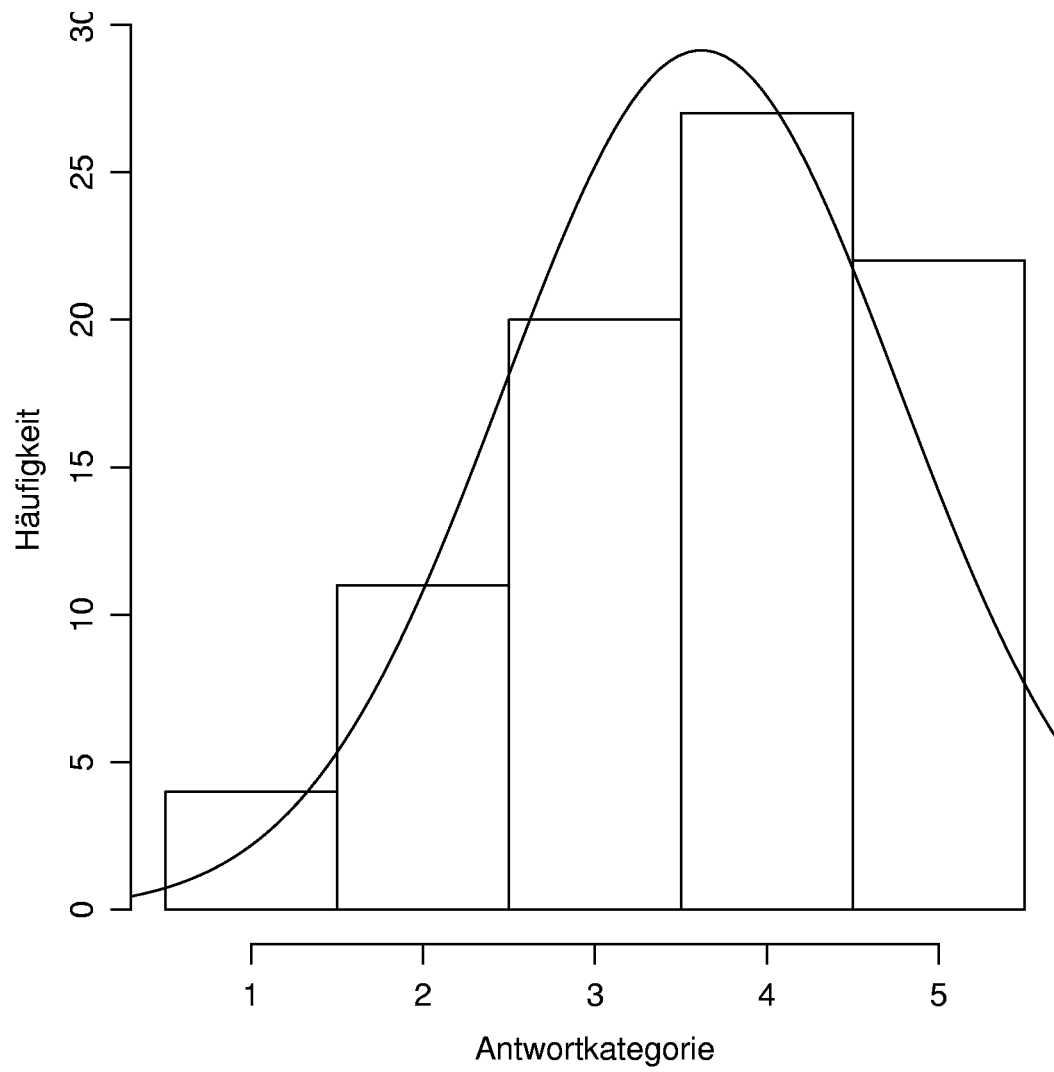


Abb. 93: Verteilung der Antworten für Item E02 zum zweiten Messzeitpunkt

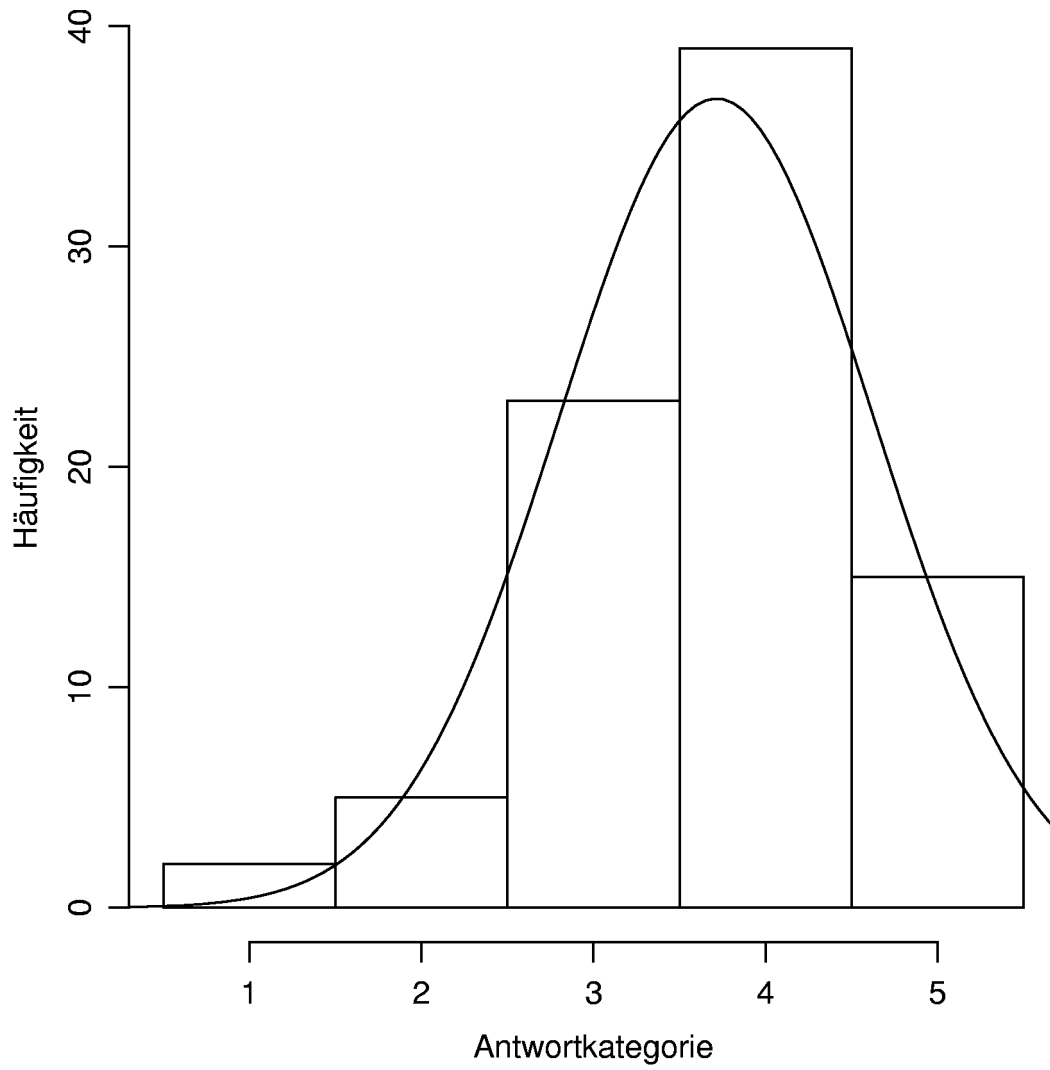


Abb. 94: Verteilung der Antworten für Item E07 zum ersten Messzeitpunkt

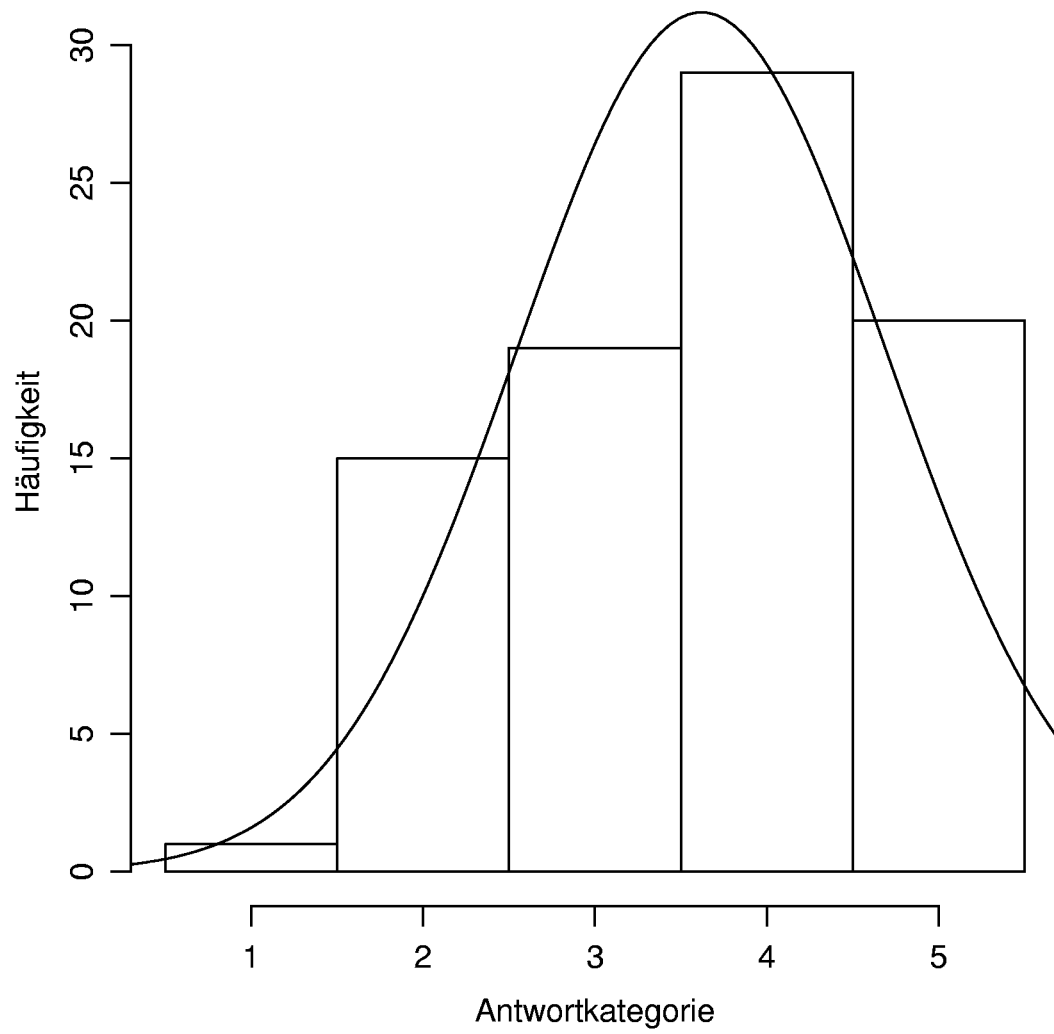


Abb. 95: Verteilung der Antworten für Item E07 zum zweiten Messzeitpunkt

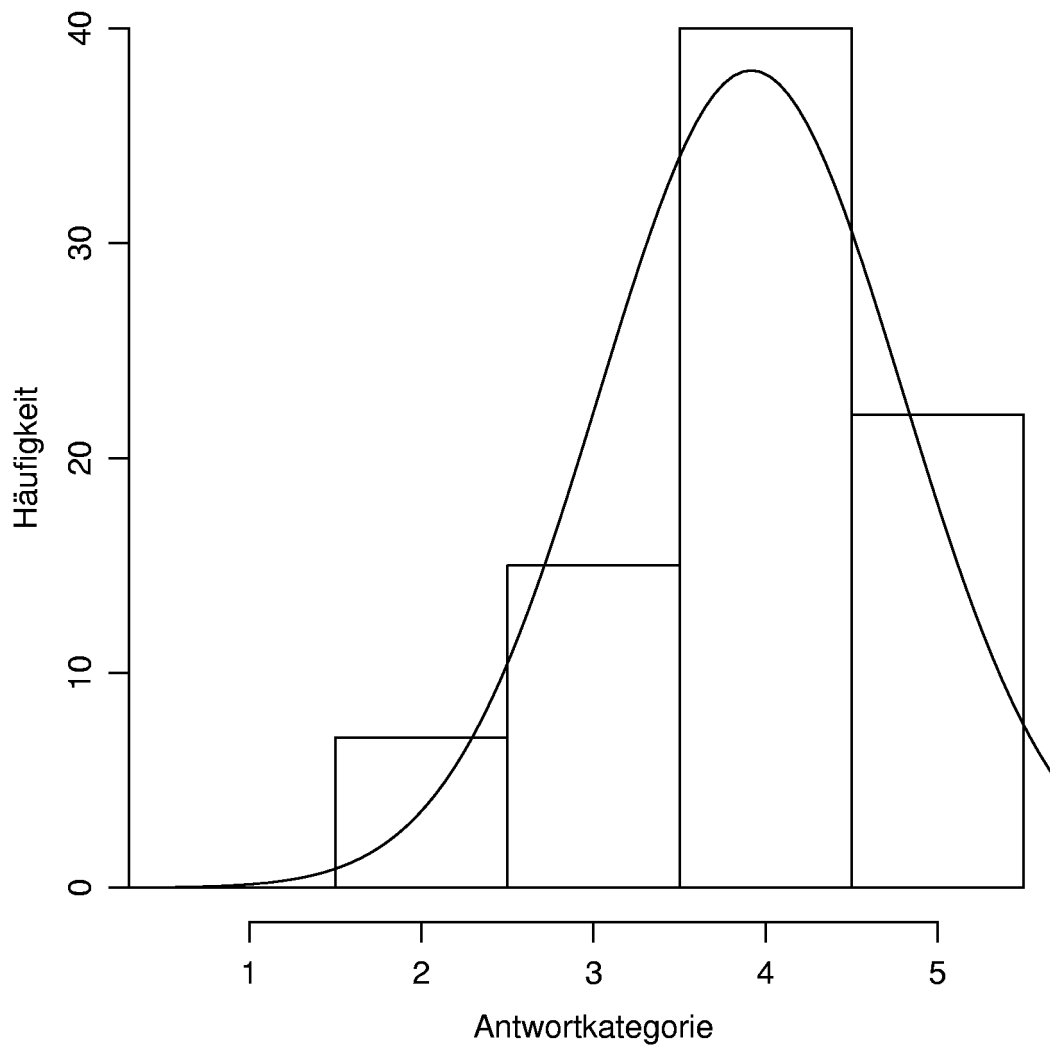


Abb. 96: Verteilung der Antworten für Item E11 zum ersten Messzeitpunkt

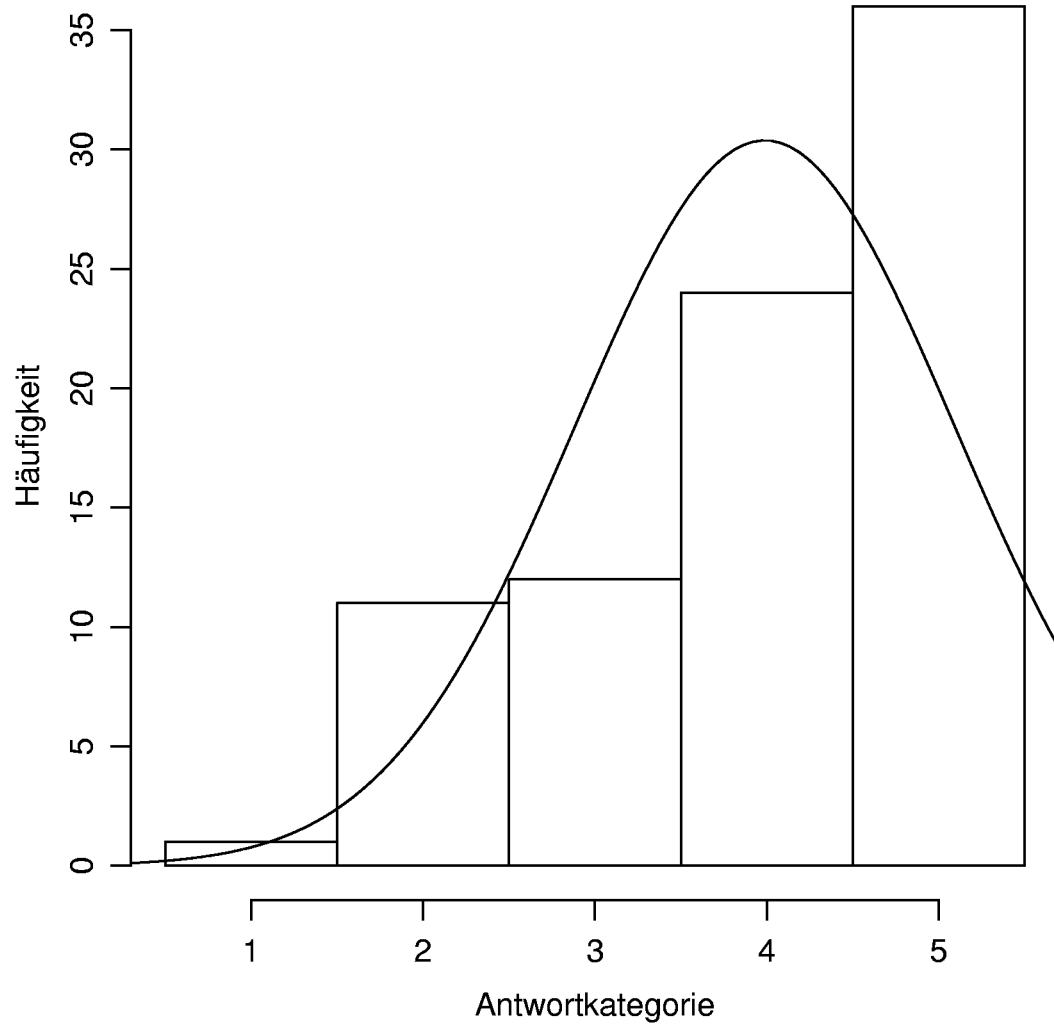


Abb. 97: Verteilung der Antworten für Item E11 zum zweiten Messzeitpunkt

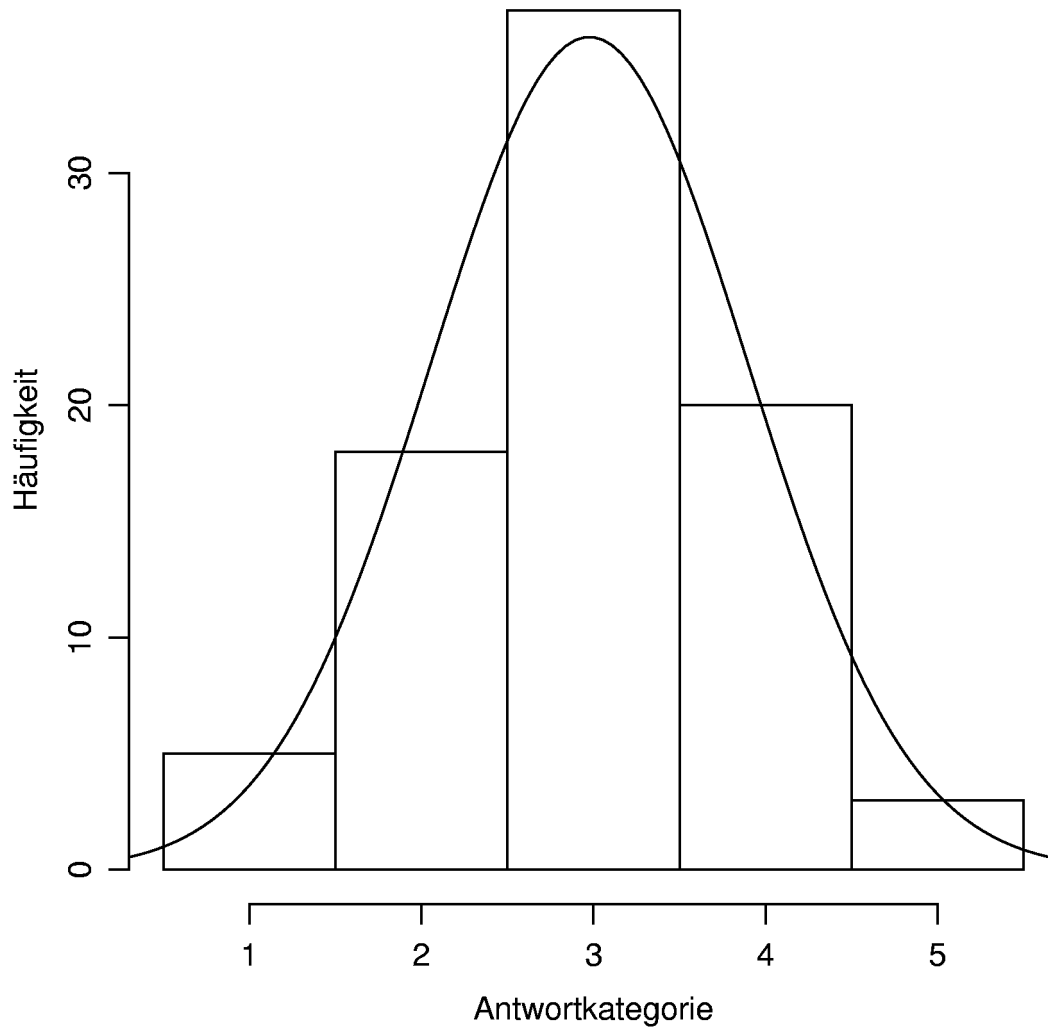


Abb. 98: Verteilung der Antworten für Item E12 zum ersten Messzeitpunkt

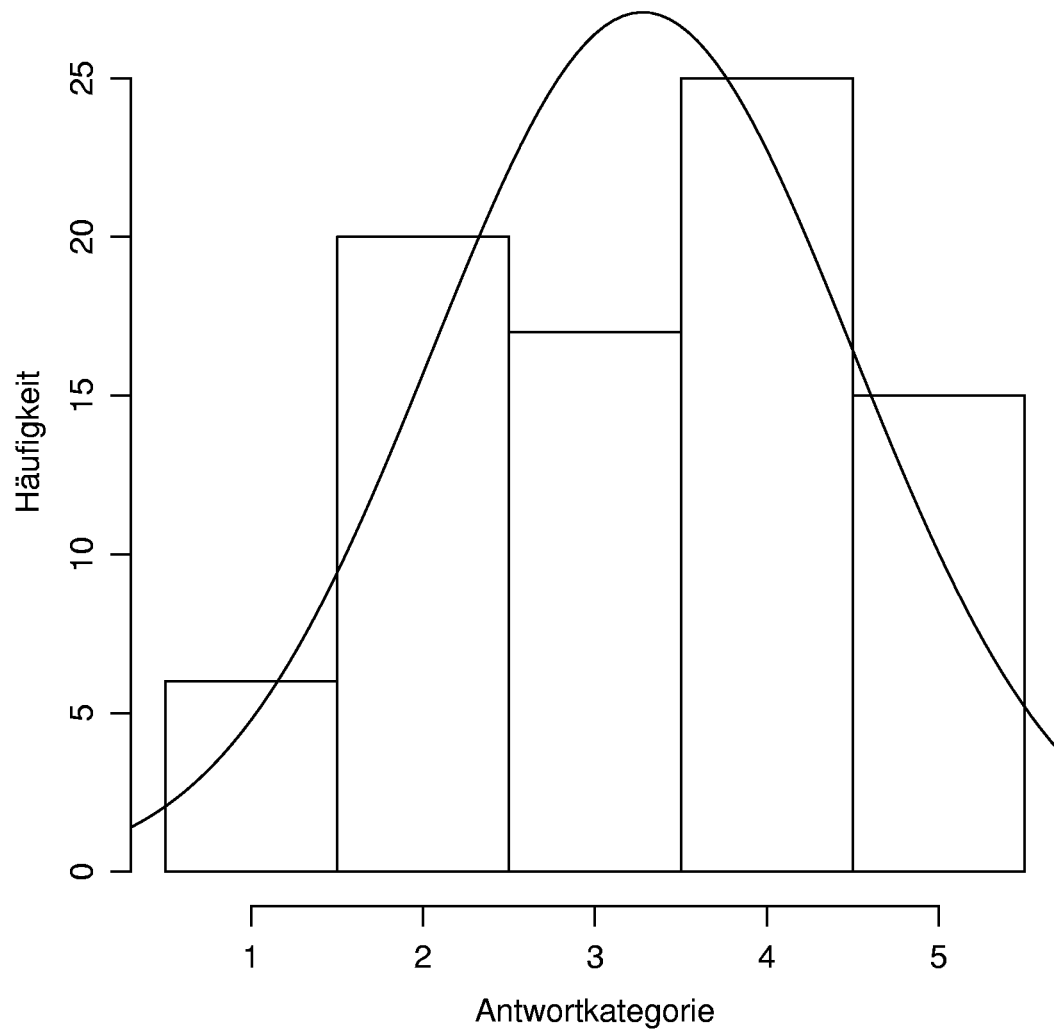


Abb. 99: Verteilung der Antworten für Item E12 zum zweiten Messzeitpunkt

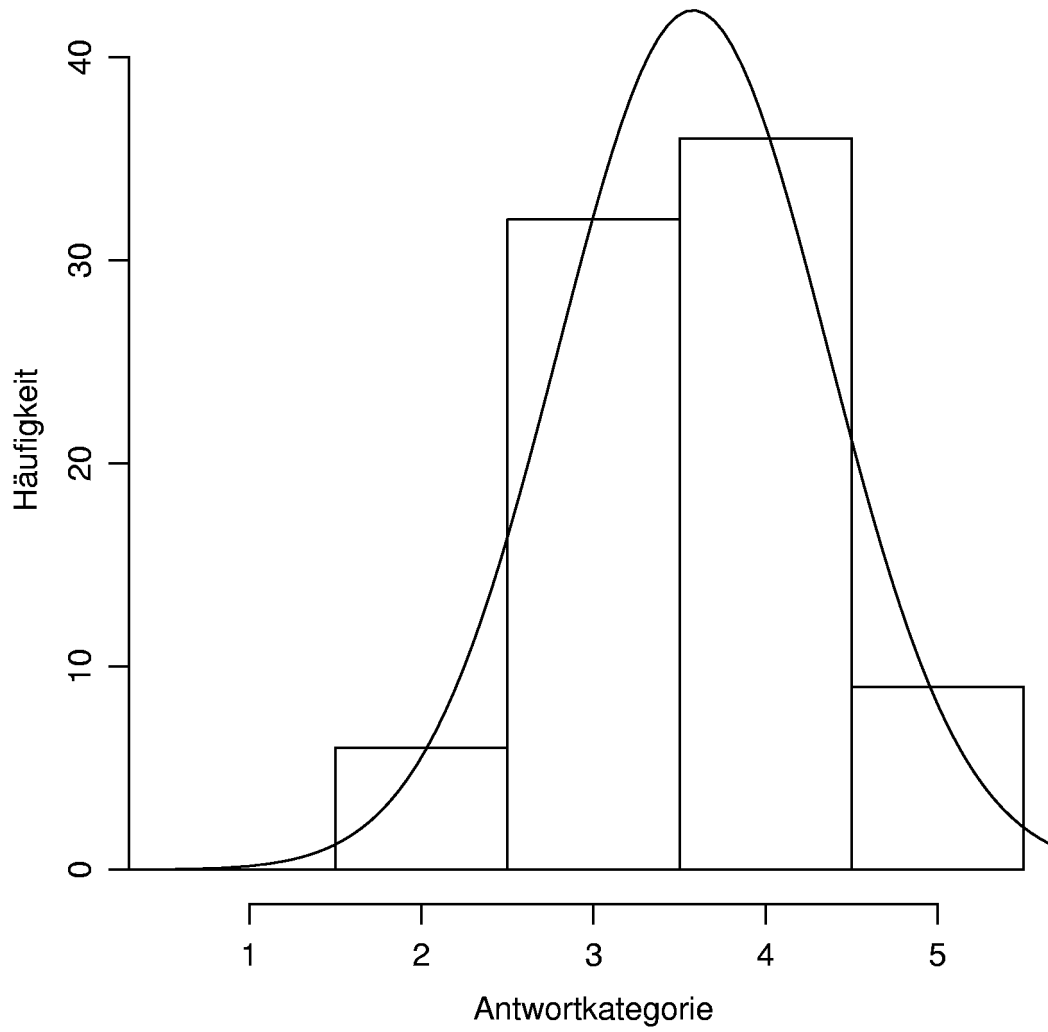


Abb. 100: Verteilung der Antworten für Item E13 zum ersten Messzeitpunkt

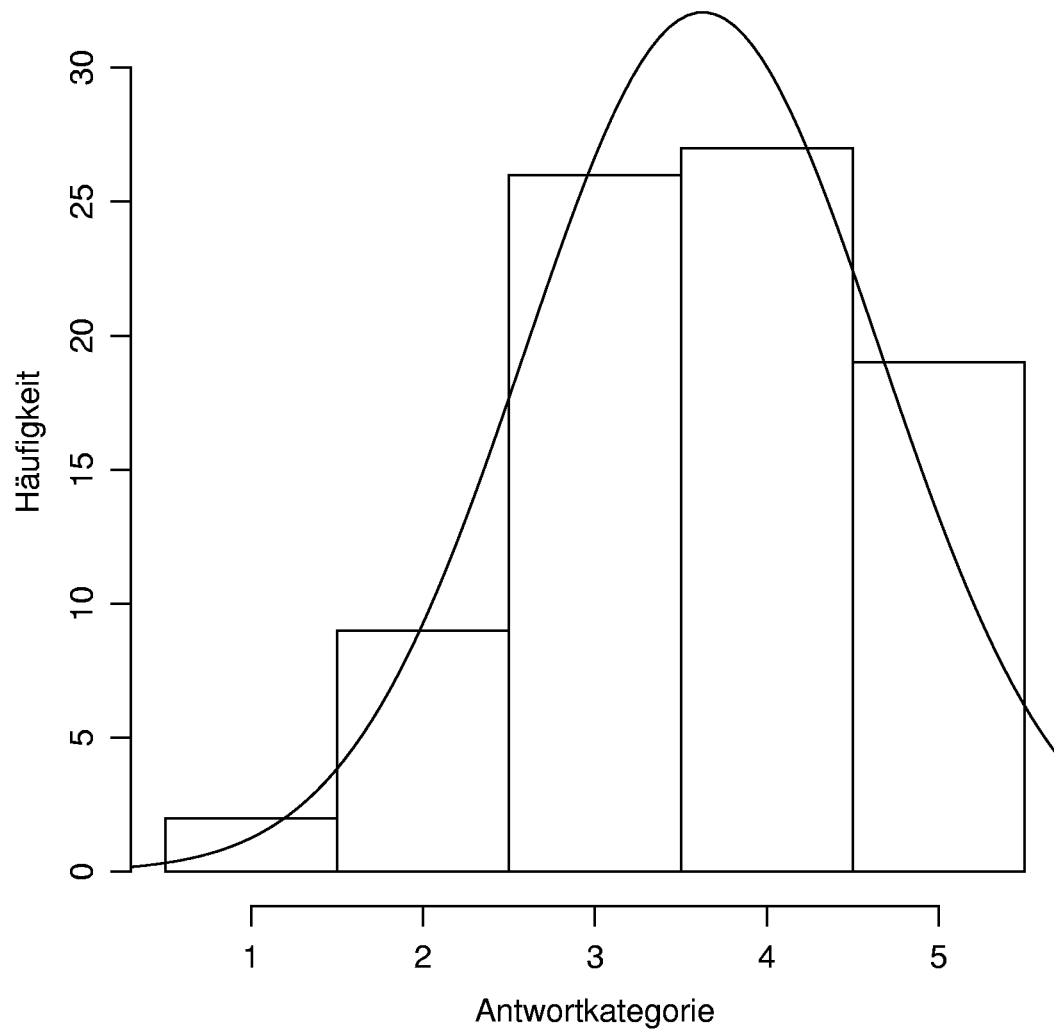


Abb. 101: Verteilung der Antworten für Item E13 zum zweiten Messzeitpunkt

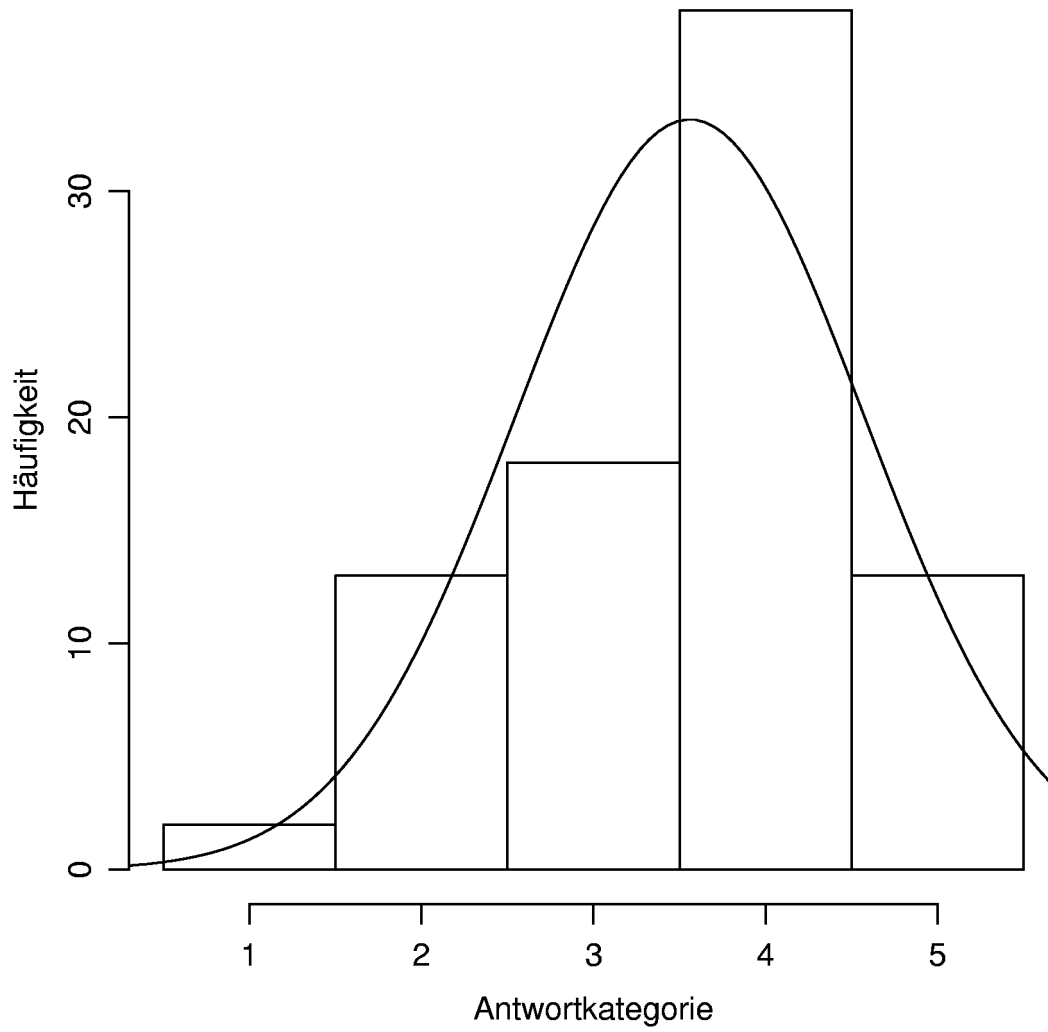


Abb. 102: Verteilung der Antworten für Item E14 zum ersten Messzeitpunkt

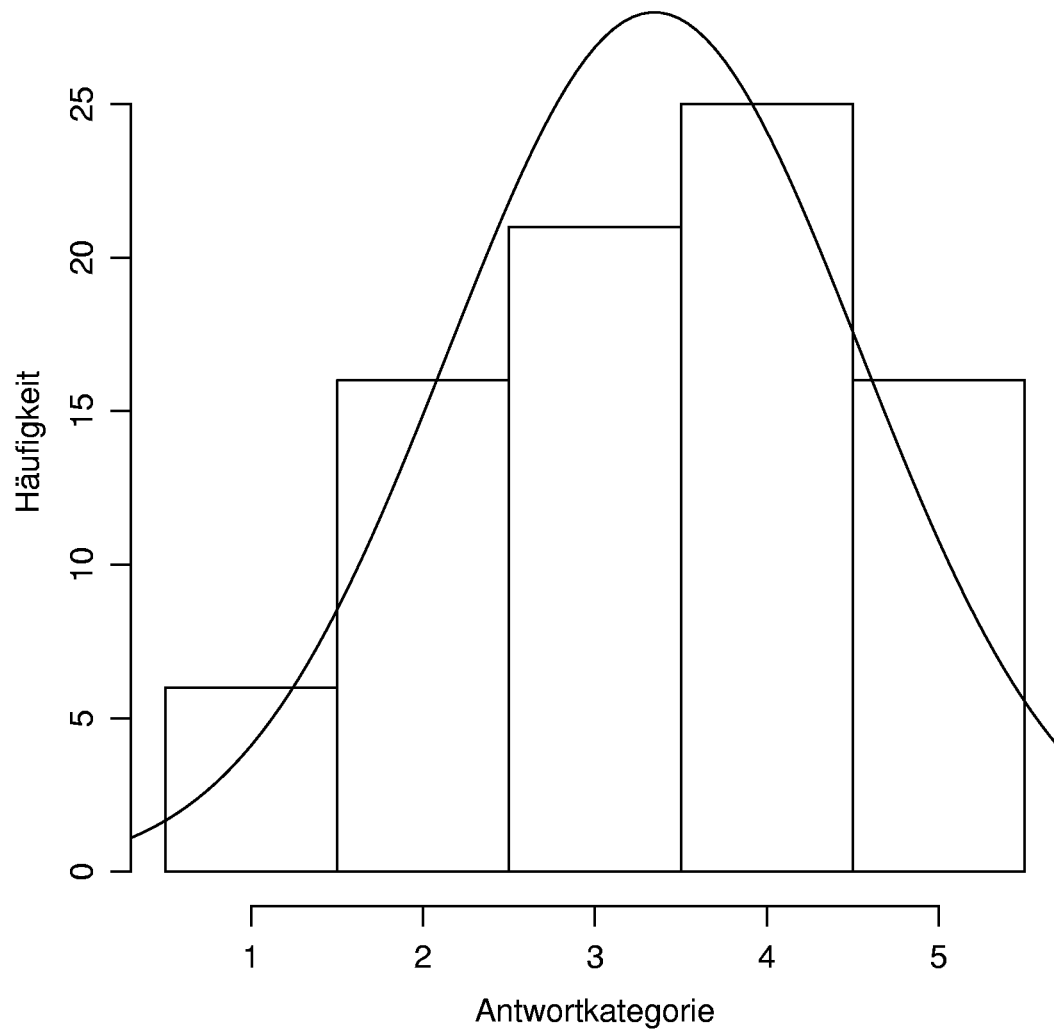


Abb. 103: Verteilung der Antworten für Item E14 zum zweiten Messzeitpunkt

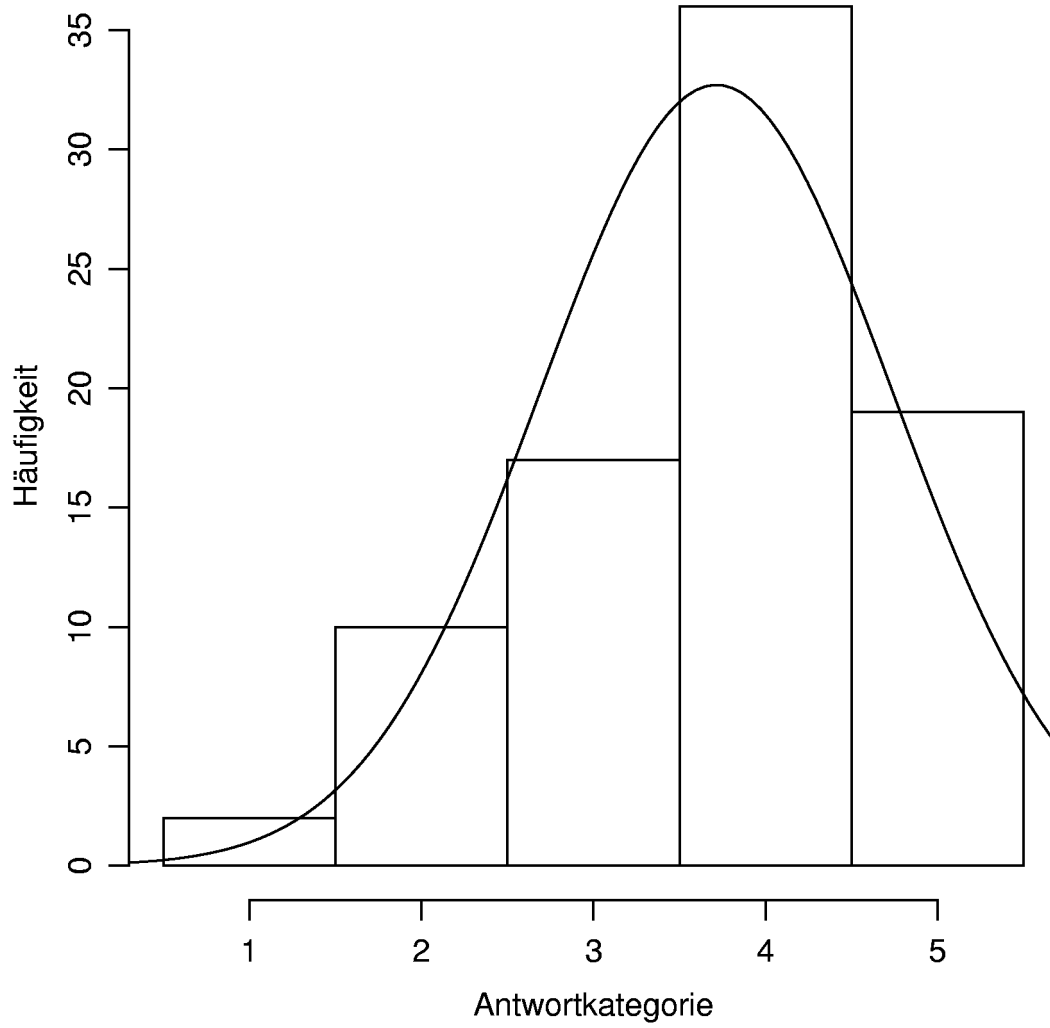


Abb. 104: Verteilung der Antworten für Item E15 zum ersten Messzeitpunkt

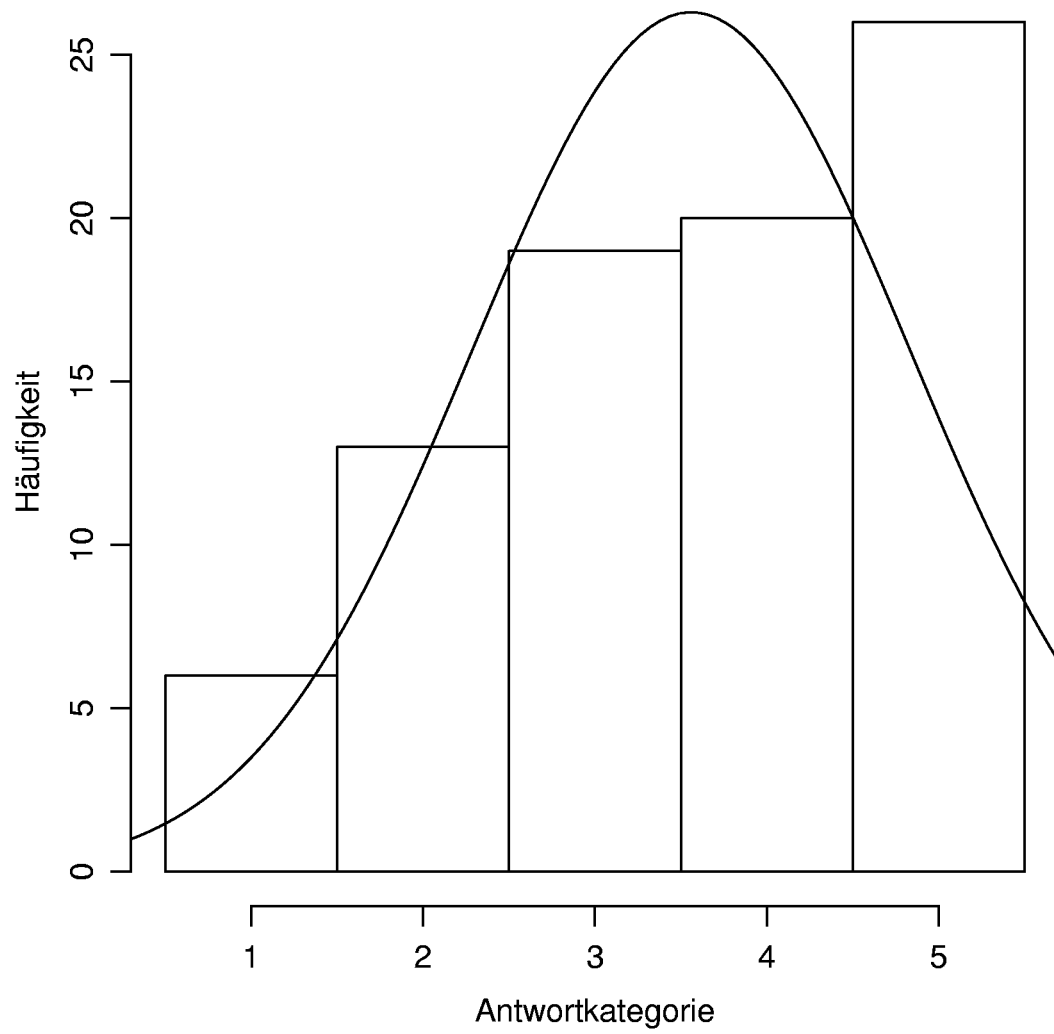


Abb. 105: Verteilung der Antworten für Item E15 zum zweiten Messzeitpunkt

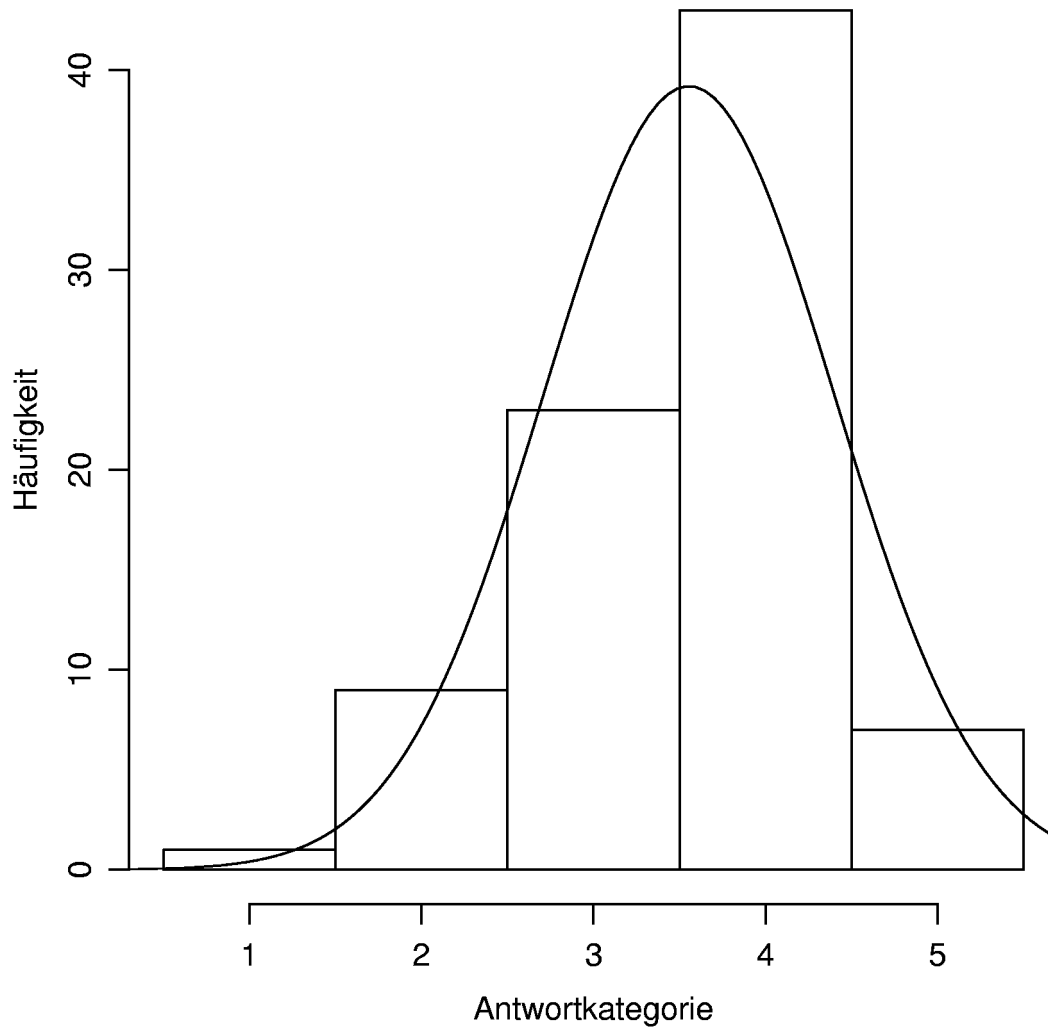


Abb. 106: Verteilung der Antworten für Item E17 zum ersten Messzeitpunkt

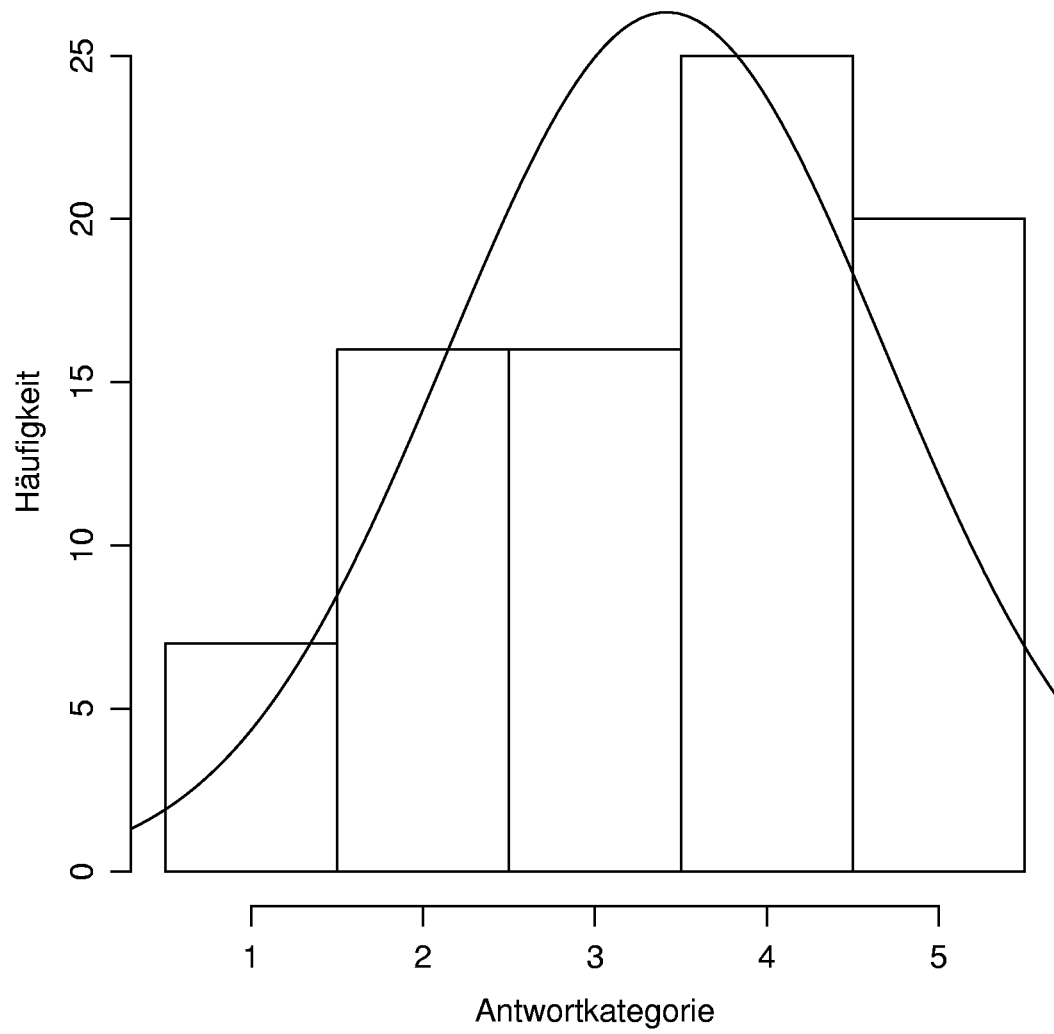


Abb. 107: Verteilung der Antworten für Item E17 zum zweiten Messzeitpunkt

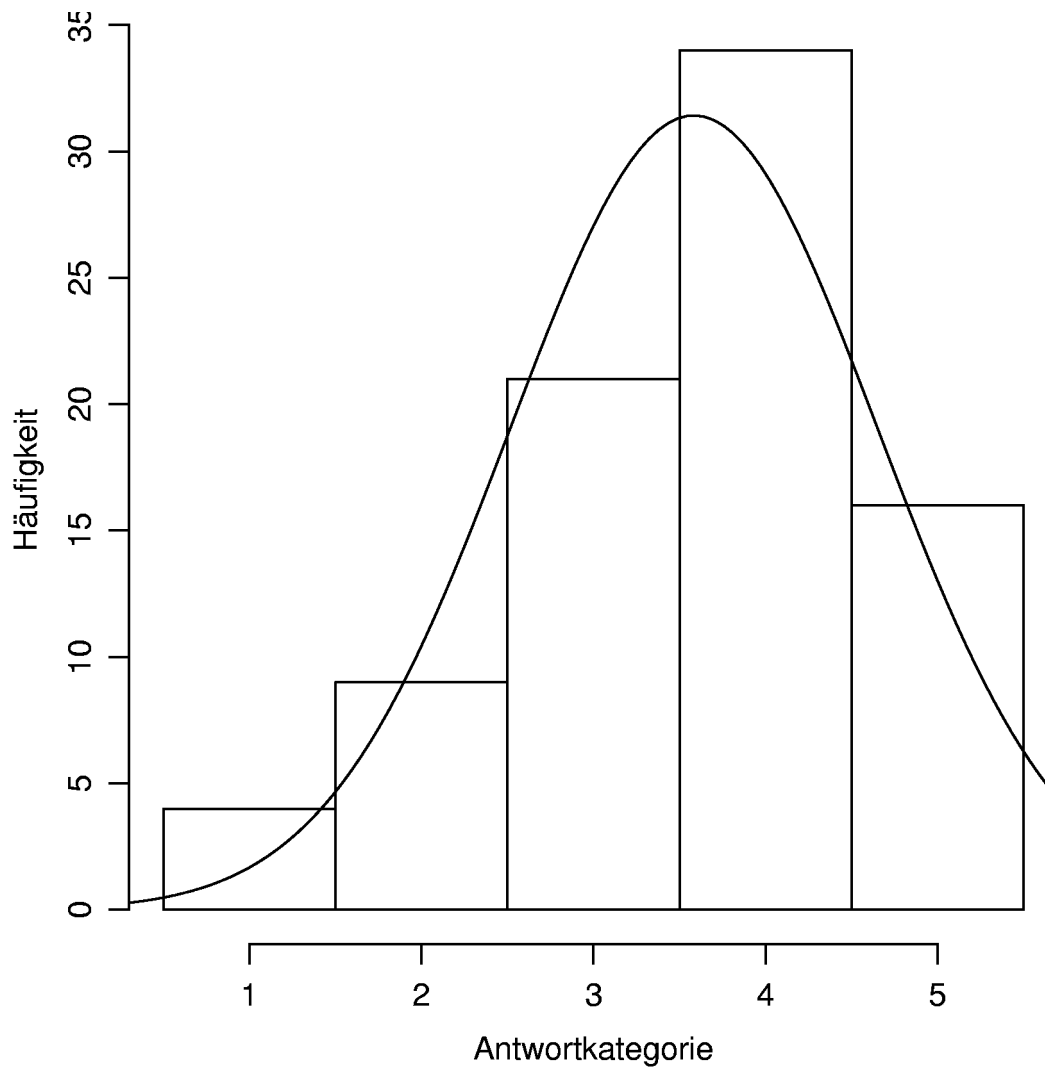


Abb. 108: Verteilung der Antworten für Item U01 zum ersten Messzeitpunkt

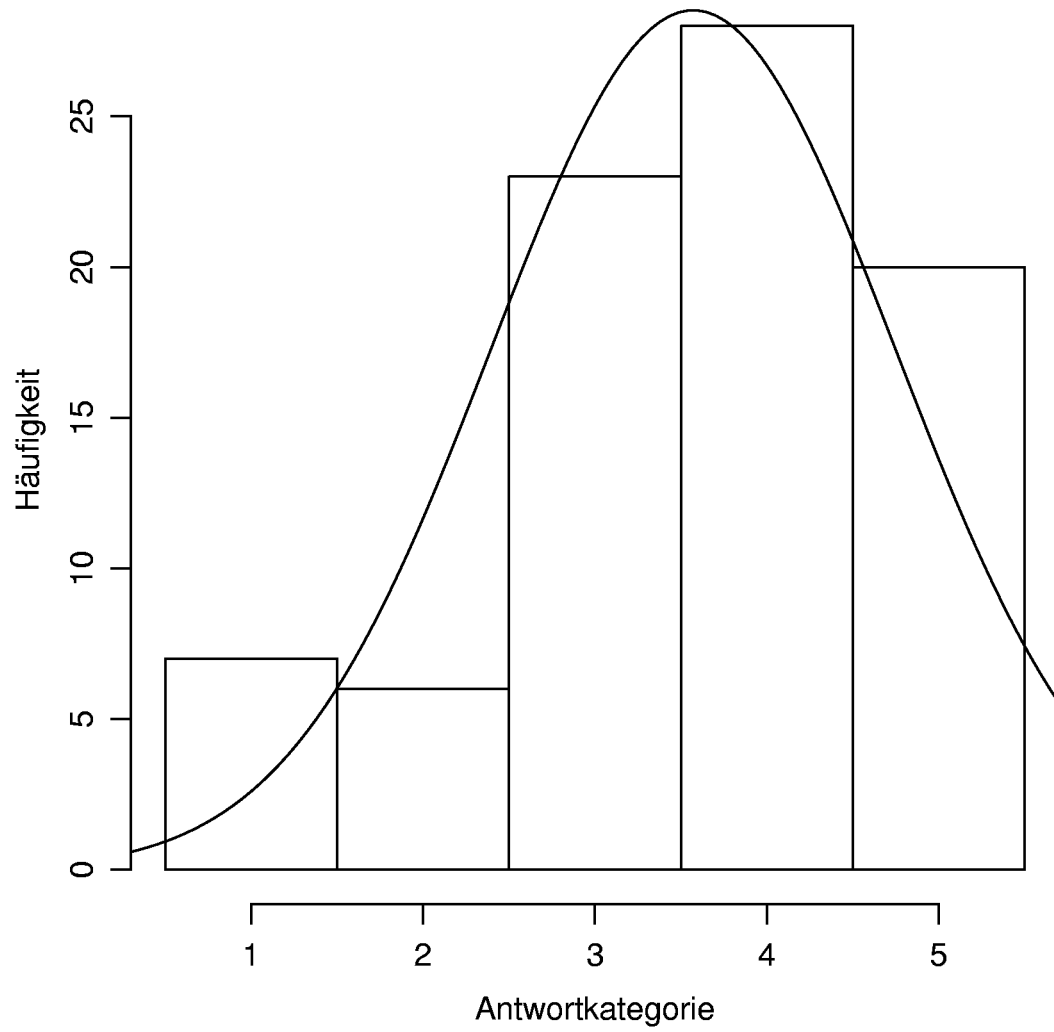


Abb. 109: Verteilung der Antworten für Item U01 zum zweiten Messzeitpunkt

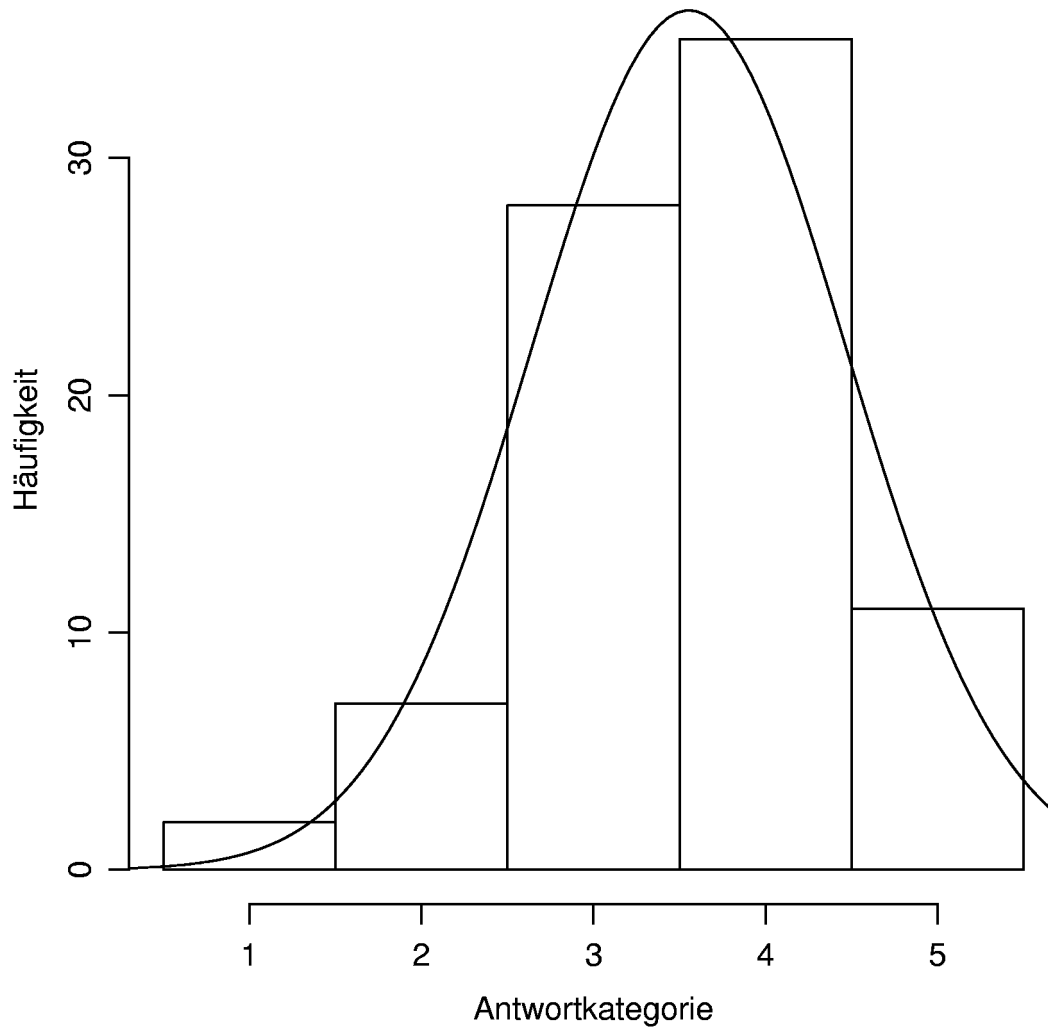


Abb. 110: Verteilung der Antworten für Item U02 zum ersten Messzeitpunkt

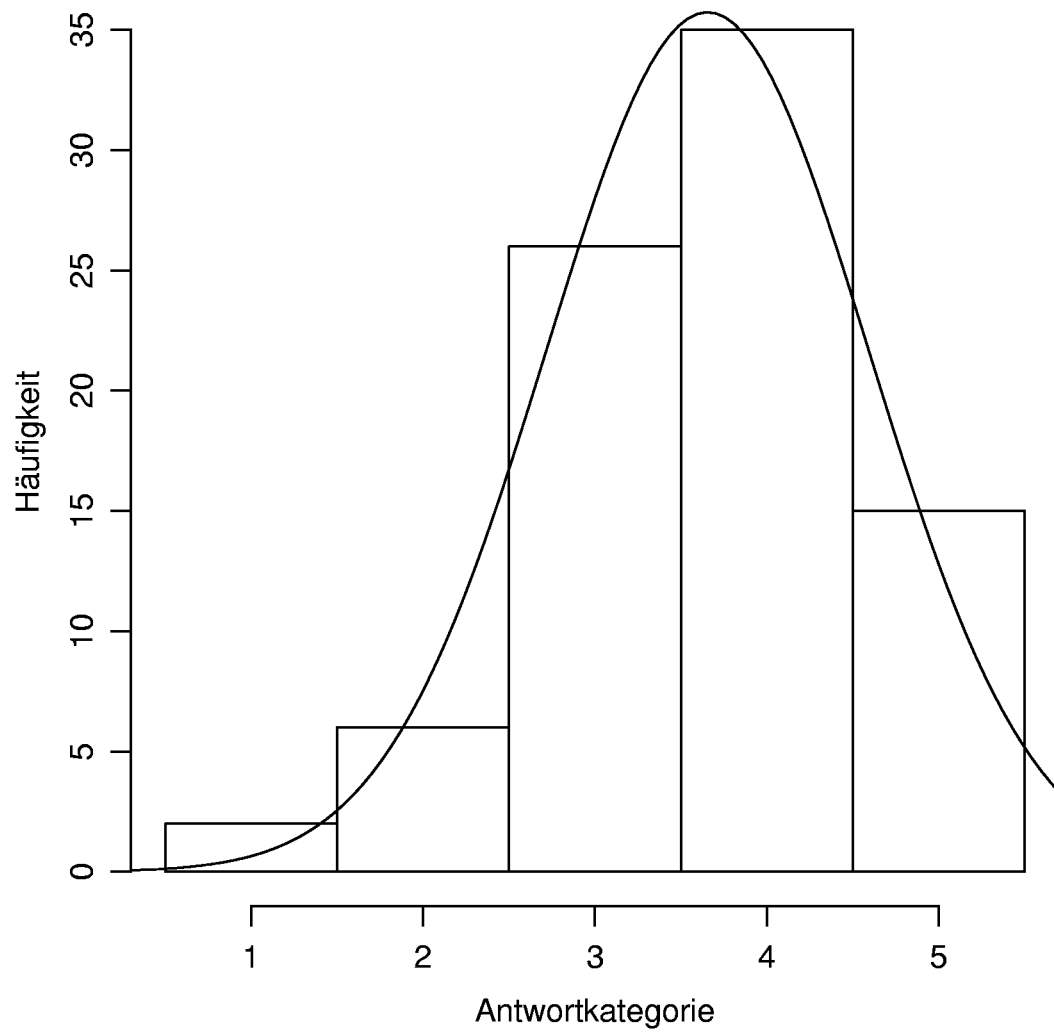


Abb. 111: Verteilung der Antworten für Item U02 zum zweiten Messzeitpunkt

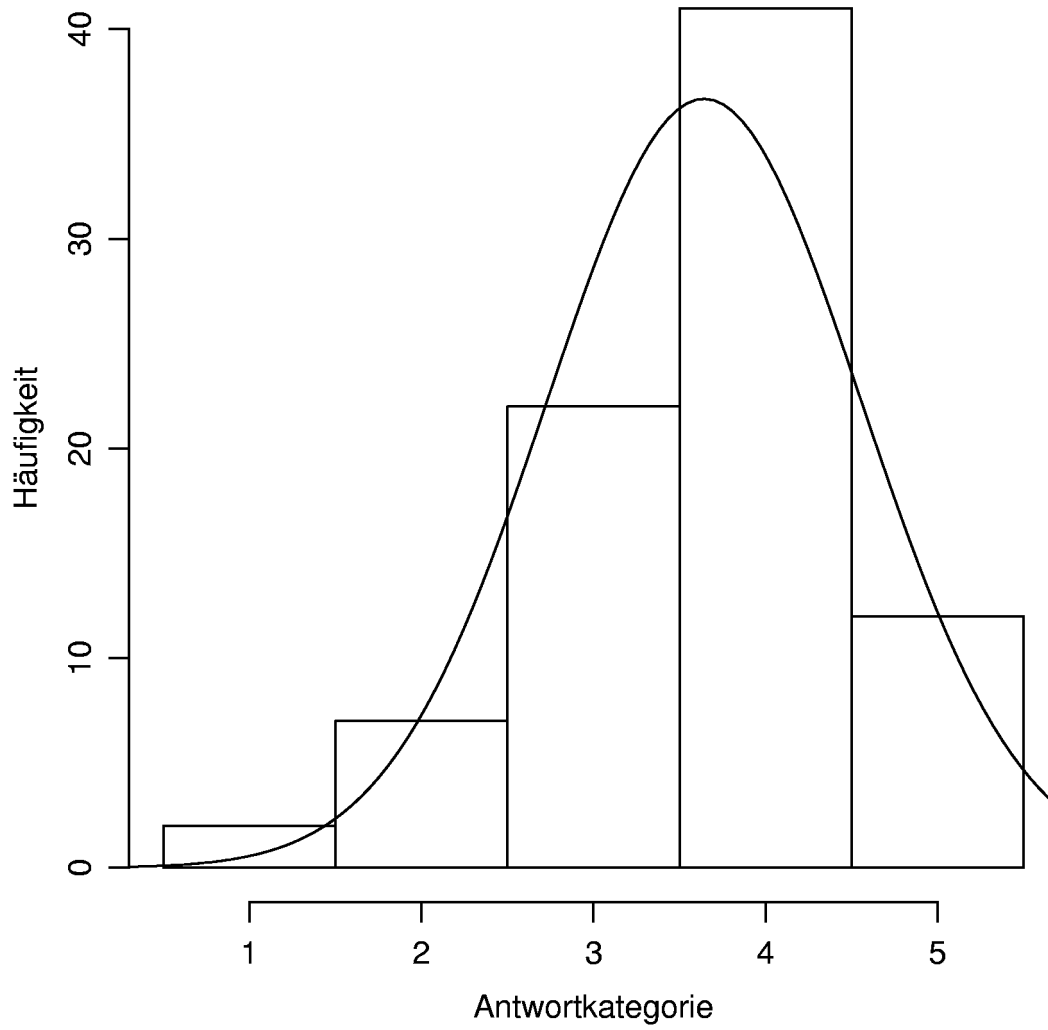


Abb. 112: Verteilung der Antworten für Item U03 zum ersten Messzeitpunkt

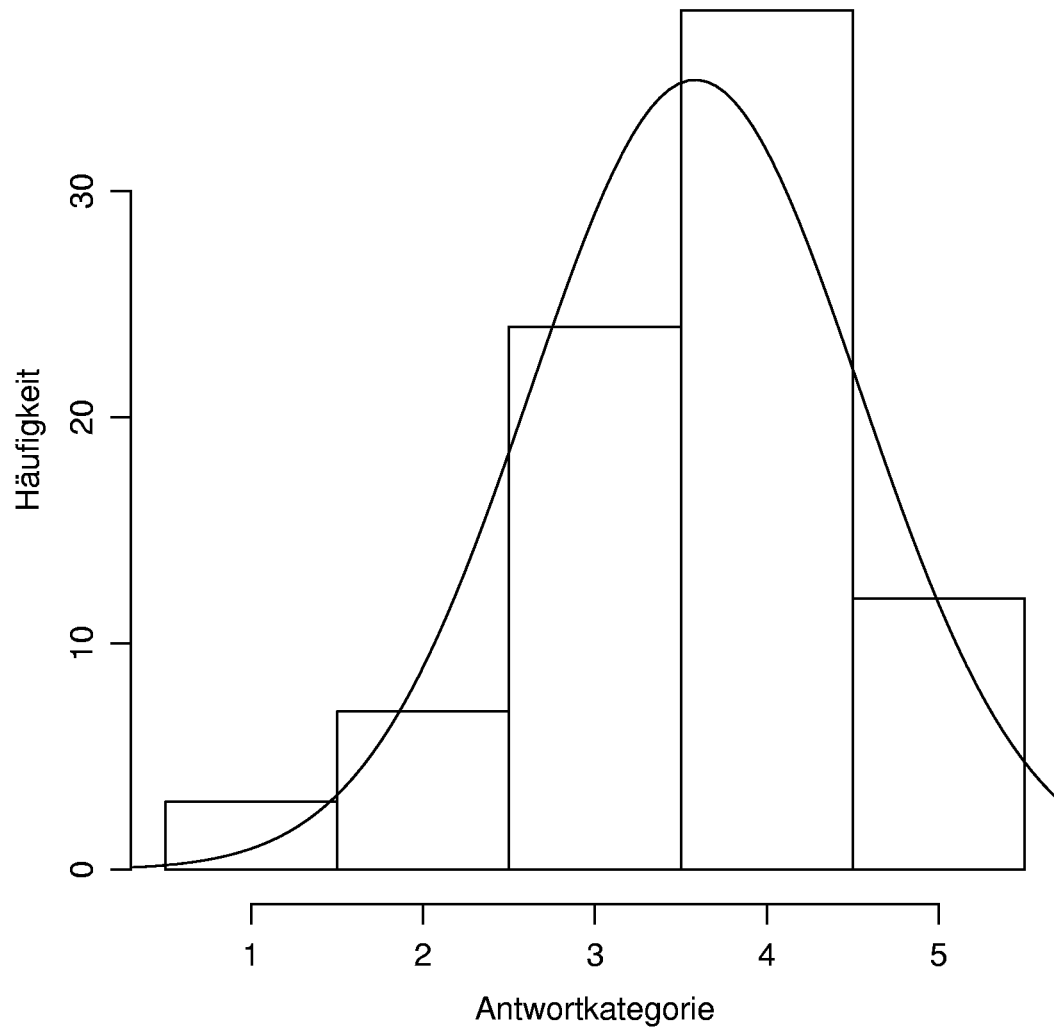


Abb. 113: Verteilung der Antworten für Item U03 zum zweiten Messzeitpunkt

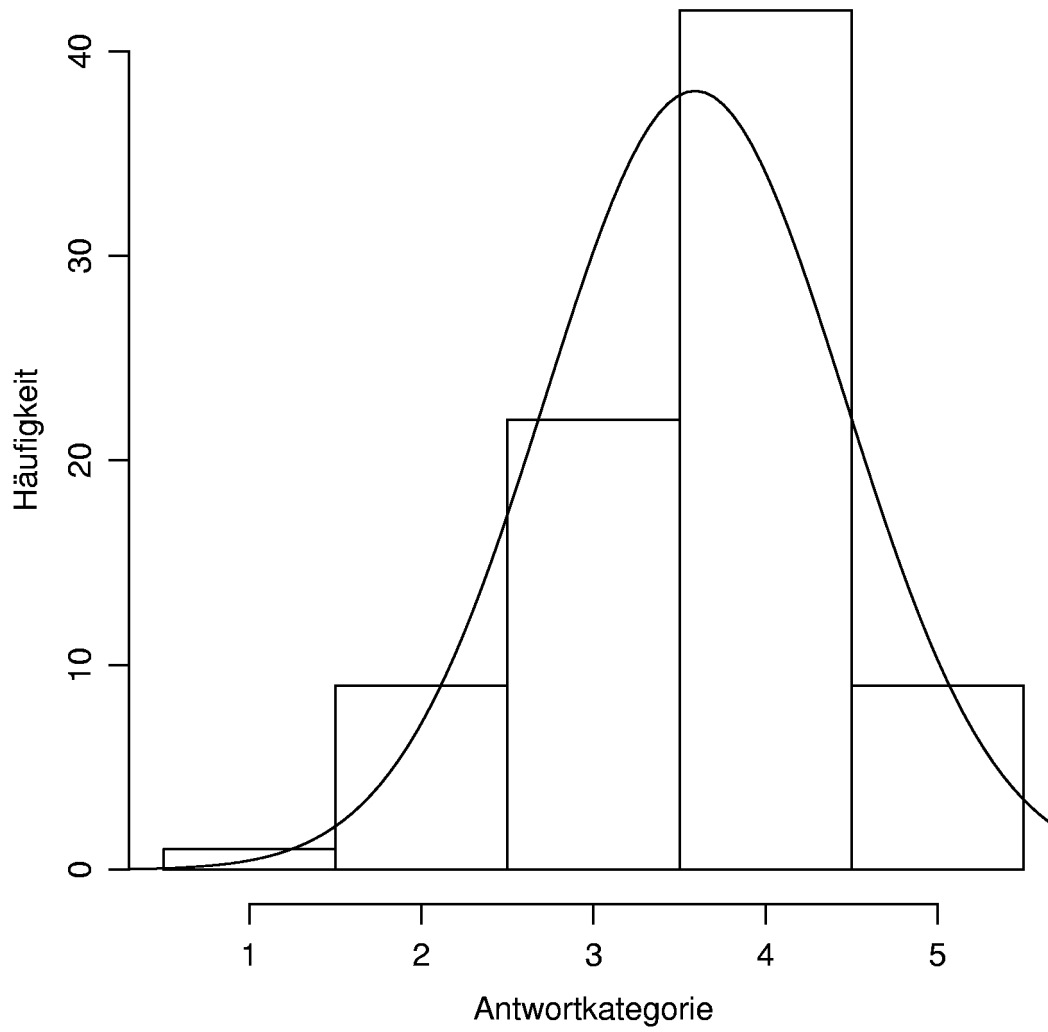


Abb. 114: Verteilung der Antworten für Item U04 zum ersten Messzeitpunkt

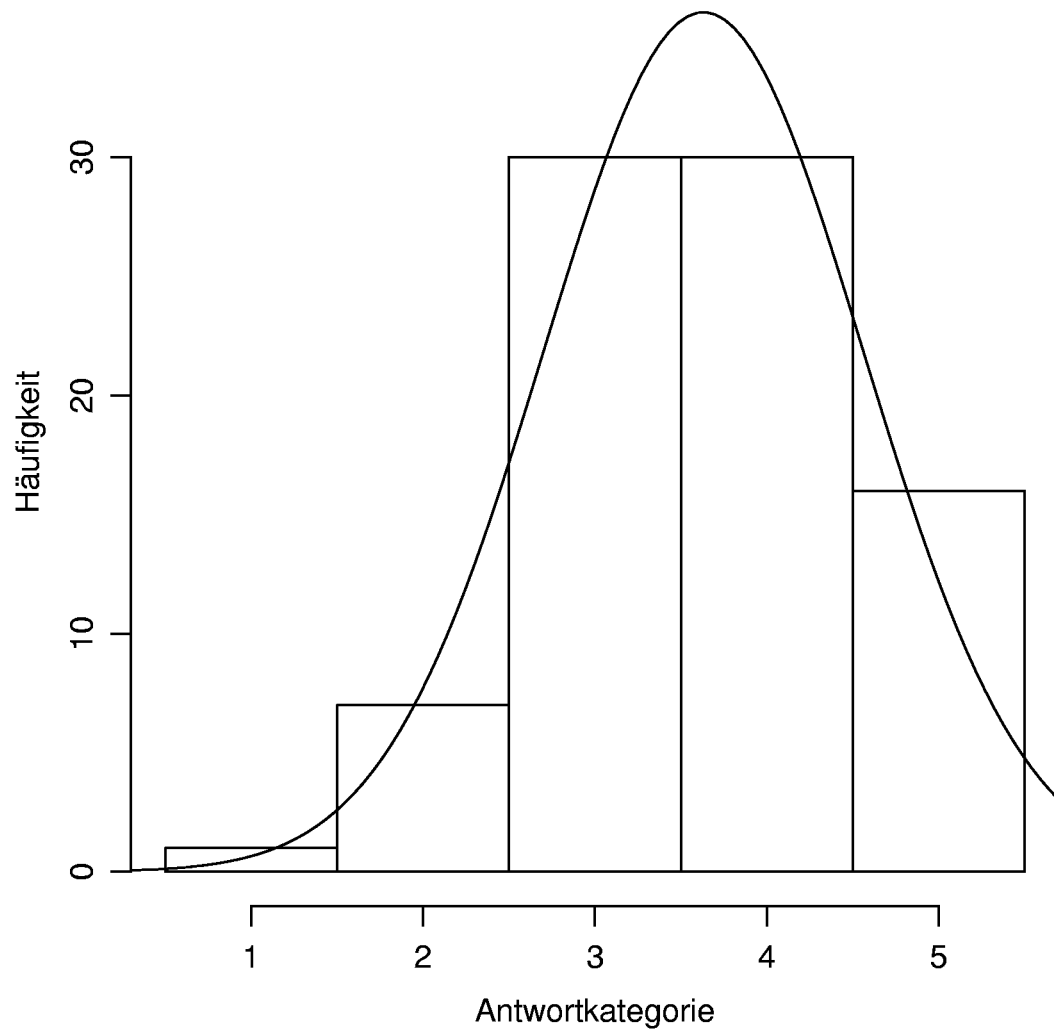


Abb. 115: Verteilung der Antworten für Item U04 zum zweiten Messzeitpunkt

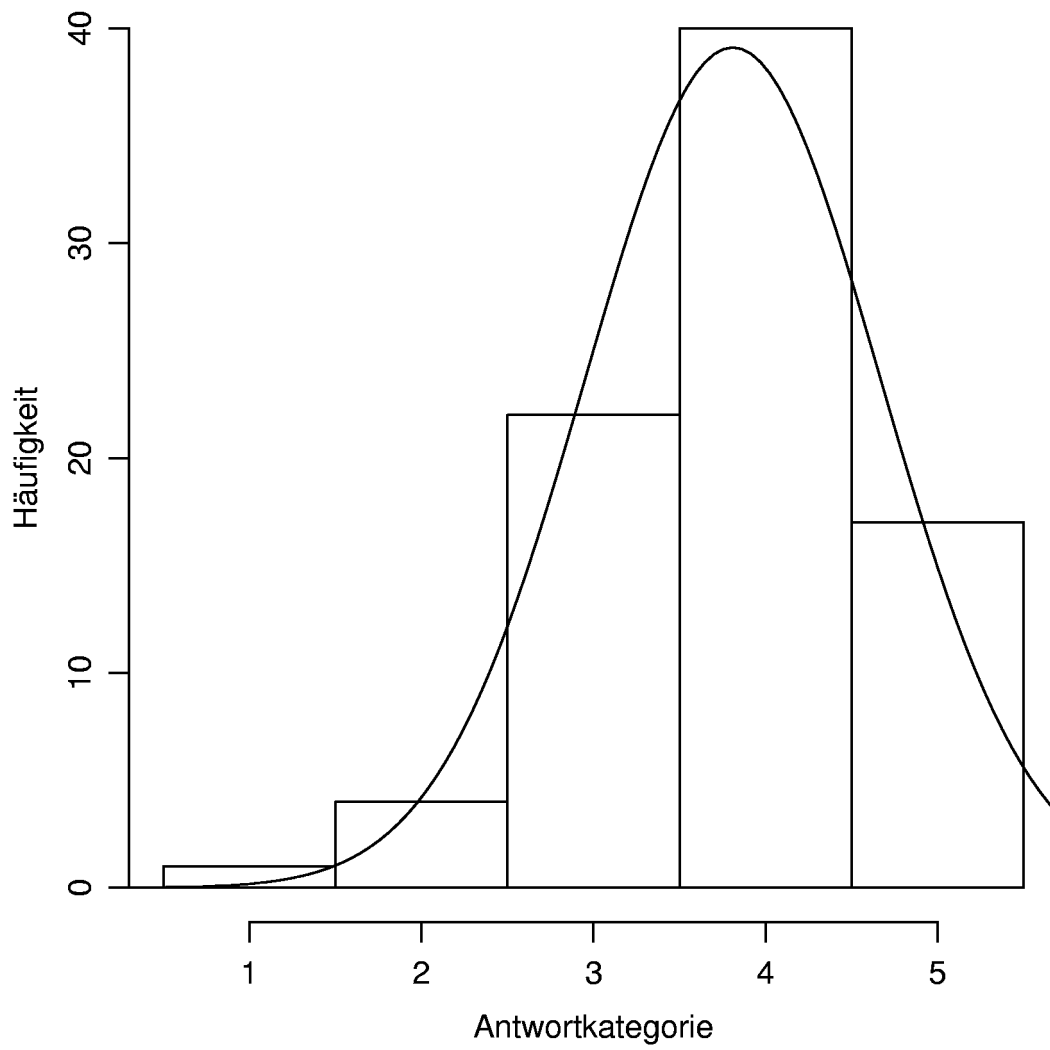


Abb. 116: Verteilung der Antworten für Item U05 zum ersten Messzeitpunkt

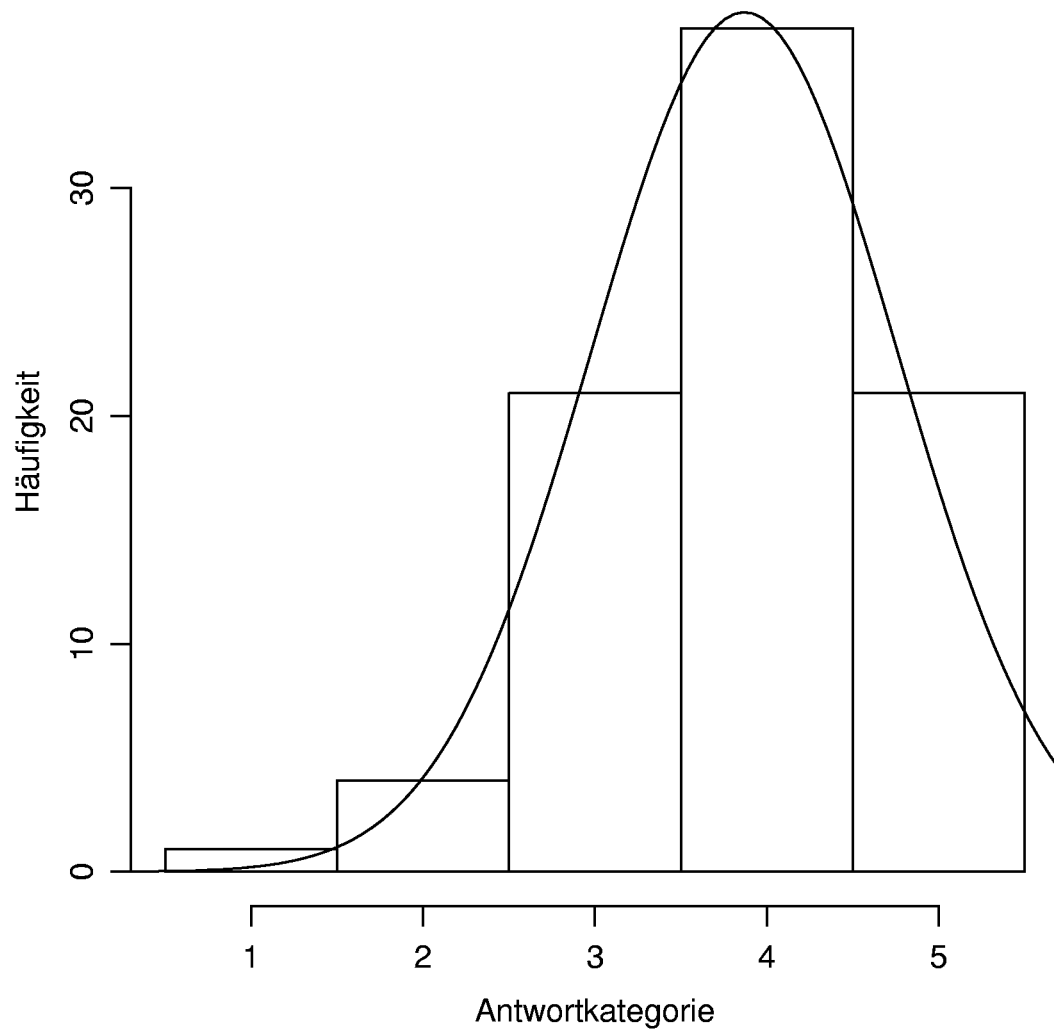


Abb. 117: Verteilung der Antworten für Item U05 zum zweiten Messzeitpunkt

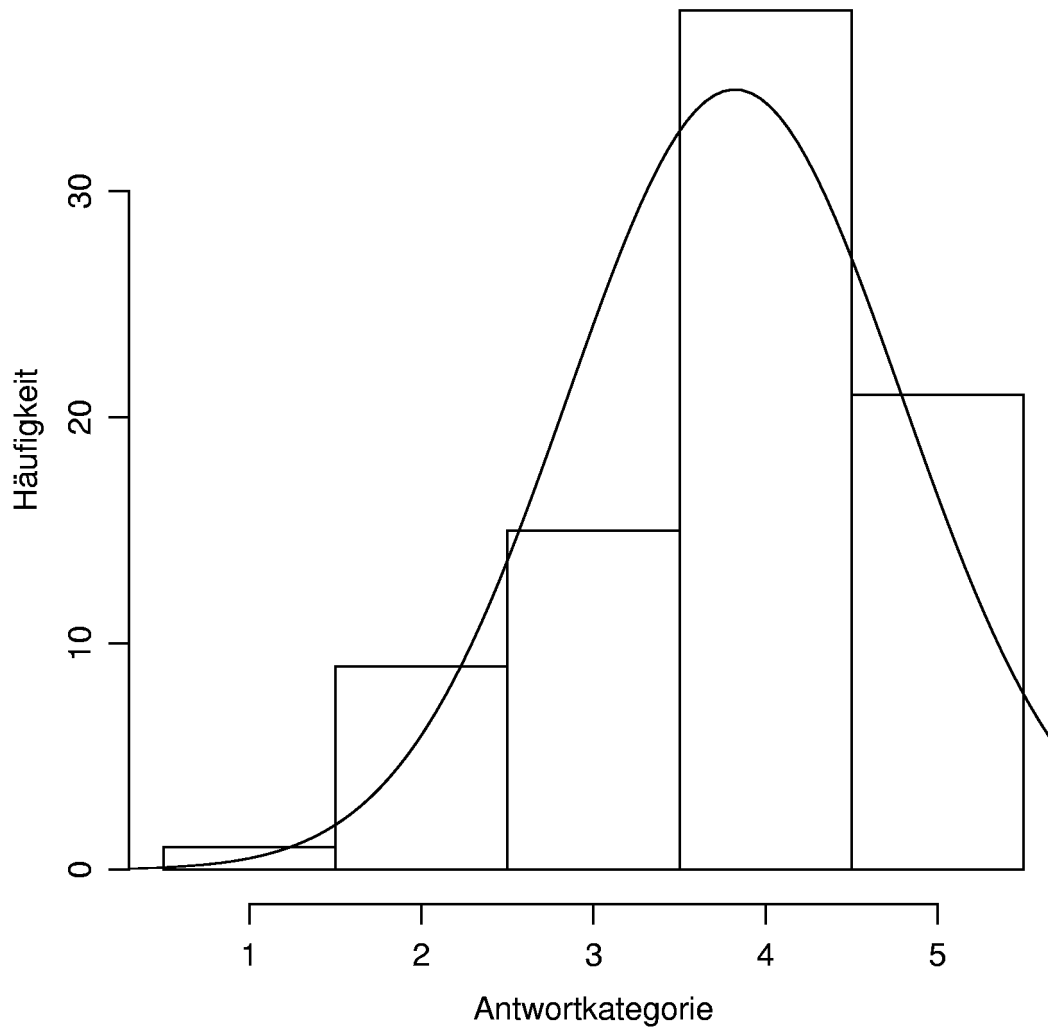


Abb. 118: Verteilung der Antworten für Item U06 zum ersten Messzeitpunkt

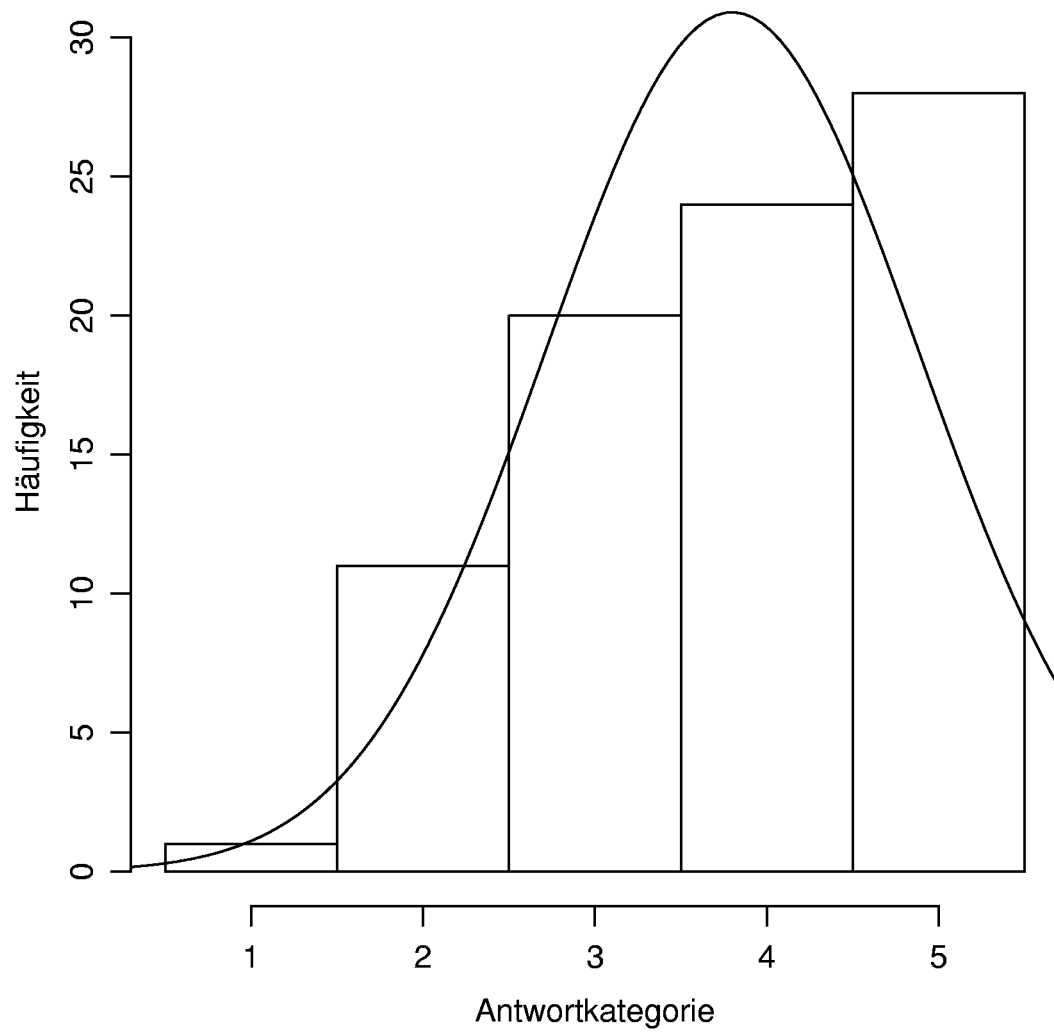


Abb. 119: Verteilung der Antworten für Item U06 zum zweiten Messzeitpunkt

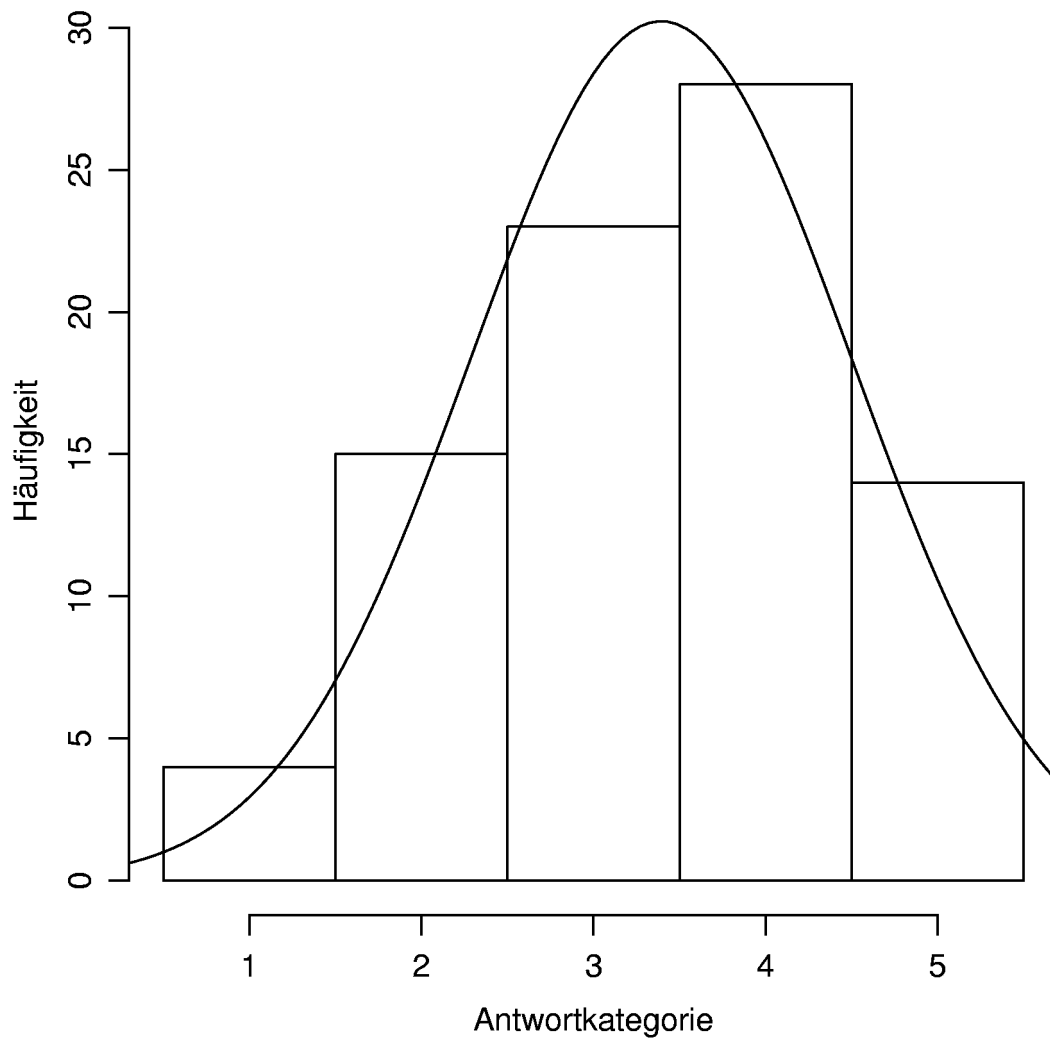


Abb. 120: Verteilung der Antworten für Item I01 zum ersten Messzeitpunkt

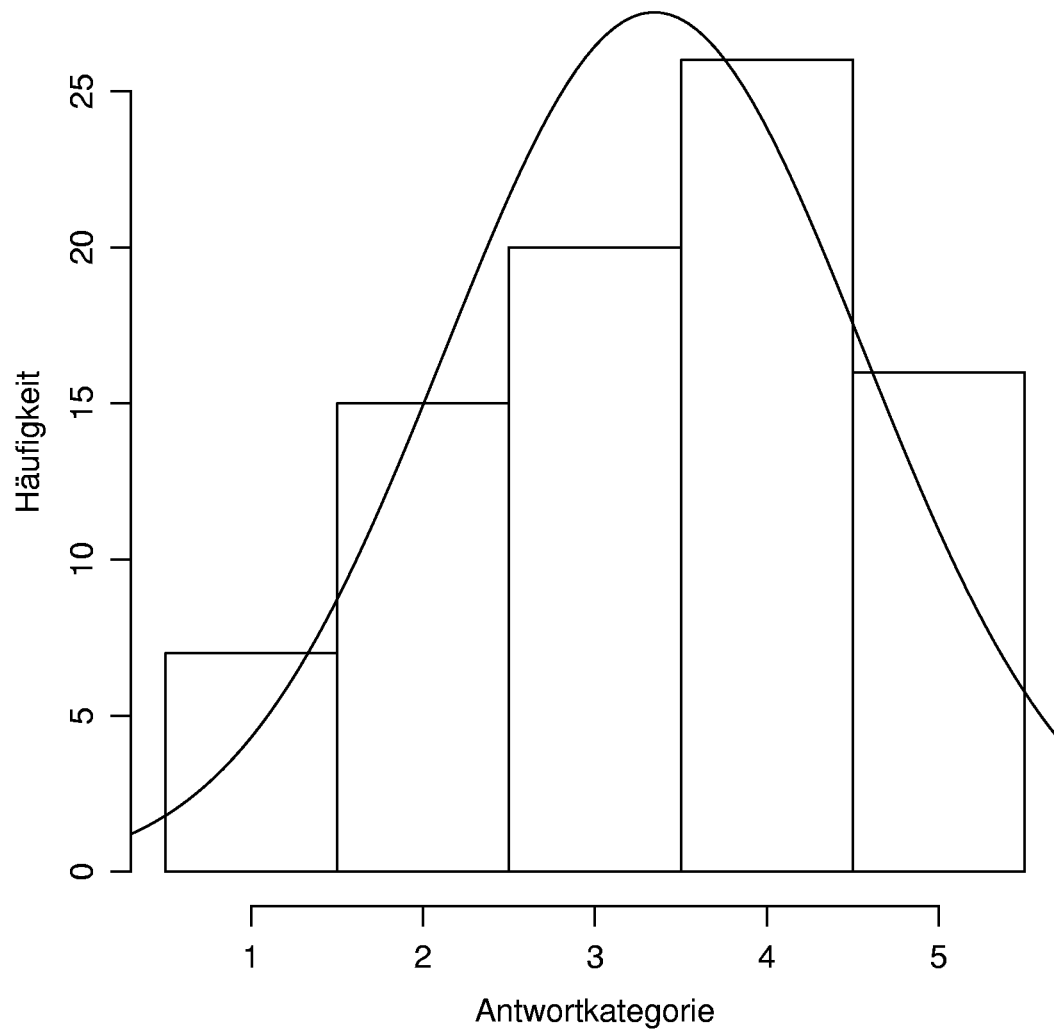


Abb. 121: Verteilung der Antworten für Item I01 zum zweiten Messzeitpunkt

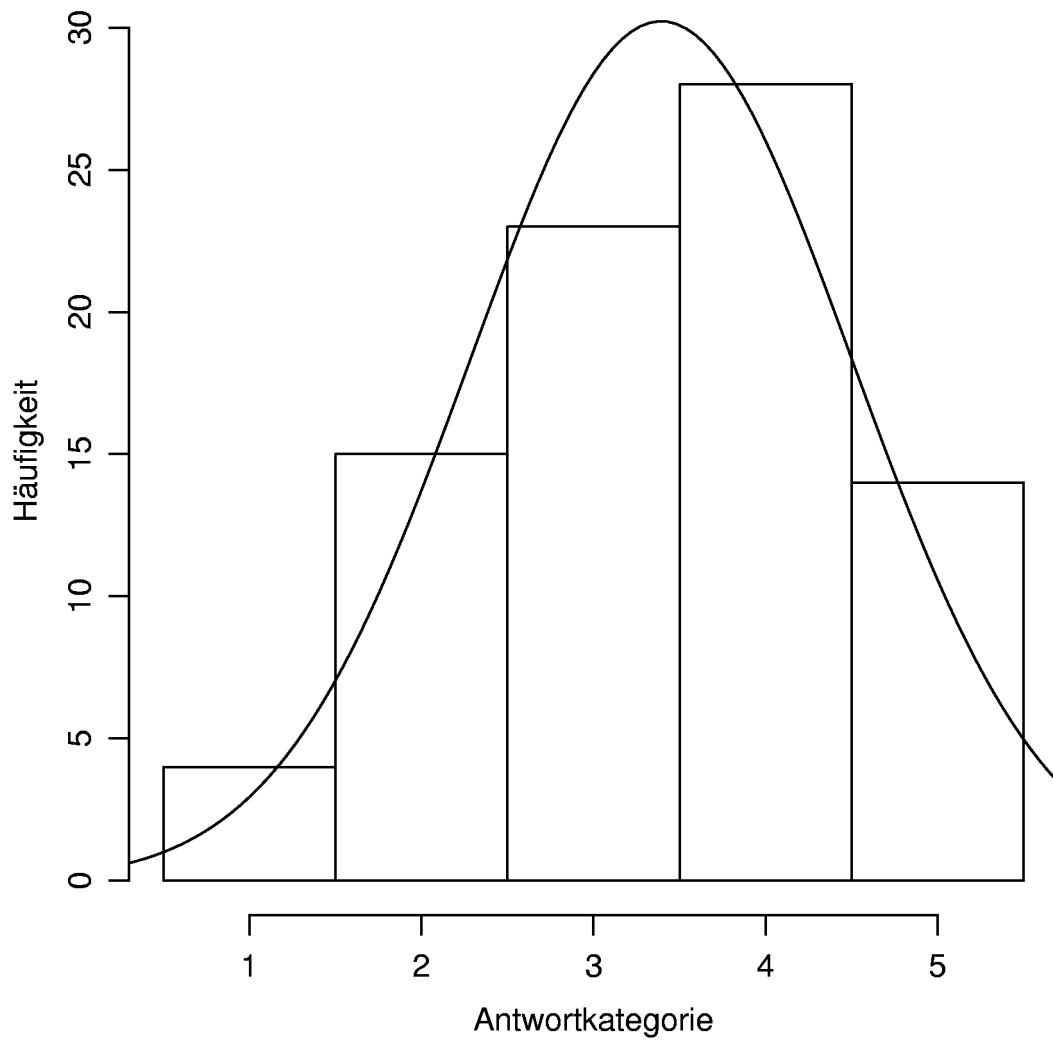


Abb. 122: Verteilung der Antworten für Item I02 zum ersten Messzeitpunkt

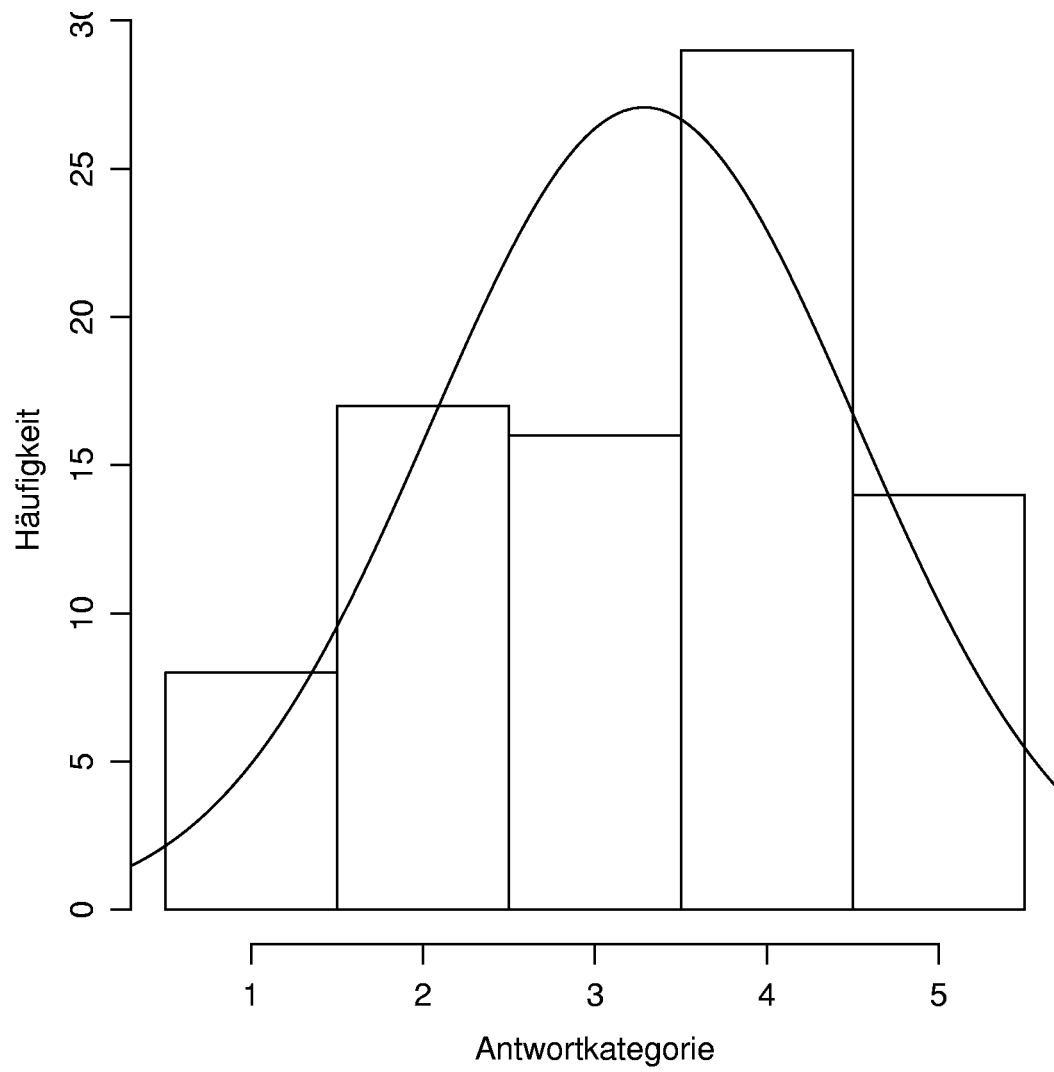


Abb. 123: Verteilung der Antworten für Item I02 zum zweiten Messzeitpunkt

Anhang M

Q-Q-Plots der gebildeten Skalenwerte im Vergleich zur Normalverteilung

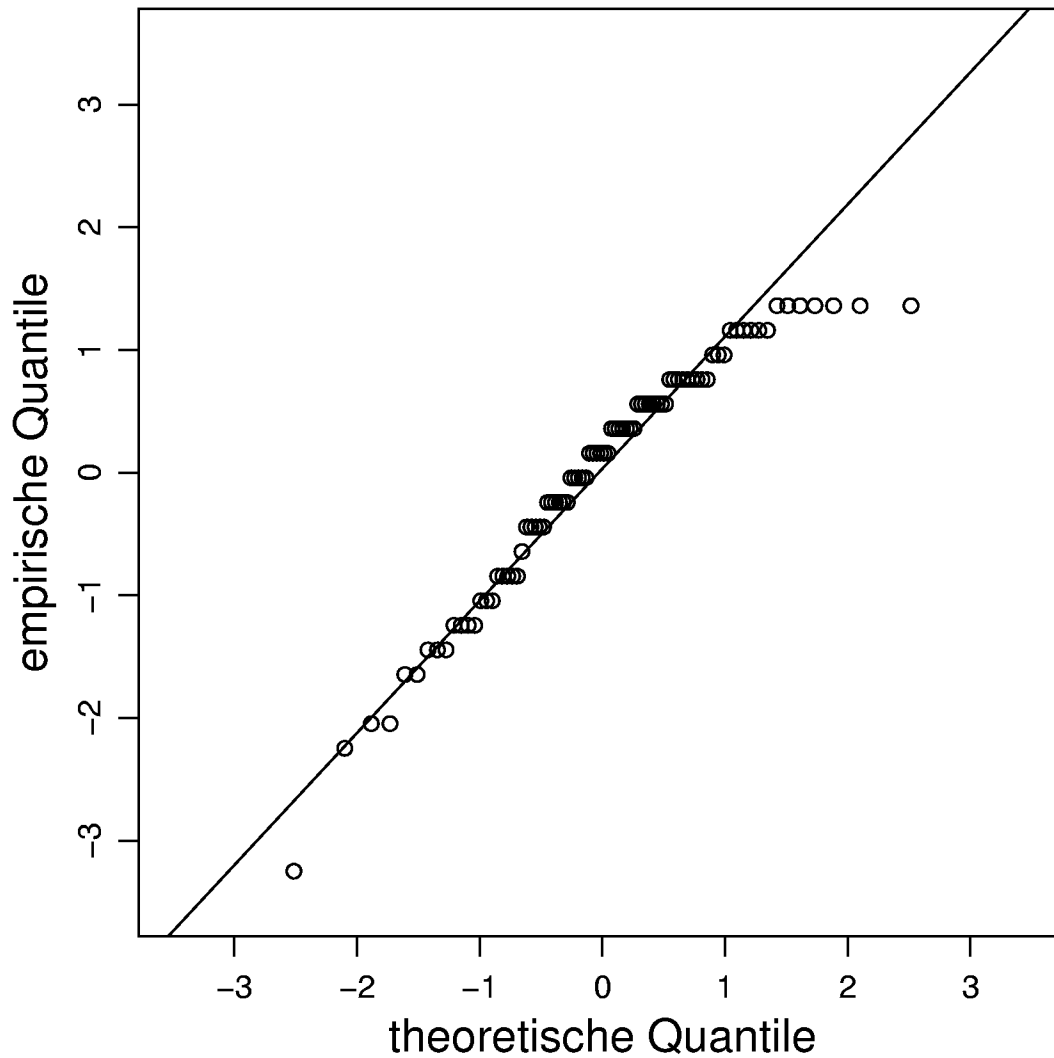


Abb. 124: Q-Q-Plot der PL-Skala (z-transformiert) für die erste Messung

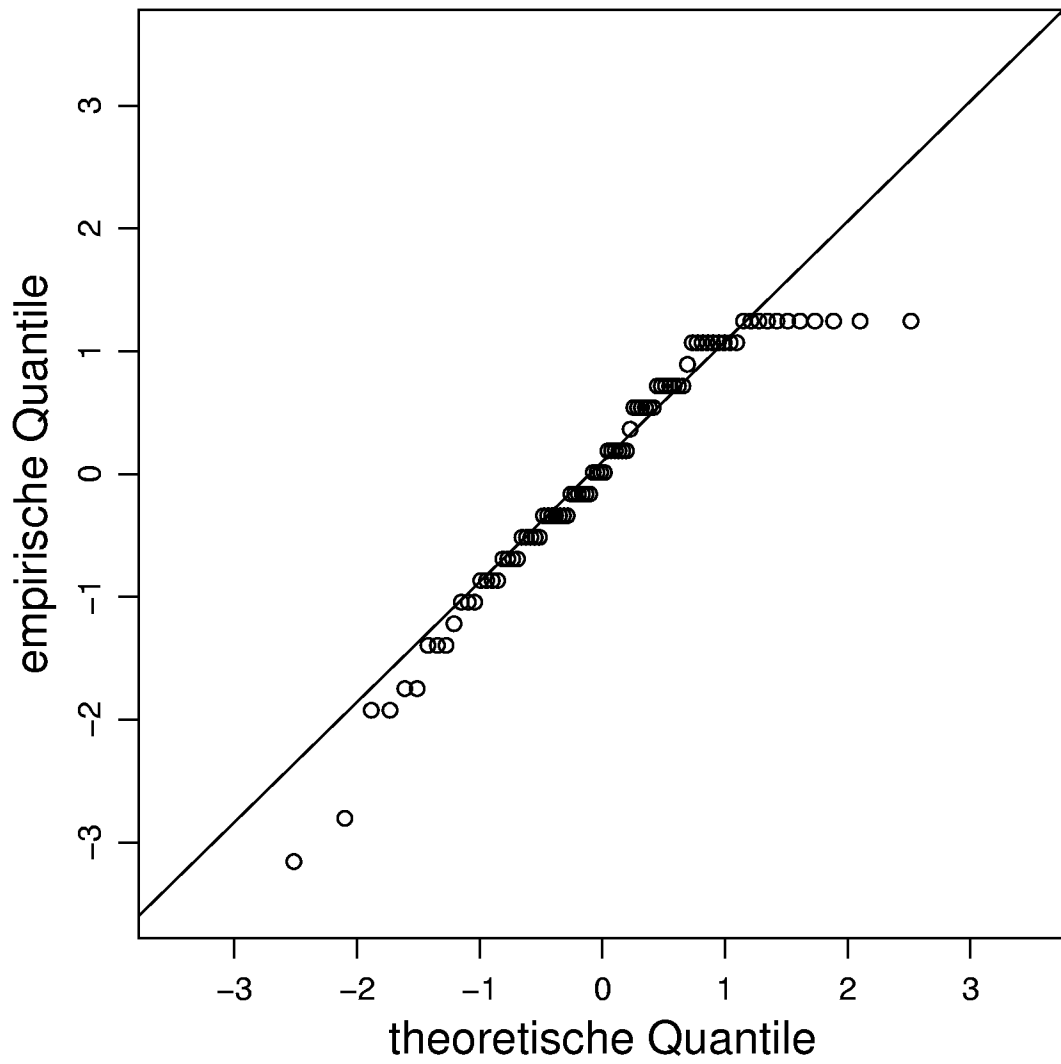


Abb. 125: Q-Q-Plot der PL-Skala (z-transformiert) für die zweite Messung

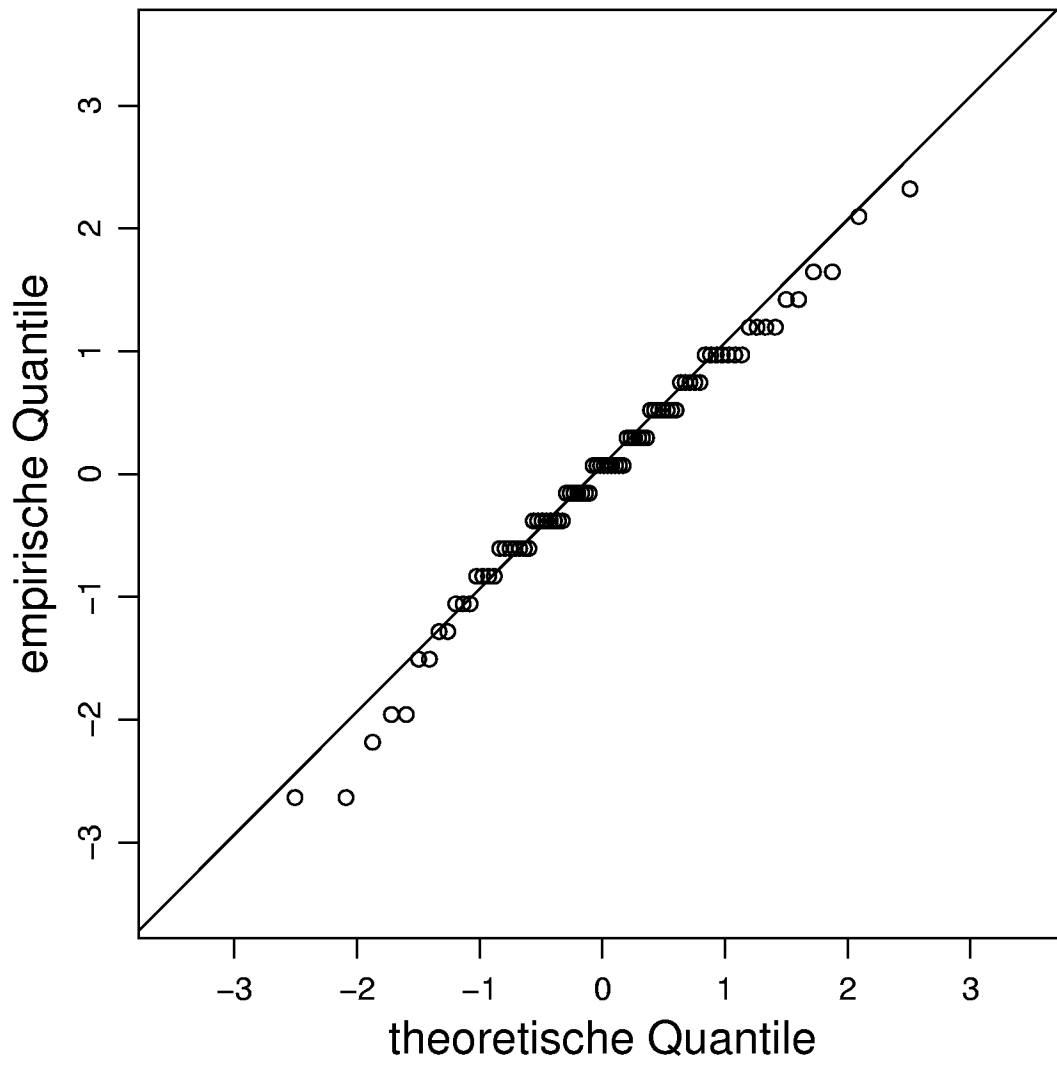


Abb. 126: Q-Q-Plot der PEUEU-Skala (z-transformiert) für die erste Messung

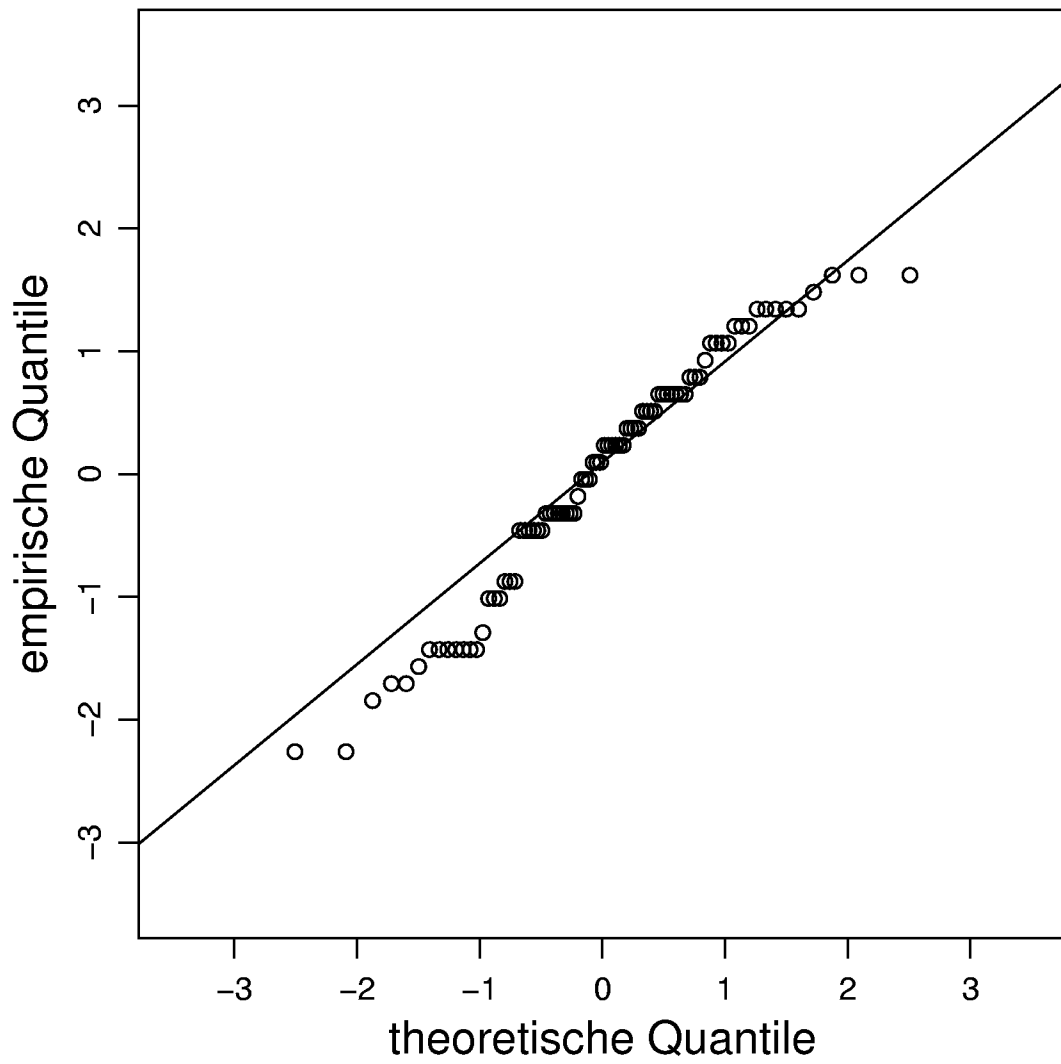


Abb. 127: Q-Q-Plot der PEUEU-Skala (z-transformiert) für die zweite Messung

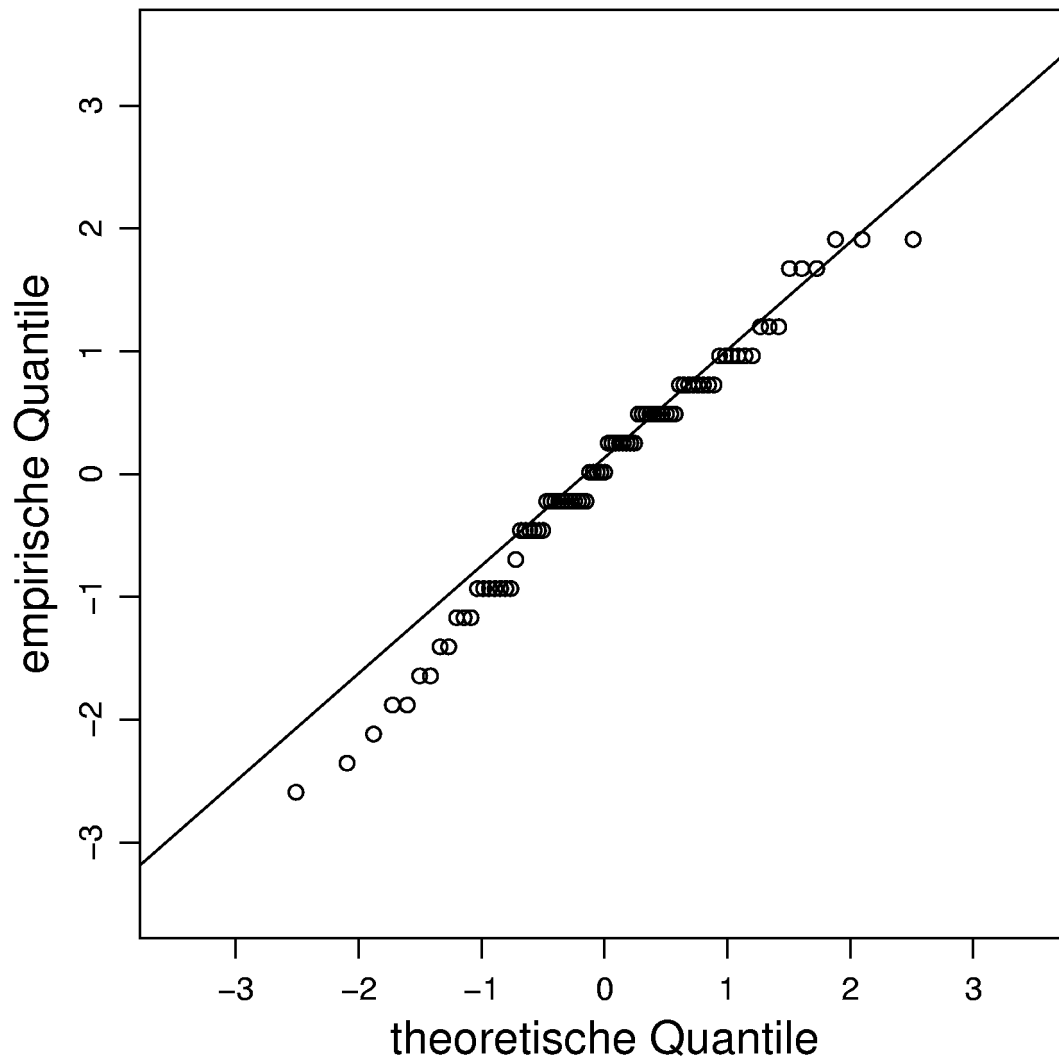


Abb. 128: Q-Q-Plot der PU-Skala (z-transformiert) für die erste Messung

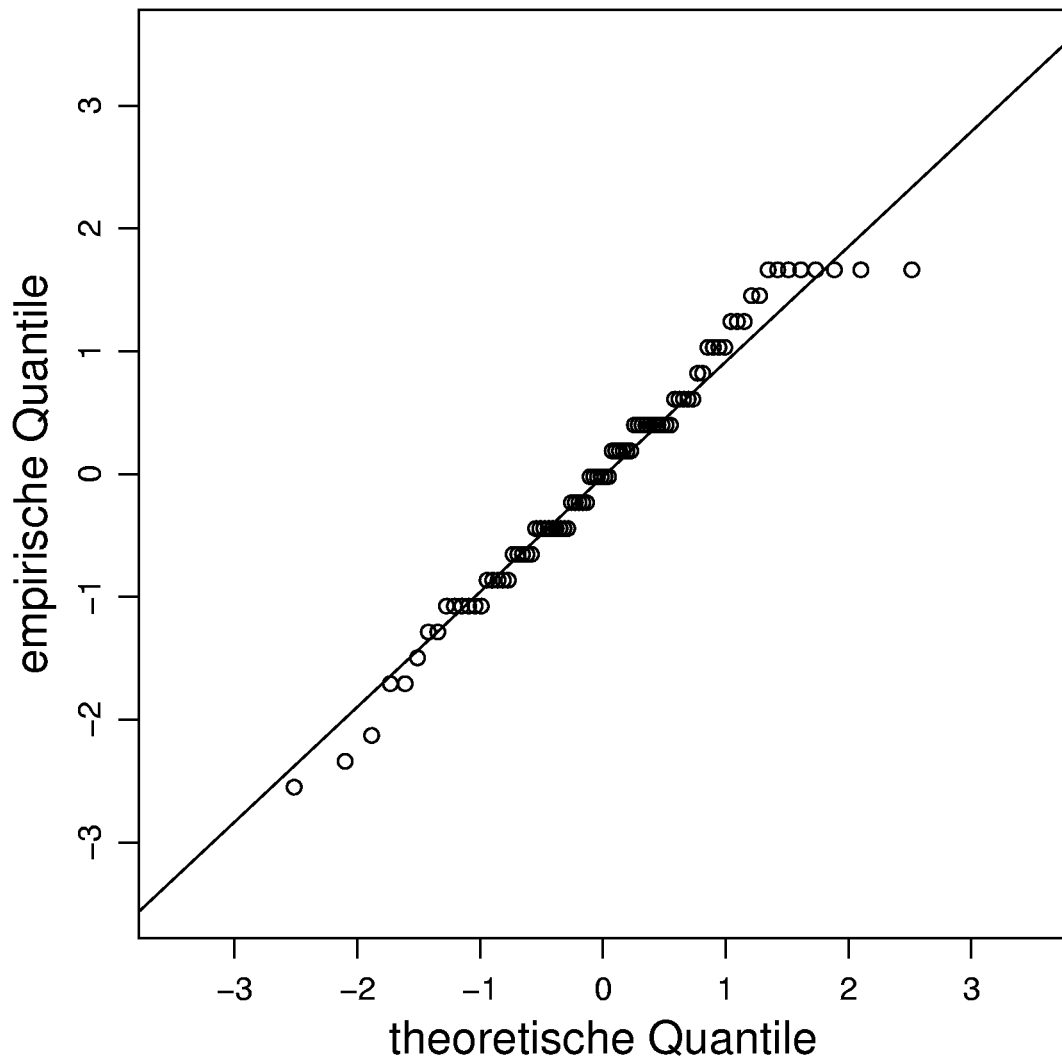


Abb. 129: Q-Q-Plot der PU-Skala (z-transformiert) für die zweite Messung

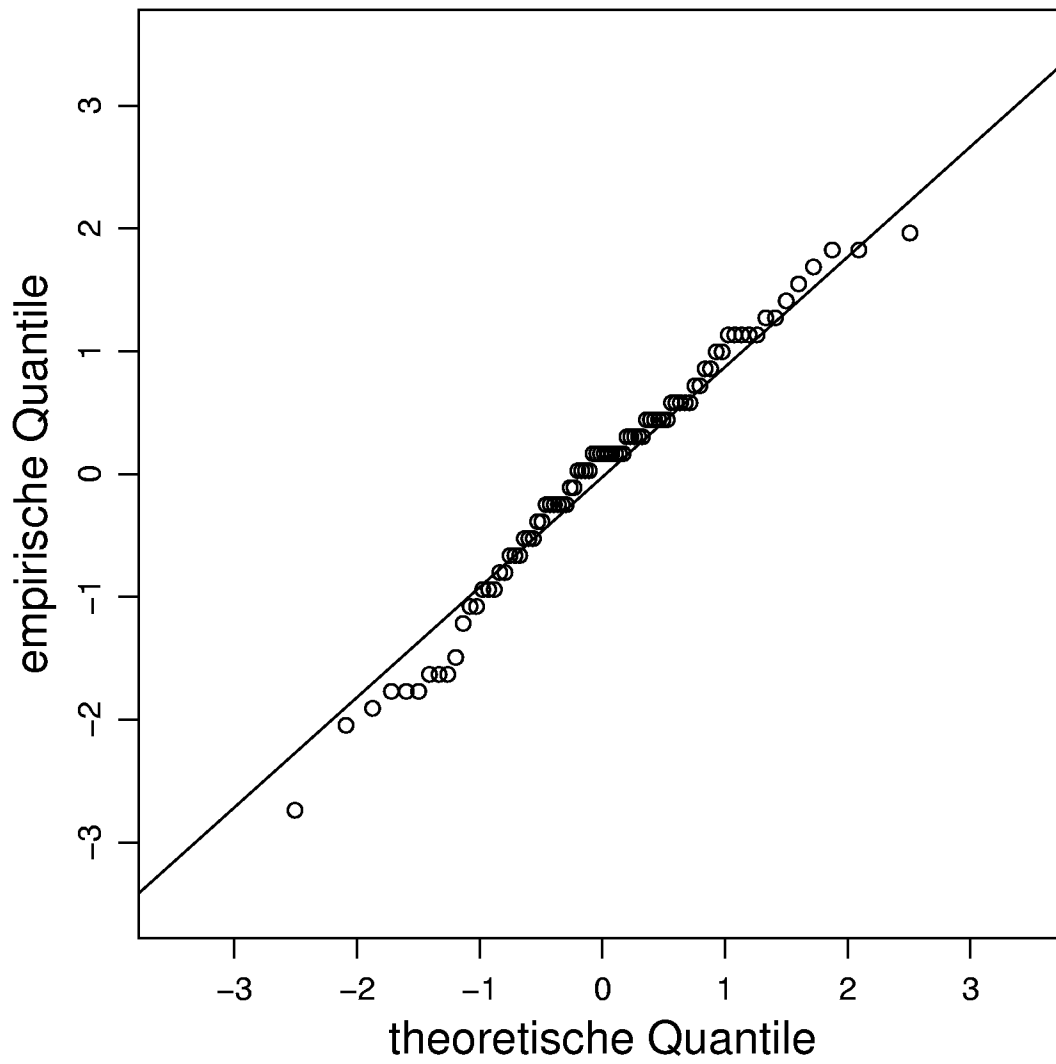


Abb. 130: Q-Q-Plot der PEOU-Skala (z-transformiert) für die erste Messung

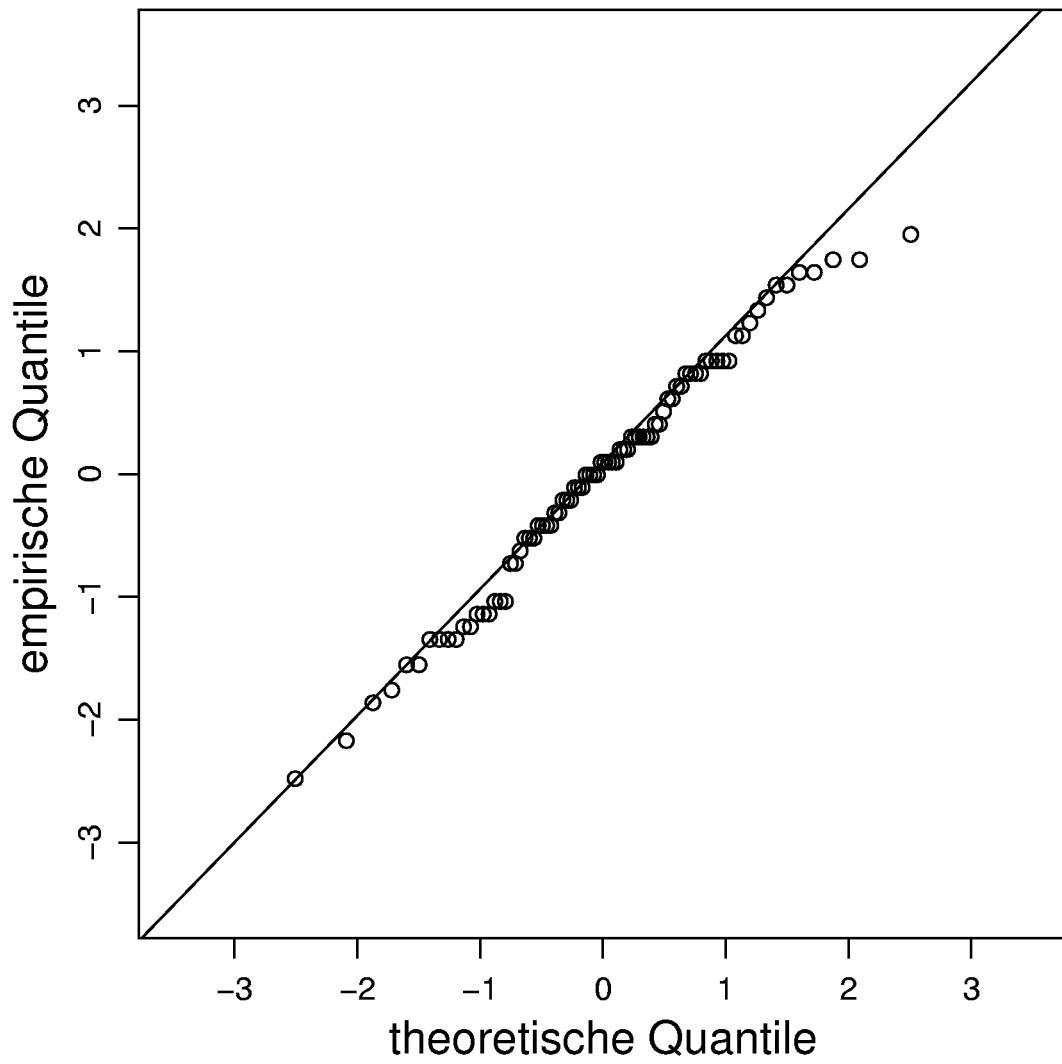


Abb. 131: Q-Q-Plot der PEOU-Skala (z-transformiert) für die zweite Messung

