

Universität Bern
Institut für Psychologie
Abteilung für Persönlichkeitspsychologie, Differentielle Psychologie und Diagnostik

und

Universität Basel
Institut für Psychologie
Forschungsschwerpunkt Mensch-Maschine Interaktion

Intranet Satisfaction Questionnaire

**Entwicklung eines Messinstruments zur Erhebung der
Benutzerzufriedenheit im Intranet**

**Lizentiatsarbeit von
Jonas Loetscher**

Eingereicht bei
Prof. Dr. Gerhard Faßnacht (Referent, Universität Bern)

Betreut durch
Lic. phil. Javier Bargas-Avila (Koreferent, Universität Basel)
Prof. Dr. Gerhard Faßnacht (Referent, Universität Bern)

Bern, Juli 2006

Danksagung

An dieser Stelle möchte ich mich bei all den Menschen bedanken, ohne welche die vorliegende Arbeit nicht möglich gewesen wäre.

Namentlich hervorheben und herzlich bedanken möchte ich mich bei lic. phil. Javier Bargas-Avila von der Universität Basel. Sein grosses Fachwissen und Engagement, seine stete Hilfsbereitschaft und die motivierenden Worte sowie die Durchsicht meines Manuskripts waren mir eine grosse Unterstützung. Zu grossem Dank verpflichtet bin ich Prof. Dr. Gerhard Fassnacht, der sich bereit erklärte, die vorliegende Arbeit gemeinsam mit der Universität Basel zu betreuen und die Arbeit zu begutachten.

Den verantwortlichen Personen des Unternehmens A, der Intranetmangerin der Direktion für Entwicklung und Zusammenarbeit (DEZA) sowie all den Angestellten, die den Fragebogen ausgefüllt haben, danke ich für die wertvolle Zusammenarbeit.

Von ganzem Herzen danke ich all den lieben Menschen in meinem Umfeld, von denen ich mich stets unterstützt wusste.

Zusammenfassung

Intranets haben bei Unternehmen in den letzten Jahren als firmeninterne Arbeits- und Informationsplattformen enorm an Bedeutung gewonnen. In diesem Kontext stellt sich zunehmend die Frage der Qualitätssicherung von Intranets. Die Erhebung der Benutzerzufriedenheit bietet hierfür einen Zugang. Zurzeit fehlt es an einem zuverlässigen Messinstrument für deren Erfassung, da vorhandene Instrumente zu kurz greifen. In der vorliegenden Arbeit wird die Entwicklung eines Messinstruments zur Erhebung der Benutzerzufriedenheit im Intranet vorgestellt. Nach zwei Testvalidierungen ($N_1 = 881$ und $N_2 = 127$) bei zwei verschiedenen Unternehmen entstand schrittweise ein aus 13 Item bestehender Fragebogen zur Messung der Benutzerzufriedenheit im Intranet. Die beiden vorgängigen Skalen mit 18 bzw. 16 Items verfügen über eine interne Konsistenz von .89 bzw. .84. Die entwickelte Endversion des *Intranet Satisfaction Questionnaire* mit 13 Items gilt es noch zu validieren.

Inhaltsverzeichnis

1 Einleitung	1
2 Theorie	3
2.1 Das Intranet: Computernetzwerk in Organisationen	3
2.2 Benutzerzufriedenheit mit dem Intranet	4
2.2 Messinstrumente zur Erhebung der Benutzerzufriedenheit	8
2.3 Qualität von Informationen aus Sicht der Benutzer	11
2.4 Benutzerfreundliche Mensch-Maschine Schnittstelle	14
2.5 Das Intranet und seine Nutzer	20
3 Methode	25
3.1 Fragebogenkonstruktion	26
3.1.1 Skalierung	26
3.1.2 Testlänge und Testzeit	27
3.1.3 Itemgenerierung und erste Testversion	28
3.2 Stichproben.....	30
3.2.1 Stichprobengewinnung	30
3.2.2 Stichprobenrepräsentativität	30
3.3 Erhebungsmethode und Durchführung	31
3.3.1 Allgemeine Bemerkungen zu webbasierten Umfragen	31
3.3.2 Durchführung und Instruktion	33
3.3.2.1 Beschreibungen der eingesetzten Online-Fragebögen	34
3.4 Anmerkungen zu den Auswertungen der Rohdaten	37
4 Testvalidierung I	39
4.1 Stichprobenbeschreibung	39
4.1.1 Datenexploration der Rohwerte	40
4.1.1.1 Verteilung der Rohdaten	40
4.1.1.2 Ausreisser	40
4.1.2 Missing Data	42
4.3 Itemanalyse.....	45
4.3.1 Statistische Kennwerte	45
4.3.2 Schwierigkeitsindizes	47
4.3.3 Trennschärfe	49
4.3.4 Homogenität	51
4.3.4.1 Interkorrelation	52
4.3.4.2 Faktorenanalyse	53
4.3.5 Reliabilität	59
4.3.4 Kriteriumsbezogene Validität	59
4.4 Diskussion	60
4.4.1 Skala	60
4.4.2 Items	60

5 Testvalidierung II	65
5.1 Stichprobenbeschreibung	65
5.2 Itemanalyse.....	66
5.2.1 Statistische Kennwerte	66
5.2.2 Schwierigkeitsindizes	68
5.2.3 Trennschärfe	69
5.2.4 Homogenität	70
5.2.4.1 Interkorrelation.....	70
5.2.4.2 Faktorenanalyse.....	71
5.2.5 Reliabilität	74
5.2.6 Kriteriumsbezogene Validität	74
5.3 Diskussion	74
5.3.1 Faktorenstruktur	74
5.3.2 Itemrevision.....	76
5.3.3 Vorläufige Endversion des Fragebogens.....	79
6 Diskussion	81
7 Literaturverzeichnis	85
8 Anhang	93

1 Einleitung

Immer mehr Unternehmen verfügen über ein Intranet oder erwägen, diese Informationstechnologie in nächster Zeit einzusetzen (Hoffmann, 2000). Die firmeninternen Arbeits- und Informationsplattformen unterstützen heute täglich tausende von Mitarbeitenden grosser und mittlerer Unternehmen bei ihrer Arbeit und sind aus dem Arbeitsalltag kaum mehr wegzudenken. In diesem Kontext stellt sich zunehmend die Frage der Qualitätssicherung von Intranets. Insbesondere zuständige Entwicklungsteams haben ein Bedürfnis nach regelmässiger „Pulsmessung“: Wie gut ist unser Intranet aus Benutzersicht? Einen möglichen Zugang hierfür bietet die Erfassung der Benutzerzufriedenheit. Die Zufriedenheit jener Personen also, für die das Intranet konzipiert wurde und die es bei ihrer Arbeit unterstützen soll.

Das Interesse am Konstrukt Benutzerzufriedenheit in der Literatur zu Informationssystemen ist nicht neu. Mit dem Aufkommen des Personal Computers (PC) am Arbeitsplatz in den 80er Jahren und der damit schnell wachsenden Anzahl von Personen, die mit Computern in Berührung kamen, wurden Fragen der Benutzerzufriedenheit und Benutzerfreundlichkeit zunehmend wichtiger. Die Geräte sollten nicht mehr nur von Profis bedient werden können, sondern auch von Laien. Mitte der 80er Jahre entstand mit *Human-Computer-Interaction (HCI)* eine Forschungsrichtung, die sich als Zusammenschluss von Informatik und kognitiven Wissenschaften verstand. Die Berücksichtigung des menschlichen Denkens bei der Entwicklung von Computersystemen sollte diese für die Menschen in ihrer Handhabung vereinfachen und leichter erlernbar machen (Preece et al., 1999).

Aus ökonomischer Sicht erhoffte man sich durch verbesserte Anpassungen an den Menschen effizientere Systeme und damit finanzielle Vorteile. Dem Konstrukt Benutzerzufriedenheit kommt hier insofern eine Bedeutung zu, da es in der Literatur gerne als Substitut für Effektivität (Melone, 1990) oder Erfolg (DeLone & McLean, 1992; Huang et al., 2004) eines Informationssystems verwendet wurde. Wobei im Fall der Effektivität die Art der Beziehung nicht unumstritten ist. Denn nicht in jedem Fall bringt ein effektives Informationssystem zwangsläufig auch zufriedene Benutzer hervor (Melone, 1990).

Seit dem Einzug des Computers am Arbeitsplatz durchlief dieser eine enorme Weiterentwicklung, deren Ende noch nicht abzusehen ist. Mit dem wachsenden Potential von Computern stiegen auch die Anforderungen an der Benutzerführung. Interaktive Software und die Verbreitung von Netzwerktechnologien veränderten die Nutzungsmöglichkeiten des

Computers nachhaltig. Messinstrumente zur Erhebung der Benutzerzufriedenheit älteren Datums wie von Bailey und Pearson (1983) erweisen sich im heutigen Kontext als nicht mehr adäquat. Sie entwickelten ihren Fragebogen in einem Umfeld, welches mit dem aktuellen nicht mehr vergleichbar ist. Etwas neuere Messinstrumente wie von Doll und Torkzadeh (1988) oder Lewis (1995) erwiesen sich als relativ valide und reliabel. Die Autoren beabsichtigten in beiden Fällen, einen allgemein gültigen Fragebogen zur Erhebung der Benutzerzufriedenheit zu entwickeln. Für die Erfassung der Benutzerzufriedenheit mit dem Intranet greifen sie jedoch zu kurz. So ist anzunehmen, dass zentrale Aspekte des Intranets wie die Mitarbeiterkommunikation oder die Suchmaschine die Zufriedenheit der Benutzer beeinflussen. Die Arbeiten von Doll und Torkzadeh (1988) sowie von Lewis (1995) bieten so zwar eine gute Grundlage für ein Messinstrument, für eine genauere Erfassung der Benutzerzufriedenheit scheint es allerdings erforderlich, intranetspezifische Komponenten stärker zu berücksichtigen.

Ziel der vorliegenden Arbeit ist deshalb die Entwicklung eines Messinstrumentes zur Erhebung der Benutzerzufriedenheit im Intranet in Form eines strukturierten Fragebogens. Für die Skalenkonstruktion wurde eine explorative Herangehensweise gewählt. Die Itemanalyse stützt sich auf die klassische Testtheorie.

Der Aufbau der vorliegenden Arbeit sieht wie folgt aus: Im 2. Kapitel werden die notwendigen Grundlagen und Begriffsklärungen vermittelt. Es bildet zugleich die theoretische Basis für die Skala und die Itemgenerierung. Kapitel 3 klärt über die methodischen Vorgehensweisen zur Skalenkonstruktion auf und erläutert methodische Aspekte der Datenerhebung. Der Fragebogen wurde für die vorliegende Arbeit zweimal validiert. In Kapitel 4 wird die erste, in Kapitel 5 die zweite Testvalidierung präsentiert. Innerhalb dieser zwei Blöcke werden die Ergebnisse jeweils besprochen. Das Kapitel 6 dient der abschliessenden Diskussion.

2 Theorie

Der erste Abschnitt dieses Kapitels dient der Klärung der für diese Arbeit zentralen Begriffe „Intranet“ und „Benutzerzufriedenheit“. Für das Intranet geschieht dies zum einen durch die Bestimmung der eingesetzten Technologie, zum anderen durch die Abgrenzung der angesprochenen Zielgruppe. Der Begriff der Benutzerzufriedenheit wird anhand der bestehenden Literatur erschlossen und dient als Ausgangspunkt für die theoretische Grundlage auf welcher das entwickelte Messinstrument basiert. Nach einer Aufführung bestehender Messinstrumente, werden abschliessend drei Aspekte näher beleuchtet, welche für die Itemgenerierung als wichtig erachtet werden. Dies sind die Qualität der angebotenen Informationen aus Benutzersicht, die Schnittstelle Mensch-Computer sowie das Intranet unter dem Blickwinkel seiner Nutzung.

2.1 Das Intranet: Computernetzwerk in Organisationen

Ein Intranet ist ein Rechnernetzwerk¹ auf Basis der Internet-Technologie, zu welchem nur eine festgelegte Gruppe von Mitgliedern einer Organisation Zugang hat (Hoffmann, 2001).

Unter Internet-Technologie bezeichnet man eine Gruppe von Software- und Hardware-Komponenten, die Protokolle (z.B. TCP/IP, HTTP), Namenskonventionen, Sprachen (z.B. HTML, XML) und Schnittstellen implementieren (Kaiser, 2000, S. 10). Kaiser (2000) nennt drei Eigenschaften, welche diese Technologie für Unternehmen äusserst attraktiv machen. Erstens ermöglicht sie einen weltweiten Zugriff durch das globale Adressierungssystem URI (*Uniform Resource Identifier*). Zweitens können nahezu beliebige Formate wie Text, Grafik, Audio, Video in einem Dokument integriert werden (multimediale Integration) und schliesslich drittens ist es möglich, über Hyperlinks beliebige Ressourcen miteinander zu verknüpfen (Vernetzung). Damit eröffnen sich für Unternehmen praktisch unbegrenzte Möglichkeiten der Nutzung, sei es als universelle Kommunikationsplattform, Informations- und Wissenspool oder für spezifische Applikationen. Ein Beispiel für die vielfältigen Einsatzmöglichkeiten illustriert das Intranet von IBM. Dort wurde das Mitarbeiterverzeichnis

¹ Netzwerkkommunikation ist eine komplexe Angelegenheit, in der zahllose Abläufe mit unterschiedlichen Protokollen involviert sind. Vereinfachend kann dieser Prozess anhand eines Schichtenmodells beschrieben werden. Jede Schicht repräsentiert dabei eine gewisse Funktion, die beim Austausch von Daten zwischen Anwendungen über ein dazwischen liegendes Netzwerk ausgeführt werden. Beim Transport von Daten wird jede dieser Schichten durchlaufen, zuerst beim Quellcomputer von oben nach unten, dann wieder beim Zielcomputer von unten nach oben. Das Kommunikationsprotokoll für das Intranet ist wie beim Internet das auf Paketvermittlung basierende TCP/IP (*Transmission Control Protocol/Internet Protocol*). Die IP-Adressen ermöglichen, dass zwei Computer miteinander eine Verbindung aufbauen können (Bundesamt für Information und Telekommunikation BIT, 2003; Pott, 1998).

mit Audio-Files ausgestattet. So können sich die Mitarbeiter vor einem Treffen oder Telefonat über die richtige Aussprache des Namens einer anderen Person erkundigen. In einem multinationalen Unternehmen wie der IBM ein nützlicher Service (Nielsen, 2006).

Je nach Benutzergruppe lassen sich die Web-Plattformen eines Unternehmens in Intranet, Extranet und Internet einteilen. Bei einem Intranet haben ausschliesslich die Mitarbeiter eines Unternehmens Zugang. Haben auch andere namentlich bekannte Benutzergruppen wie z.B. Kunden, Lieferanten oder Geschäftspartner Zugriff, spricht man von einem Extranet. Steht der Zugang grundsätzlich allen Benutzern offen, wird im Allgemeinen der Begriff Internet verwendet (Kaiser, 2000; Kortzfleisch, 1997). Während sich die drei Formen auf technischer Ebene nicht unterscheiden, bestehen deutliche Unterschiede auf funktionaler Ebene. Bei Intranets und Extranets steht vor allem eine effiziente Unterstützung von Prozessen im Vordergrund, ein Internetauftritt wird meist als Instrument für das Marketing genutzt. Die Benutzeroberflächen von Intra- und Extranets werden so nach funktionalen Kriterien konzipiert, die Benutzeroberflächen von Internets werden inhaltlich, optisch und zeitlich auf das Marketing ausgerichtet (Nielsen, 1997).

Zusammenfassend wird in der vorliegenden Arbeit ein Intranet als Teil des Informationssystems eines Unternehmens begriffen, welches auf der Web-Technologie basiert und dessen Nutzung nur den Mitarbeitern der Unternehmung vorbehalten ist.

2.2 Benutzerzufriedenheit mit dem Intranet

Das Konstrukt Benutzerzufriedenheit im Zusammenhang mit Computern wird in der Literatur meist als eine affektive Einstellung beschrieben. Bailey und Pearson (1983, S. 531) definieren Benutzerzufriedenheit als „sum of one’s positive and negative reactions to a set of factors.“ Doll und Torkzadeh (1988, S. 261) beschreiben es als „the affective attitude toward a specific computer application by someone who interacts with the application directly.“ Ähnlich Melone (1990, S. 81), welche es als „a predisposition to respond favorably or unfavorably to a computer system“ darlegt oder Harrison und Rainer (1996, S. 80), die Benutzerzufriedenheit als “an affective attitude toward all the various activities surrounding an end user’s interaction with a computer-based information system” umschreiben. Auch Fisher und Howell (2004, S. 247) betrachten Benutzerzufriedenheit als eine affektive Reaktion auf ein Informationssystem.

In der vorliegenden Arbeit wird Benutzerzufriedenheit ebenfalls als eine Einstellung begriffen. Einstellung wird dabei im Sinne von Eagly und Chaiken (1998, S. 296) als

„psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor“ verstanden. Nach Rosenberg und Hovland (1960) können Einstellungen dabei als Produkte aus affektiven, kognitiven und verhaltensbezogenen Prozessen aufgefasst werden, welche in affektiven, kognitiven und verhaltensbezogenen Reaktionen Ausdruck finden. Unter der kognitiven Komponente verstehen sie Überlegungen (Meinungen) über ein Einstellungsobjekt, zur affektiven Komponente zählen sie Gefühle und Emotionen, die vom Einstellungsobjekt ausgelöst werden, die Verhaltenskomponente beinhaltet für sie Handlungen und Handlungsabsichten in Bezug auf ein Einstellungsobjekt. Dies bedeutet jedoch nicht, dass bei der Entstehung (Prozess) oder beim Ausdruck (Reaktion) einer Einstellung alle drei Komponenten beteiligt sein müssen (Eagly und Chaiken, 1998).

Nach Huang und Mitarbeitern (2004) ist Benutzerzufriedenheit das wohl meist verwendete Konstrukt, um den Erfolg eines Informationssystems (*information system success*) zu messen. Eine Stellvertreterfunktion, auf die in der Literatur vielfach hingewiesen wird. Dem gegenüber existieren vergleichsweise wenige Forschungsarbeiten dazu, unter welchen Bedingungen Benutzerzufriedenheit generell entsteht. Es herrscht auch nur bedingt Konsens darüber, wie Benutzerzufriedenheit operationalisiert werden sollte. So lässt bspw. die ISO-Norm 9241-11 sehr viel Spielraum offen, wie Benutzerzufriedenheit gemessen werden kann.² In vielen Studien fehlt den Operationalisierungen des Konstrukts Benutzerzufriedenheit eine theoretische Basis (Melone, 1990).

² “Satisfaction can be specified and measured by subjective rating on scales such as discomfort experienced, liking for the product, satisfaction with product use, or acceptability of the workload when carrying out different tasks, or the extent to which particular usability objectives (such as efficiency or learnability) have been met. Other measures of satisfaction might include the number of positive and negative comments recorded during use. Additional information can be obtained from longer-term measures such as rate of absenteeism, observation of overloading or underloading of the user's cognitive or physical workload, or from health problem reports, or the frequency with which users request transfer to another job” (International Standardization Organisation, 1998, S. 6).

Die Operationalisierung für das vorliegende Messinstrument stützt sich im Kern auf den Erwartungs-mal-Wert-Ansatz von Fishbein (1967a, 1967b). Nach Fishbein kann eine Einstellung als Summe der Produkte von Erwartungen und Bewertungen verstanden werden. Diese Gleichung lässt sich durch folgende Formel darstellen:

$$A_o = \sum_{i=1}^n b_i e_i$$

Abbildung 2.1. Formel des Erwartungs-mal-Wert-Ansatzes von Fishbein übernommen aus Eagly und Chaiken (1998, S. 275).

Auf die vorliegende Arbeit adaptiert, steht das A_o der Formel stellvertretend für das Konstrukt Benutzerzufriedenheit mit dem Intranet. b_i ist die Erwartung, dass das Intranet gewisse Merkmale bzw. Attribute besitzt und e_i ist die Bewertung ebendieses Merkmals. Das n entspricht der Anzahl von Attributen. Damit A_o repräsentativ für das Konstrukt Benutzerzufriedenheit betrachtet werden kann, müssten deshalb theoretisch alle relevanten b 's bekannt sein. Der Theorieteil dient in diesem Sinne der Identifikation der Variablen b_1, b_2, \dots, b_i .

Die Formel setzt für b_1, b_2, \dots, b_i eine gleiche Bedeutung für die Benutzerzufriedenheit voraus. Das heisst, alle verwendeten Variablen müssten für den Benutzer dieselbe Relevanz besitzen. Einige Forscher begegnen diesem Problem, indem sie jede Variable oder ein Set von Variablen von den Probanden auch in Bezug auf die subjektive Wichtigkeit (Gewichtung) einschätzen lassen (Bailey & Pearson, 1983; Eagly & Chaiken, 1998). Bei der Entwicklung des vorliegenden Messinstruments wird eine solche subjektive „Gewichtungskategorie“ nicht eingesetzt. Dies wird damit begründet, dass für alle eingesetzten Variablen (Items) angenommen wird, dass sie für alle Intranetbenutzer von hoher Bedeutung sind. Die Erhebung von Gewichten würde so gesehen keinen Gewinn an Informationen beitragen. Anders formuliert wird also davon ausgegangen, dass die Items die Gewichtungen schon in sich tragen. Zudem bringt die Itemanalyse in Form der Item-Test-Korrelation und der Faktorenanalyse hier natürlich ein gutes Instrument für die Überprüfung der Konstruktvalidität jedes einzelnen Items.

Noch nicht beantwortet ist damit die Frage, welche Elemente b_i enthalten sollte, um Benutzerzufriedenheit valide zu messen. Einen interessanten Ansatz bietet die bereits erwähnte ISO-Norm 9241-11. Darin definiert sie *Usability* mit: „Extent to which a product

can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Effektivität wird dabei verstanden als "accuracy and completeness with which users achieve specified goals" und Effizienz beschrieben als "resources expended in relation to the accuracy and completeness with which users achieve goals" (International Standardization Organisation, 1998, S. 2). Auf den Intranetkontext bezogen bedeutet dies, dass ein Mitarbeiter eines Unternehmens das Intranet immer mit einer bestimmten Absicht benutzt wie bspw. die Telefonnummer eines anderen Mitarbeiter suchen, den er dringend erreichen sollte, das Protokoll einer Sitzung abrufen, an der er nicht teilnehmen konnte oder die Firmenmitteilungen aufsuchen, weil er von Übernahmegerüchten durch eine andere Firma gehört hat. Ein Intranet ist dann effektiv, wenn der Mitarbeiter mit dem Intranet sein Ziel erreichen kann. Die Ressourcen, die er dafür einsetzen muss, sei es durch psychische, physische oder zeitliche Beanspruchung, werden hier noch nicht beachtet. Dies wird in der ISO-Definition im Begriff Effizienz berücksichtigt. Effizienz bedeutet in diesem Sinne, dass der Mitarbeiter sein Ziel durch geringstmöglichen Aufwand erreichen kann. Der Aufwand muss dabei geringer sein als der gewonnene Nutzen. So dürfte ein Mitarbeiter das Intranet als wenig effizient empfinden, wenn er für das Auffinden der Telefonnummer länger braucht, als zu Fuss ins Büro seines Arbeitskollegen zwei Stockwerke tiefer.

Nach Gizycki (2002) besitzen Web-Benutzer Erwartungen in Bezug auf Effektivität und Effizienz. Werden diese von einem System erfüllt, kann damit gerechnet werden, dass der Benutzer mit dem System zufrieden ist.³ Diese Auffassung von Zufriedenheit wird im Bereich des Marketings als sog. Confirm/Disconfirm-Praradimga bezeichnet (Töpfer & Mann, 1999, S. 70). Dementsprechend kann Kundenzufriedenheit als das Ergebnis eines Vergleichsprozesses von Erwartungen (Soll-Leistung) und einem tatsächlich wahrgenommenen Leistungsniveau (Ist-Leistung) (Töpfer & Mann, 1999; Schneider, 2000, S. 39) verstanden werden.

Zufriedenheit wird in diesem Sinne in der vorliegenden Arbeit durch die Erfüllung von Erwartungen, die Benutzer in Bezug auf das Intranet haben, definiert. Dies bedeutet, dass das Messinstrument Zufriedenheit in erster Linie über die kognitive Komponente der Einstellung

³Einen sehr ähnlichen Ansatz verfolgen McKinney, Yoon und Zahedi (2002) zur Messung von Zufriedenheit von Internet-Kunden (*Web-Customer*). In ihrem Messmodell, welches sie in ihrer Studie bestätigen konnten, unterscheiden sie zwischen den zwei Ebenen Information (*Content*) und System. Die Autoren der Studie gehen dabei davon aus, dass bei den Internetbenutzern dann Zufriedenheit erreicht wird, wenn die Erwartungen sowohl in Bezug auf die Qualität der Informationen, die ihnen bereitgestellt werden, wie auch in Bezug auf die Qualität des Systems, mit dem sie interagieren müssen, erfüllt werden.

erfasst und sich die vorliegende Arbeit bewusst auf solche Variablen beschränkt, die auf diese Weise erfassbar sind. Variablen, die die Benutzerzufriedenheit ebenfalls beeinflussen können, wie bspw. Partizipation bei der Entwicklung von Informationssystemen (McKeen et al., 1996), Nutzung (Al-Gahtani & King, 1999) oder „Computerängstlichkeit“ (Harrison & Rainer, 1996), werden nicht berücksichtigt.⁴

2.2 Messinstrumente zur Erhebung der Benutzerzufriedenheit

Im folgenden Abschnitt werden umfragebasierte und validierte Messinstrumente vorgestellt, die zur Erfassung von End-Benutzerzufriedenheit entwickelt wurden. Die Darstellung erhebt keinen Anspruch auf Vollständigkeit.

Bailey und Pearson (1983) interviewten 32 mittlere Führungskräfte und entwickelten einen auf dem semantischen Differential basierenden 39 Items umfassenden Fragebogen zur Benutzerzufriedenheit mit Computern. Dem Fragebogen ist sein Alter von über zwanzig Jahren deutlich anzumerken. Ein Computersystem konnte damals nur in sehr engen Grenzen von Benutzern eingesetzt werden. Kontakte mit Programmierern und Datenverarbeitungsspezialisten gehörten damals zur Tagesordnung. Mehrere Items des Fragebogens von Bailey und Pearson (1983) erfassen denn auch die Zufriedenheit mit dem Personal der Datenverarbeitungsabteilung.

Mit der technischen Weiterentwicklung, insbesondere der Entwicklung von interaktiver Software nahm die (direkte) Abhängigkeit und Bedeutung von Programmierern für die Benutzer ab. Bei Doll und Torkzadeh (1988) finden sich bereits keine Items mehr zur Zufriedenheit mit dem Personal der Datenverarbeitungsabteilung. Gleichzeitig mit dem wachsenden Potential von Computern, wuchsen auch die Anforderungen an die Benutzerführung. Das Bereitstellen von benutzerfreundlichen Anwendungen gewann deshalb zunehmend an Bedeutung. Wurde leichte Benutzerführung (*Ease of Use*) bei Bailey und Pearson nur indirekt und am Rand tangiert, ist es im Fragebogen von Doll und Torkzadeh (1988) bereits ein elementarer Bestandteil zur Erfassung von Benutzerzufriedenheit. In ihrer

⁴ Die Beziehung zwischen Zufriedenheit und Nutzung ist ohnehin unklar und zweideutig. Al-Gahtani und King (1999) konnten zwar einen hohen positiven Zusammenhang feststellen, in ihren Untersuchungen mit Strukturgleichungsmodellen liess sich jedoch die Richtung des Zusammenhangs nicht bestimmen. Auch die Beziehung zwischen Zufriedenheit und Partizipation bei der Entwicklung ist nicht eindeutig. So fanden McKeen und Mitarbeiter (1996) sowohl Fälle, in denen hohe Partizipationsmöglichkeiten beim Entwicklungsprozess bei den Usern zu höherer Benutzerzufriedenheit führten, sie fanden aber auch Fällen mit hoher Partizipation und tiefer Zufriedenheit bzw. Fälle von geringer Partizipation und hoher Benutzerzufriedenheit.

explorativen Studie (N = 618) entwickelten Doll und Torkzadeh einen 12-Item langen und likert-skalierten Fragebogen zur Messung von Benutzerzufriedenheit mit spezifischen Computer-Applikationen. Nach einer Faktorenanalyse erhielten sie drei Faktoren mit einem Eigenwert von grösser Eins. Für am besten interpretierbar hielten sie jedoch eine erzwungene Fünf-Faktoren-Struktur, welche sie als Modellannahme beibehielten. Das Modell ist im Anhang A1 mit den dazugehörenden Items abgebildet. Nach Doll und Torkzadeh (1988) setzt sich Benutzerzufriedenheit aus den fünf Faktoren *Content*, *Accuracy*, *Format*, *Ease of Use* und *Timeliness* zusammen. Für die 12-Item-Skala erhielten sie ein Cronbach Alpha von .92. Harrison und Rainer (1996) bestätigten die Validität und Reliabilität des Messinstruments und konnten zeigen, dass sich das Messinstrument, welches ursprünglich für spezifische Computer-Applikationen konzipiert wurde, auch als generelles Messinstrument für Benutzerzufriedenheit mit Computern einsetzen lässt. In einer späteren Untersuchung von Doll und Mitarbeitern (2002) konnte die Stabilität des Modells bestätigt und dem Fragebogen eine relativ breite Verwendbarkeit in Bezug auf verschiedene Benutzergruppen, Applikationen und Hardware Plattformen attestiert werden.

Lewis (1995) entwickelte mit dem *IBM Computer Usability Satisfaction Questionnaires (CSUQ)* ebenfalls einen Fragebogen zur Messung der Benutzerzufriedenheit. Usability bzw. die Benutzerfreundlichkeit eines Systems wird von ihm dabei als das zentrale Element der Benutzerzufriedenheit erachtet. Lewis (1995) validierte den Fragebogen bei 377 relativ versierten Computerbenutzern. Die Hauptkomponentenanalyse (Varimax) ergab eine Drei-Faktoren-Struktur, deren Ladungen er als *System Usefulness*, *Information Quality* und *Interface Quality* interpretierte. Als Selektionskriterium für die Faktorladungen wählte Lewis .50, womit sich alle Items einem Faktor zuordnen liessen. Die Items des CSUQ's sind im Anhang A2 abgedruckt. Bei Betrachtung der rotierten Hauptkomponentenanalyse erweist sich das gewählte Selektionskriterium von .50 zum Teil als fraglich. So lädt bspw. Item 7, welches er zum ersten Faktor zählt, mit .49 auf dem zweiten Faktor. Auch andere Items weisen mittlere Ladungen auf anderen Faktoren auf. Für die von Lewis (1995) vorgeschlagene Zuteilung spricht, dass jeweils 50% der aufgeklärten Varianz auf einen Faktor fällt, womit nach Fürntratt (1969, zit. nach Bortz, 1999, S. 535) für die gewählte Zuordnung argumentiert werden könnte. Folgt man Fürntratt müsste allerdings auch das von Lewis eingesetzte Overall-Satisfaction-Item 19, dem ersten Faktor *System Usefulness* zugeteilt werden. Lewis führte die Faktorenanalyse mit dem Overall-Satisfaction-Item 19 durch, verwendet dieses jedoch unabhängig von den anderen drei Faktoren. Dies ist methodisch problematisch, da das

Item auf die Berechnungen der Faktorenanalyse einwirkt und damit die Höhen der Faktorenladungen der anderen Items beeinflusst.

Eine weitere interessante Skala stammt von Davis (1989). Sein Fokus lag bei der Entwicklung allerdings nicht auf dem Konstrukt Zufriedenheit, sondern auf dem Konstrukt Benutzerakzeptanz eines Informationssystems. Als zwei Determinanten von Akzeptanz fand er wahrgenommenen Nutzen (*Perceived Usefulness*) und wahrgenommene Leichtigkeit des Gebrauchs (*Perceived Ease of Use*). Die zwei konstruierten Skalen mit je sechs Items zeigen eine hohe Reliabilität von .98 für *Perceived Usefulness* und .94 für *Perceived Ease of Use*. Die Items der zwei Skalen sind im Anhang A3 aufgeführt. Phleps und Mok (1999) untersuchten Variablen, die bei der Einführung eines Intranets Auswirkungen auf die Zufriedenheit der Benutzer haben könnten. Die höchsten Zusammenhänge fanden sie für die zwei Skalen von Davis. Sie kommen zum Schluss, dass *Perceived Usefulness* und *Perceived Ease of Use* als zwei primäre Faktoren von Benutzerzufriedenheit betrachtet werden sollten. Ihre Ergebnisse müssen jedoch mit Vorsicht interpretiert werden. Phleps und Mok (1999) operationalisierten in ihrer Studie Benutzerzufriedenheit mit Items wie bspw. die Absicht der Probanden, das Intranet für die Arbeit zu nutzen, ob die Einführung eines Intranets sich positiv auf die Firma auswirken würde oder ob die Probanden denken, ihre Arbeit besser mit oder ohne einem Intranet erledigen zu können. Die Items in ihrem Fragebogen ähneln damit zum Teil der *Perceived Usefulness*-Skala von Davis (1989), was die Aussage über den Zusammenhang verzerrt.

Weitere Autoren entwickelten Zufriedenheitsskalen für spezifische Bereiche wie Online-Shopping (McKinney et al., 2002), Firmen-Websites (Muyelle et al., 2004), Business-to-Employee⁵ (Huang et al., 2004) oder Informationssysteme in Kleinunternehmen (Palvia, 1996). Die Autoren beziehen sich zum Teil direkt auf die oben vorgestellten Messinstrumente und ihre Skalen sind so im Kern meist auf diese zurückführen. Die Unterschiede zeigen sich in Formulierungen, die besser auf den jeweiligen Kontext angepasst sind oder Ergänzungen von spezifischen Items wie bspw. im Zusammenhang mit Preis, Sicherheit oder Lieferbedingungen.

⁵ *Business-to-employee (B2E)* steht für die (elektronische) Kommunikationsbeziehung zwischen einem Unternehmen und seinen Mitarbeitern. In der Absicht, die Motivation und Zufriedenheit der Mitarbeiter zu fördern wird diese Form der Kommunikation von einigen Firmen genutzt, um den Angestellten attraktive Dienstleistungen anzubieten (bspw. Vergünstigungen bei Partnerfirmen wie Autoverleiher, Hotelketten, Buchhändler, Airlines usw.).

Vergleicht man die Items bzw. die verschiedenen Skalen der genannten Messinstrumente, lassen sich zwei Schwerpunkte heraus kristallisieren. Der erste Schwerpunkt umfasst Items, welche sich auf die angebotenen Informationen des Systems beziehen. Der zweite Schwerpunkt schliesst Items ein, die die Gestaltung der Mensch-Maschine-Schnittstelle betreffen. Neuere Arbeiten deuten zudem darauf hin, dass es durchaus Sinn machen kann, neben generellen und allgemeingültigen Items (Bailey & Pearson, 1983; Doll & Torkzadeh, 1988; Davis, 1989; Lewis, 1995), auch kontextspezifische Items zu berücksichtigen (McKinney et al., 2002; Muyelle et al., 2004; Huang et al., 2004). Im Folgenden werden diese drei Ebenen (Qualität von Information, Qualität der Mensch-Maschine-Schnittstelle sowie die Nutzung des Intranets in Organisationen) deshalb in Bezug auf die Konstruktion des Fragebogens näher ergründet.

2.3 Qualität von Informationen aus Sicht der Benutzer

Einen Ansatz für die Bestimmung hoher Qualität von Informationen, zeigt eine Studie von Wang und Strong (1996). Darin untersuchten sie, was Datenqualität⁶ in einem geschäftlichen Kontext aus Benutzersicht bedeutet. Sie gingen dabei implizit davon aus, dass Qualität von Daten sich ähnlich bestimmen lässt wie die Qualität von anderen Produkten. In einer ersten Umfrage sammelten sie Attribute von Datenqualität. In einer zweiten Umfrage liessen sie diese insgesamt 118 gesammelten Attribute nach ihrer Wichtigkeit einstufen. Die Attribute, die als die zwanzig wichtigsten bewertet wurden, liessen sie anschliessend von Probanden in Gruppen zusammenfassen (*Open Card Sorting*⁷). Auf Basis dieser Ergebnisse bestimmten sie vier Ebenen, die sie *Intrinsic Data Quality*, *Contextual Data Quality*, *Representational Data Quality* und *Accessibility Data Quality* benennen. Abbildung 2.2 veranschaulicht ihr Modell und gibt die Attribute an, welche von Benutzern als die wichtigsten bezeichnet wurden.

⁶ Die Begriffe Daten und Informationen werden in dieser Arbeit als synonym betrachtet.

⁷ Bei der *Open Card Sorting* Methode lässt man verschiedene Benutzer mit Begriffen versehene Kärtchen zu (offenen) Gruppen ordnen. Mittels der Clusteranalyse wird anschliessend versucht, möglichst übereinstimmende Kategorien zu bilden und diese sinnvoll zu benennen (*Wording*). Bei *Closed Card Sorting* werden Kategorien bereits vorgegeben und die Benutzer müssen die Kärtchen diesen zuordnen. Die Methode wird auch gerne zur Konzeption von Informationsarchitekturen beigezogen. *Card Sorting* lässt sich äusserst günstig realisieren und nach einer Studie von Tullis und Wood (2004) können bereits bei Stichproben von 20 bis 30 Probanden aussagekräftige Resultate erwartet werden.

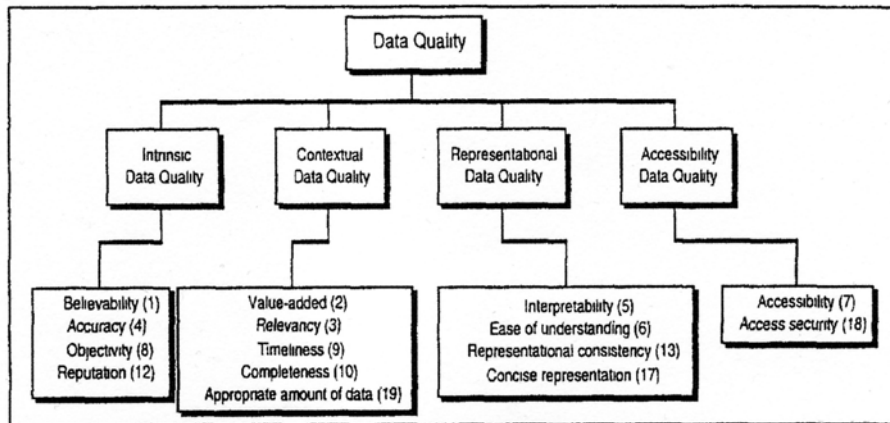


Abbildung 2.2. Modell für Datenqualität aus Wang und Strong (1996, S. 20). Die Zahl in den Klammern gibt Auskunft über den Rang, den die Attribute bei der Einstufung der Wichtigkeit durch die Benutzer im Schnitt erreichte (N = 355). *Relevancy* wurde so bspw. im Schnitt als drittwichtigstes Attribut gewertet.

„Intrinsische“ Datenqualität umfasst Attribute, die in den Daten an sich liegen wie bspw. Genauigkeit, Objektivität, Glaubwürdigkeit. „Kontextuelle“ Datenqualität liegt vor, wenn die Daten abrufbar sind, wie sie in einem Arbeitsprozess benötigt werden. Nach Wang und Strong handelt es sich hierbei um Attribute wie Relevanz, Mehrwert erbringend, Rechtzeitigkeit oder Angemessenheit.⁸ „Gegenständliche/Darstellungs-Qualität“ der Daten schließt Attribute mit ein wie leichte Interpretierbarkeit- oder Verständlichkeit. Das Intranet bietet sich als Medium an, Sachverhalte neben Text auch mit Bildern oder im geschäftlichen Kontext vor allem mit Diagrammen zu ergänzen. Um die *Representational Quality* zu erhöhen, können Befunde aus der Kognitions- und Wahrnehmungspsychologie wertvolle Inputs liefern. So zeigt bspw. Schnotz (2002), wie die Lesbarkeit von Diagrammen erhöht werden kann, wenn man sich bei der Gestaltung an die Gestaltgesetze⁹ hält und damit Prozesse der Wahrnehmungsorganisation berücksichtigt. Zudem zeigt er auf, dass sich nicht alle Diagrammtypen gleich gut eignen, um gewisse Sachverhalte dem Leser klar darzustellen (Schnotz, 2002, S. 77ff.). Die Wahl eines Diagrammtyps (Kreis-, Balken-, Säulen-, Kurven- oder Punktediagramm) sollte sich deshalb stets nach der beabsichtigten Mitteilungsfunktion richten.

⁸ In der Direktion für Entwicklung und Zusammenarbeit DEZA beschwerten sich bspw. die Koordinationsbüros im Ausland, dass sich die Mitteilungen im Intranet oft nur auf die Zentrale bezögen und für sie irrelevant wären. Dies war für sie insofern störend, da sie zuerst immer wichtige von unwichtigen Mitteilungen trennen mussten. In diesem Sinne war also keine hohe kontextuelle Datenqualität gegeben. Um das Problem zu beheben, wurde ein Filter eingerichtet, welcher je nach eingestelltem Benutzerprofil Mitteilungen im Intranet erschienen liess oder nicht. Die kontextuelle Datenqualität der Mitteilungen konnte so erheblich erhöht werden.

⁹ Die Gestaltgesetze beschreiben einen Satz von Regeln, die zu erklären versuchen, welche Wahrnehmungen zu erwarten sind, wenn bestimmte Reizbedingungen gegeben sind. Diese gestaltheoretische Erklärung der Wahrnehmungsorganisation geht auf die sogenannten Gestaltpsychologen um Max Wertheimer in den 20er Jahren zurück. Für eine übersichtliche Darstellung mit Illustrationen der Gestaltgesetze sei auf Goldstein (1997), für vertiefende Informationen auf Katz (1969) oder die Originaltexte von Wertheimer (1921, 1923) verwiesen.

Studien weisen darauf hin, dass illustrierte Texte oft bessere Behaltensleistungen hervorbringen als nicht-illustrierte Texte (Schnotz & Bannert, 1999). Dies wird meist durch die duale Kognitionstheorie von Paivio (1986) erklärt. Sie nimmt an, dass Texte und Bilder auf unterschiedlichen kognitiven Subsystemen verarbeitet werden. Während verbale Information nur im verbalen System enkodiert wird, werden Bilder sowohl im verbalen wie auch in einem imaginalen System enkodiert. Bessere Behaltensleistungen werden deshalb oft damit begründet, dass die Doppelcodierung gegenüber der einfachen Codierung überlegen ist und Texte mit Bildern generell zu elaborierteren Wissenstrukturen führen. Neuere Studien wie die von Schnotz und Bannert (1999) stellen die duale Kodierungstheorie in Frage, weil sie unter anderem nicht erklären kann, wieso verschiedene (informationsäquivalente) Abbildungen unterschiedliche Effekte erzielen. In einem Experiment konnten sie belegen, wie zwei informationsäquivalente Abbildungen in einem Fall die Konstruktion des mentalen Modells unterstützte, im anderen Fall aber mit der mentalen Modellkonstruktion interferierten. Daraus kann geschlussfolgert werden, dass die Vermischung von Text und Bild (Grafik) zwar meist vielversprechend ist, nicht in jedem Fall aber zu einer Verbesserung der Verarbeitungseffizienz führt. Eine schlecht umgesetzte Grafik kann einen Mitarbeitenden beim Aufnehmen eines Sachverhalts unter Umständen also mehr ‚verwirren‘ als unterstützen.

Die letzte Ebene bei Wang und Strong (1996) ist die „Zugänglichkeits-Qualität“ der Daten. Hier geht es nach den Autoren primär darum, wie gut der Zugang zu den Daten gewährleistet ist: Sind die Daten bspw. mit Hilfe einer Suchmaschine leicht auffindbar oder sind sie aufgrund einer klaren Informationsstruktur gut zugänglich. Gemeint ist zudem, auf welche Weise auf die Daten zugegriffen werden kann (PC am Arbeitsplatz, Laptop unterwegs, Mobiltelefone usw.).¹⁰

Nach Wang und Strong (1996, S. 22) kann hohe Datenqualität wie folgt zusammengefasst werden: „High-quality data should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the data consumer“. Für die Fragebogenkonstruktion wird davon ausgegangen, dass die Benutzer im Intranet grundsätzlich eine hohe Datenqualität erwarten, woraus sich mehrere Items ableiten lassen:

¹⁰ Im Kontext des Internets taucht der Begriff „*Accessibilty*“ oft auch in Diskussionen zum Internetzugang von Behinderten und älteren Menschen auf. Es geht dabei darum, Internetseiten (technisch) so zu gestalten, dass z.B. auch Blinde oder sehbehinderte Menschen eine Internetseite nutzen können. Die *Web Content Accessibility Guidelines* des World Wide Web Consortiums (1999) legen Richtlinien fest, wie möglichst allen Menschen der Zugang ins Internet gewährt werden kann. In Deutschland müssen Internet-Seiten der Bundesverwaltung per Gesetz in diesem Sinne barrierefrei sein. Im Kontext von Intranets wird dieser Aspekt in der Literatur nicht thematisiert, was natürlich auch auf die praktisch nicht vorhandene Integration von behinderten Menschen in der Arbeitswelt zurückzuführen ist.

(1) Das Intranet stellt für mich arbeitsrelevante Inhalte zur Verfügung, (2) Die Informationen auf dem Intranet sind aktuell, (3) Die Art und Weise, wie Informationen auf dem Intranet vermittelt werden, ist klar und gut verständlich, (4) Bei Bedarf kann ich jederzeit auf das Intranet zugreifen, (5) Informationen, die ich aus dem Intranet beziehe, kann ich ohne Modifikationen für meine Arbeit verwenden, (6) Ich kann mich auf die Informationen im Intranet verlassen und darauf vertrauen, dass sie richtig sind.

2.4 Benutzerfreundliche Mensch-Maschine Schnittstelle

Das Intranet unterstützt Arbeitende bei der Erfüllung von bestimmten Tätigkeiten. Dies kann vom simplen Abfragen einer Telefonnummer bis zu komplexeren Anwendungen wie der Verwaltung von Personaldaten reichen. Zu diesem Zweck muss der Arbeitende mit dem Computer in einen Dialog treten. Der Inhalt dieses Dialogs besteht im Wesentlichen im Aufsuchen oder Erstellen, in manchen Fällen auch im Kontrollieren oder Mutieren von Informationen. Zu einer eigentlichen Primäraufgabe (Aufgabe, für welche das System entwickelt wurde) kommt so die Dialogführung als Sekundäraufgabe hinzu (Spinas, 1987, S. 8). Rauterberg, Spinas, Strohm, Ulich und Waeber (1994) unterscheiden dabei zwischen Anwendungsoperationen, welche ein Anwendungsobjekt direkt verändern (z.B. ein Textdokument) und Dialogoperationen, welche nicht zur unmittelbaren Aufgabenbearbeitung beitragen (z.B. die Veränderung von Fenstern, wie Öffnen, Schliessen, Verkleinern usw.). Durch die Dialogführung entstehen dementsprechend zusätzliche kognitive Anforderungen. Diese sollten möglichst gering sein, da sonst die Gefahr besteht, dass die Beherrschung eines Dialogsystems zu einer eigenständigen Aufgabe wird.

Für einen Benutzer sollte es möglich sein, sich eine adäquate mentale Repräsentation des Dialogsystems zu erstellen (Spinas, 1987). Das heisst, dass sich ein Benutzer über Inhalt (Informationen und Funktionen) und Organisation (Ordnungskriterien und Zugriffswege) ein gedankliches Modell, eine Art „geistige Landkarte“ machen kann. Nach Nievergelt (1982, S. 203, zit. nach Baitsch et al., 1989, S. 78) sollte es für einen Benutzer jederzeit möglich sein, die fünf Fragen: „Wo bin ich? Was kann ich hier tun? Wie kam ich hierhin? Wo kann ich hin, und wie komme ich dorthin?“ zu beantworten. Damit ein Dialogsystem einen Benutzer in seinen Arbeitsprozessen optimal unterstützt, sollte es deshalb möglichst gut an die Prozesse der menschlichen Informationsverarbeitung (Wahrnehmung, Gedächtnis, Denken, Handeln) angepasst sein (Spinas, 1987; Baitsch et al., 1989; Rauterberg et al., 1994; Herczeg, 1994; Preece et al., 1999; Rosson & Carroll, 2002). Diese Anpassung der Benutzerschnittstelle auf

den Menschen wird dabei meist unter dem Begriff „Benutzerfreundlichkeit“ (engl. *Usability*) behandelt.¹¹

In der Literatur wurde verschiedentlich versucht, allgemein gültige Richtlinien und Standards für benutzerfreundliche Schnittstellen festzulegen. Es finden sich dabei grob unterteilt zwei Arten von Publikationen, zum Einen Listen mit Gestalthinweisen, zum Anderen generelle Gestaltungsprinzipien. Die erste Gruppe gibt Anleitungen nach Art von Gebrauchsanweisungen, die Hinweise reichen dabei von allgemeinen Aussagen bis hin zu detaillierten Anweisungen eines bestimmten Gestaltungsaspekts (z.B. wo Links platziert werden sollten, wie man Fehlermeldungen formulieren sollte oder aus wievielen Kategorien eine Informationsstruktur bestehen darf, usw.). So erarbeiteten Nielsen und Tahir (2002) 113 Regeln zur Gestaltung von benutzerfreundlichen Internetseiten und bei Lynch und Horton (2002) finden sich Anleitungen zum Bau eines benutzerfreundlichen Internetauftritts. Die Regeln und Hinweise werden nur selten vertieft begründet oder gar empirisch belegt, der Richtigkeitsanspruch gründet so hauptsächlich in der intuitiven Plausibilität und der jeweiligen Fachautorität der Autoren.

Zur zweiten Gruppe kann die erwähnte ISO-Norm 9241, diesmal Part 10, gezählt werden. Darin werden sieben Gestaltungsgrundsätze für die Dialogschnittstelle beschrieben. Genannt werden Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlerrobustheit, Individualisierbarkeit und Lernförderlichkeit (ISO-Norm 9241 Part 10, zit. nach Herczeg, 1994, S. 105). Ein anderes, relativ bekanntes Beispiel stammt von Shneiderman (1998), welcher acht ‚goldene‘ Regeln des Dialogdesigns formulierte. Im Kern sind sie mit den Grundsätzen der ISO-Norm 9241 Part 10 vergleichbar, er beschreibt sie allerdings weniger abstrakt, weshalb sie leichter anzuwenden sind. Sie lauten: *1. strive for consistency; 2. enable frequent users to use shortcuts; 3. offer informative feedback; 4. design dialog to yield closure; 5. offer simple error handling; 6. permit easy reversal of actions; 7. support internal locus of control; 8. reduce short-term memory load* (Shneiderman, 1998, S. 124–150).

Das kommende Modell stammt aus dem Lehrstuhl für Arbeits- und Organisationspsychologie der ETH Zürich und lehnt sich stark an ein Modell für

¹¹ Mittlerweile werden im Zusammenhang mit der Schnittstelle Mensch-Maschine mehrere Richtungen bzw. Bereiche unterschieden (wie *Human Factors*, *User Centerd Desgin*, *Software Ergonomy*, *Usability Engineering* oder *Human Computer Interaction (HCI)*). Diese Bereiche setzen unterschiedliche Schwerpunkte bei der Interaktionsgestaltung, überschneiden sich jedoch zum Teil stark. Aus einzelnen Bereichen entwickelten sich auch verschiedene praxisorientierte Methoden zur Gestaltung von benutzergerechten Schnittstellen. Auf die Darstellung einzelner Methoden oder der Unterschiede zwischen den Bereichen wird hier jedoch verzichtet.

benutzerorientierte Dialoggestaltung von Ulich (siehe Ulich 2001, S. 369) an, welches als handlungspsychologisch begründbar und empirisch abgesichert gilt (Rauterberg et al., 1994, S. 16). Es wird folgend näher betrachtet, wobei sich viele Überschneidungen zu den beiden oben genannten Gestaltungsgrundsätzen bzw. Kriterien ergeben und deshalb zuvor auf eine Erläuterung verzichtet wurde.

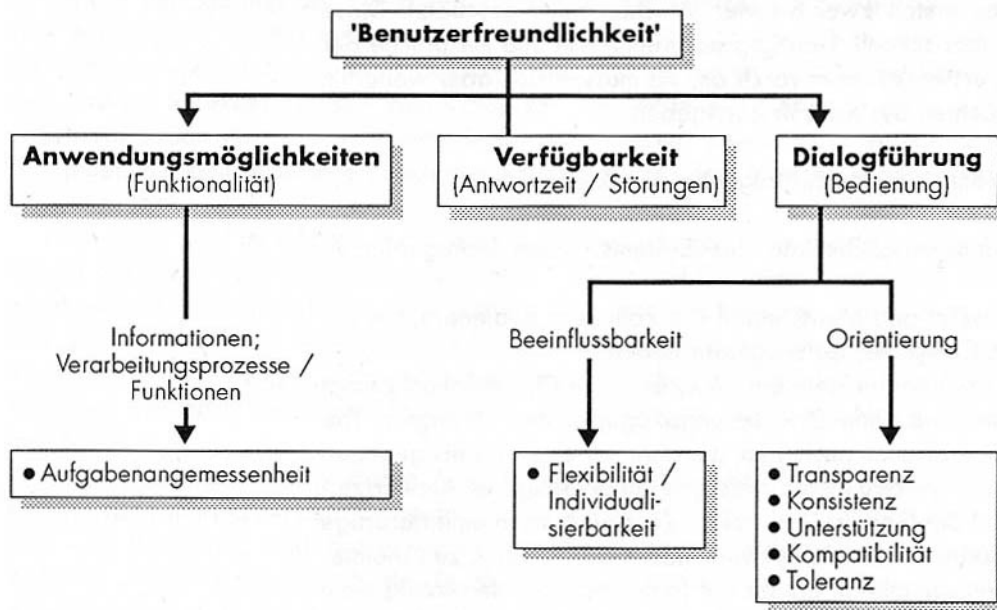


Abbildung 2.3. Aspekte und Kriterien der Benutzerfreundlichkeit aus Baitsch und Mitarbeiter (1989, S. 78).

Unter Anwendungsmöglichkeiten (Funktionalität) werden die vom System zur Verfügung gestellten Verarbeitungsprozesse und Informationen begriffen, welche den Benutzer bei der Erledigung seiner Aufgaben unterstützen sollen. Der Aspekt der Aufgabenangemessenheit wurde bereits unter dem Thema Datenqualität angeschnitten und betont die Wichtigkeit, dass Informationen und Funktionen des Systems auf die Arbeitshandlungen der Benutzer ausgerichtet sind.

Häufige Störungen oder gar Systemausfälle können die Verfügbarkeit des Computers einschränken. Die Benutzer sind bei Störungen dem Computer hilflos ausgesetzt, meist können sie weder die Ursachen durchschauen noch den weiteren Verlauf voraussehen oder gar beeinflussen. Besonders wenn Benutzer für ihre Arbeit auf den Computer angewiesen sind, können solche Situationen als enormer Stress erlebt werden. Die Zuverlässigkeit eines Systems kann deshalb als eine Grundvoraussetzung für den Einsatz an einem Arbeitsplatz

bezeichnet werden. Ähnliches gilt auch für die Antwortzeiten des Computers. Lange oder stark variierende und dadurch unberechenbare Antwortzeiten stören den Gedankenfluss der Benutzer und zwingen ihm einen unregelmässigen Arbeitsrhythmus auf (Baitsch et al., 1989; Rauterberg et al., 1994). Bei den Antwortzeiten gilt es dabei, einen optimalen Mittelweg zu finden, denn sowohl ‚zu schnelle‘, als auch ‚zu langsame‘ Antwortzeiten können bei Benutzern zu Belastungen führen (Kohlisch & Kuhmann, 1997). Ein Benutzer sollte so einen Maskenwechsel jeweils noch gut erkennen können, er sollte jedoch auch das Gefühl haben, mit dem System liesse sich zügig arbeiten.

Mit Flexibilität ist gemeint, dass die Steuerung eines Dialogablaufs soweit wie möglich einem Benutzer überlassen werden sollte. Dies ermöglicht es ihm, unterschiedliche Vorgehensweisen zu wählen und gibt ihm Freiheiten bei der Planung seiner Arbeitshandlungen. Einem Benutzer sollte es zudem möglich sein, individuelle Anpassungen vorzunehmen und eine Benutzeroberfläche nach persönlichen Bedürfnissen einrichten zu können (Individualisierbarkeit) (Baitsch et al., 1989; Rauterberg et al., 1994). Der Intranet Report 2003 (Stimmt AG, 2003) weist zu diesem Aspekt darauf hin, dass zwar in vielen Intranets Möglichkeiten zu Personalisierungen angeboten werden, diese aber nur von einem relativ kleinen Anteil von Personen genutzt werden. Daraus lässt sich folgern, dass die Mehrheit der Personen mit Standardeinstellungen arbeitet und keine Anpassungen vornimmt. Die Gründe dafür können unterschiedlicher Art sein: Die Benutzer wissen nichts von diesen Möglichkeiten, sie nehmen sich keine Zeit, um die Einstellungen vorzunehmen, die Einstellung vorzunehmen ist zu kompliziert oder die Benutzer sehen keinen Nutzen darin. Personalisierungen scheinen unter diesem Gesichtspunkt zwar ein interessantes Mittel, ein System benutzerfreundlicher zu gestalten, sie scheinen aber nicht per se die wahrgenommene Benutzerfreundlichkeit zu erhöhen.

Die nächsten Punkte betreffen die Orientierung: „Die Transparenz eines Dialogsystems soll dem Benutzer die Bildung eines Struktur- und Prozessmodells des Systems im Gedächtnis erleichtern, was ihm die notwendigen Orientierungsgrundlagen für die Benutzung bietet“ (Rauterberg et al., 1994, S. 16). Eine gute Möglichkeit, in einem Intranet die Transparenz zu erhöhen, ist die sog. Brotkrummennavigation. Diese gibt dem Benutzer stets Auskunft darüber, wo er sich gerade befindet und auf welchem Weg er dorthin gelangt ist. Auch eine klar (nach Gestaltgesetzen) strukturierte und damit verständlichere Benutzeroberfläche hilft einem Benutzer, eine innere Landkarte des Intranets zu bilden. Ein anderes Beispiel für Transparenz betrifft die Eingabe von Befehlen. So sollte ein Benutzer

erkennen können, ob ein eingegebener Befehl behandelt wird oder das System noch weitere Eingaben erwartet. Transparenz wird als ein zentrales Kriterium bezeichnet, es ist allerdings schwierig, es isoliert von den anderen Kriterien zu betrachten. Zudem hängt es auch stark von der Erfahrung eines Benutzers ab (Rauterberg et al., 1994).

Mit Unterstützung sind verschiedene Formen von Hilfsfunktionen gemeint, die einem Benutzer in Problemsituationen einen Ausweg zeigen können. Dies kann bspw. ein elektronisch zugängliches Handbuch, eine informative Fehlermeldung, eine Rubrik mit oft gestellten Fragen (FAQ) oder andere Formen der Hilfe sein. Wichtig dabei ist, dass diese Hilfe von Benutzern jederzeit abgerufen werden kann und deren Inhalte möglichst kontextsensitiv bzw. problemspezifisch sind. Zudem sollte die Hilfe nicht in einem Computerjargon, sondern in der Sprache der Benutzer verfasst sein (Baitsch et al., 1989; Rauterberg et al., 1994; Ulich 2001).

Das Kriterium der Konsistenz bezieht sich auf die Berechenbarkeit des Systemverhaltens. Beispielsweise sollten gewisse Funktionstasten oder Menüoptionen über ein System hinweg immer einheitlich sein. Ein konsistent-regelhafter Aufbau von Dialogstruktur, Semantik und Syntax der Operationen entlastet das Gedächtnis des Benutzers, da er bspw. gelernte Kommandos im ganzen System verwenden kann (Baitsch et al., 1989; Rauterberg et al., 1994).

Mit dem Kriterium der Kompatibilität versucht man bei der Dialoggestaltung äussere Analogien oder Metaphern zu nutzen. Bei den Benutzern soll damit ein unmittelbares Verständnis erreicht werden. Aus kognitiver Sicht führt dies beim Benutzer zu einer Reduktion von geistigen Transformationsschritten, womit das Kurzzeitgedächtnis entlastet werden kann (Baitsch et al., 1989; Rauterberg et al., 1994). Ulich (2001, S. 371) unterscheidet dabei zwischen Sprach- und Darstellungskompatibilität.

Das letzte Kriterium Toleranz meint den Verzicht auf genaue Eingabeerfordernisse. Ein elektronisches Telefonbuch sollte so bei der Eingabe von ‚Meier‘ auch die Meyers und Maiers aufführen. Ein weiterer Aspekt von Toleranz betrifft die Möglichkeit, Operationen ohne Verlust von Daten rückgängig machen zu können (UNDO-Funktion). Durch UNDO-Funktionen wird auch erreicht, dass Benutzer eine grössere Sicherheit im Umgang mit dem Computer erhalten (Baitsch et al., 1989; Rauterberg et al., 1994).

Die genannten Kriterien können als wichtige Grundlage für eine benutzerfreundliche Dialoggestaltung betrachtet werden. Wie wichtig die einzelnen Kriterien für die Dialoggestaltung sind, hängt immer auch von der konkreten Anwendungssituation ab (Baitsch et al., 1989).

In einer Studie untersuchte die Stimmt AG¹² (2003) in einem heuristischen Ansatz die Benutzeroberflächen von 22 Intranets anhand von 24 Usability-Kriterien. Die Kriterien bezogen sich auf die Bereiche Navigation, Inhalt, Technisches, Screen Design und Suche. Es zeigte sich, dass nur etwa die Hälfte der 24 Usability-Kriterien durchschnittlich erfüllt wurde. Intranets, welche in den letzten Jahren neu gestaltet wurden, schnitten dabei tendenziell etwas besser ab. Bei 13 der 22 Intranets war kein Organisationsschema bzw. einheitliches Konzept zur Organisation der Inhalte ersichtlich (bspw. Firmenzentrierung, Aufgaben-, Themen-, Benutzergruppenorientierung oder eine Kombination von diesen). Zehn Intranets besaßen zudem keine eindeutige Informationsarchitektur. In mehreren Intranets wurde dem Benutzer die Orientierung und die Bildung eines mentalen Modells erschwert, weil die aus der Kognitionspsychologie bekannte Grenze von 7 ± 2 aufnehmbaren Informationseinheiten nicht befolgt wurde (Miller, 1956). Bei einem Drittel der Intranets war es für einen Benutzer nicht ersichtlich, wo er sich gerade befindet. Bei den Usability-Kriterien zum Inhalt wurde bei mehreren Intranets bemängelt, dass Neuerungen nicht ersichtlich gemacht werden. Benutzer können so nie wissen, ob es neue Dateien oder Seiten auf dem Intranet gibt. Mehrere Intranets erfüllten auch nicht das Kriterium, Inhalte mit einem Aktualisierungsdatum zu versehen oder jeweils eine Kontaktperson anzugeben. Die Angabe einer Kontaktperson ist aus Sicht eines Benutzers deshalb wünschenswert, da sie ihm Auskunft darüber gibt, von wem eine Information stammt und ihm zudem ermöglicht, bei Fragen direkt mit einer verantwortlichen Person in Kontakt treten zu können. Als Fazit ihrer Untersuchung zieht die Stimmt AG (2003, S. 59), dass viele Intranets selbst elementare Regeln der Benutzerfreundlichkeit missachten. Die Gründe hierfür können unterschiedlich sein. Ein Grund dürfte darin liegen, dass Intranets vielfach organisch gewachsen sind. Besonders Zusammenschlüsse von mehreren kleinen Intranets zu einem Grossen sind in der Regel äusserst komplex und schwierig umzusetzen. Ein anderer Grund ist, dass Funktionalität und Aussehen oft von den technischen

¹² Bei der Stimmt AG handelt es sich um eine in Zürich ansässige Firma aus der Privatwirtschaft. Intranets und Mitarbeiterportale gehören seit einigen Jahren zu einem ihrer Kompetenzfelder. Der Einbezug von Studien aus der Privatwirtschaft ist aus wissenschaftlicher Sicht grundsätzlich heikel. An der Studie arbeitete jedoch auch Bargas-Avila, der Betreuer der vorliegenden Arbeit, mit, weshalb diesbezügliche Bedenken ausgeräumt werden konnten.

Möglichkeiten bestimmt werden und nicht Kriterien der Benutzerfreundlichkeit im Vordergrund stehen.

Für den Fragebogen wird angenommen, dass die Mitarbeiter eines Unternehmens erwarten, dass ihnen für die Erledigungen ihrer Arbeit ein benutzerfreundliches und damit effizientes Intranet bereitgestellt wird. Entsprechend dürften sie unzufrieden reagieren, wenn sie das Intranet als ein benutzerunfreundliches Arbeitsinstrument wahrnehmen. Als eine gute Quelle für die Itemgenerierung bieten sich hier die Items des CSUQ's von Lewis (1995) an.

2.5 Das Intranet und seine Nutzer

Der Fokus der vorgestellten Messinstrumente lag auf der Entwicklung eines allgemeingültigen Fragebogens unabhängig von der jeweiligen Computeranwendung bzw. vom jeweiligen Computersystem. Es ist jedoch anzunehmen, dass Benutzer kontext- bzw. systemspezifische Erwartungen haben, die sich aus dem Gebrauch und den Möglichkeiten eines Informationssystems, hier dem Intranet, ergeben. Im Folgenden wird deshalb geklärt, ob es intranetspezifische Punkte gibt, die der Fragebogen enthalten sollte.

Es ist schwierig, allgemein gültige Aussagen über Verbreitung und Verwendung von Intranets zu machen. Zahlen und Fakten aus systematisch durchgeführten Untersuchungen sind eher die Ausnahme. Ein Grund dafür dürfte auch in der Auffassung vieler Unternehmen liegen das Intranet als Wettbewerbsfaktor zu verstehen und Auskünfte, wenn überhaupt, nur unter Vorbehalten zu erteilen. Der Untersuchungsgegenstand Intranet ist zudem an eine Technologie gekoppelt, die sich rasant entwickelt; Befunde veralten daher schneller als in anderen Bereichen. Der folgende Abschnitt versucht trotzdem, einige Charakteristiken des Intranets in Organisationen zu vermitteln. Dies geschieht in erster Linie anhand von Studien der Stimmt AG (2003, 2005), von Hoffmann (2001) und von Meier (2002), die alle den Untersuchungsgegenstand Intranet nicht exemplarisch, sondern über mehrere Unternehmen hinweg untersuchten, die beiden letztgenannten in erster Linie in Bezug auf das Intranet als Instrument für die Interne Kommunikation.

Grundsätzlich können alle Bereiche eines Unternehmens zur Unterstützung von betrieblichen Anwendungen auf das Intranet zurückgreifen (Kortzfleisch, 1997). In der Praxis werden so je nach Organisationseinheit und Bedürfnis verschiedene Intranet-Anwendungen realisiert. Vielfach publizieren Unternehmen auf dem Intranet zu Beginn allgemeine Informationen wie Geschäftsberichte, Pressemitteilungen und Informationen, die für alle

Mitarbeiter gedacht sind wie Adressen, Verzeichnisse, Archive, Dokumentationen oder Arbeitshilfen (Hoffmann, 2001, S. 90). In 95 von Hoffmann (2001, S. 160) befragten deutschen Unternehmen wird das Intranet neben der Mitarbeiterkommunikation bevorzugt in den Bereichen Verwaltung, Personalwesen sowie Marketing und Vertrieb eingesetzt.

Die Stimmt AG (2003) untersuchte Intranets von mehreren Schweizer Firmen. Sie interviewte dazu 27 Intranetverantwortliche, befragte über tausend Benutzer aus 14 Unternehmen mit dem CSUQ (Lewis, 1995) und analysierte 22 Intranets in Bezug auf Usability-Kriterien (siehe oben). Ein Jahr später führte die Stimmt AG (2005)¹³ eine ähnliche Untersuchung durch und interviewte dazu 42 Intranetverantwortliche und befragte 900 Benutzer aus 11 Unternehmen mit dem CSUQ. Die Ergebnisse publizierten sie in sog. Intranet Reports. Nach dem Intranet Report 2004 (Stimmt AG, 2005) wird das Intranet von den Benutzern vorwiegend für Informationszwecke aufgesucht. Noch relativ gering ist der Anteil, in dem es für Kommunikationszwecke (Interaktion) oder Workflows (vom Intranet gesteuerte Interaktion) genutzt wird. Das elektronische Telefonbuch stellt nach Ansicht der meisten Intranetmanager eine Killer-Applikation¹⁴ für das firmeneigene Intranet dar. Als weitere Killer-Applikationen werden die News (Interne Kommunikation), die Suche, Anwendungen im Personalwesen und Bestellapplikationen genannt. Oft eingesetzt, aber von Intranetmanagern nicht primär als Killer-Applikationen bezeichnet, werden Marktplatz/Schwarzes Brett, Dokumenten-Management, Wörterbücher/Nachschlagewerke, Offene Stellen, Weiterbildungsangebote, Wissensdatenbanken, Operationelle Applikationen, Projektmanagement, Veranstaltungskalender und Diskussionsforen (Stimmt AG, 2005).

Dass bereits viele Unternehmen das Intranet als Instrument der Internen Kommunikation nutzen, hält auch Meier (2002) fest. Für seine Dissertation interviewte er Kommunikationsverantwortliche von 165 Unternehmen in Bezug auf die Interne Kommunikation. 77% der von Meier (2002) befragten Unternehmen gaben dabei an, für die Interne Kommunikation auf das Intranet zurückzugreifen.¹⁵ Seine Umfrage zeigte zudem, dass der Internen Kommunikation in der Praxis eine grosse Bedeutung beigemessen wird. Praktisch alle befragten Kommunikationsfachleute waren dabei der Meinung, dass sich gute Interne Kommunikation auf die Motivation und Leistungsbereitschaft des Personals auswirkt.

¹³ Die Studie wurde 2004 durchgeführt, jedoch 2005 publiziert.

¹⁴ Als Killerapplikationen werden Anwendungen bezeichnet, die einer Technologie zum Durchbruch verhelfen. Oder anders formuliert, hätte es eine Technologie sehr schwer von den Benutzern angenommen zu werden, wenn es diese Anwendung nicht anbieten würde.

¹⁵ Die Dissertation von Meier (2002) verfasste er im Jahr 2000. Es ist realistisch anzunehmen, dass sich der Anteil von Unternehmen, die das Intranet für ihre Interne Kommunikation nutzen, in den letzten 6 Jahren noch angestiegen ist.

Als Motivationsgrund, das Intranet für die Interne Kommunikation zu benutzen, nennt die Studie von Hoffmann (2001, S. 179) die Verbesserungen und Erleichterungen, die diese Technologie mit sich bringt. Mitarbeiter können so nicht nur äusserst schnell und sehr bequem erreicht werden, die Mitteilungen lassen sich auch sehr flexibel gestalten (bspw. durch den Einsatz von Hyperlinks). Die Schnelligkeit des Mediums zahlt sich auch in jenen Fällen aus, in denen man verhindern will, dass Mitarbeiter von Vorfällen in und um das Unternehmen zuerst von der Presse erfahren und nicht über interne Kommunikationswege.

Die 95 Kommunikationsverantwortlichen in der Studie von Hoffmann (2001) bewerteten auf einer fünfstufigen Skala die Befriedigung von Kommunikationsbedürfnissen der Mitarbeiter im Schnitt als das wichtigste Ziel der Mitarbeiterkommunikation ($M = 4.40$, $SD = .67$). Gefolgt von Public Relation (PR) nach innen ($M = 4.19$, $SD = .88$), Integration von Mitarbeitern ($M = 4.19$, $SD = .78$) und Motivation von Mitarbeitern ($M = 4.12$, $SD = .81$). Alle weiteren erhobenen Variablen besitzen Mittelwerte von unter 4.0. Aufgrund dieser Ergebnisse argumentiert Hoffmann (2001, S. 184), dass ausser bei der Befriedigung der Mitarbeiterkommunikation, primär unternehmensorientierte Ziele (wie PR nach innen, Integration und Motivation der Mitarbeiter usw.) und nicht mitarbeiterorientierte Ziele (wie Unternehmenskultur, Betriebsklima, Entwicklung und Qualifizierung der Mitarbeiter usw.) mit der Mitarbeiterkommunikation verfolgt werden. Für die Itemgenerierung interessant ist in diesem Zusammenhang das folgende Zitat von Hoffman (2001, S. 185): „Auffällig an den vorliegenden Untersuchungsergebnissen ist die herausragende Bedeutung der Befriedigung von Kommunikationsbedürfnissen und die damit angestrebte Zufriedenheit der Mitarbeiter.“

Für die Itemgenerierung wird angenommen, dass Benutzer sowohl die Erwartungen haben über das Intranet informiert zu werden, als auch, dass das Intranet die interne Kommunikation erleichtert (z.B. gutes elektronisches Telefonbuch).

Die Untersuchung der Benutzeroberflächen von 22 Intranets anhand von Usability-Kriterien wies bei den Intranets Schwächen bei der Gestaltung der Informationsarchitektur auf (Stimmt AG, 2003). Diese Beobachtung bestätigte sich auch in den CSUQ-Umfragen der Stimmt AG (2003, 2005). Die Auswertung der Kommentarfelder¹⁶ des CSUQ ergab, dass sich viele Intranetbenutzer an einer unklaren Informationsarchitektur stören. Vielfach kommen sie mit der Navigation nicht zurecht und können sich hinter den Bezeichnungen der Rubriken nichts vorstellen oder erwarten dahinter andere Inhalte. Dies verunmöglicht es ihnen, eine inhaltliche Struktur des Intranets zu erkennen. Das Auffinden von Inhalten kann so äusserst

¹⁶ Bei jedem Item des CSUQ's (Lewis, 1995) hatten die Befragten die Möglichkeit, einen Kommentar abzugeben. Diese wurden von der Stimmt AG (2003, 2005) anschliessend ausgewertet.

bemühend und unbefriedigend empfunden werden. Dieses Problem wird zusätzlich verschärft, da Suchmaschinen oftmals nicht die gewünschte Brauchbarkeit bieten können. Die Intranet Reports (Stimmt AG, 2003 und 2005) bezeichnen die Implementierung einer guten Suchmaschine als ein schwer zu erreichendes Ziel. Suchmaschinen werden so zwar von den meisten Intranets angeboten, die Benutzer sind damit jedoch selten zufrieden (Stimmt AG, 2003, S. 65). Bei einem Vergleich der Benutzerkommentare von Intranets mit hohen CSUQ-Werten mit Benutzerkommentaren von Intranets mit tiefen CSUQ-Werten zeigte sich, dass sich die Benutzer von Intranets mit tiefen CSUQ-Werten häufiger über schlechte Informationsarchitekturen, ungenügende Suchmaschinen, unübersichtliche Interfaces, suboptimale Workflows, fehlende/unvollständige oder veraltete Inhalte und zu wenig relevante Arbeitsinhalte ärgern (Stimmt AG, 2005).

Aus diesen Beobachtungen lassen sich erneut Items ableiten. Sowohl die wahrgenommene Qualität der Informationsarchitektur wie auch die wahrgenommene Qualität der Suchmaschine dürften die Benutzerzufriedenheit beeinflussen. Zudem reagieren Benutzer negativ, wenn sie das Gefühl haben, das Intranet unterstütze keine effiziente Arbeitsweise (suboptimale Workflows) oder beinhalte zu wenige arbeitsrelevante Inhalte. Von den befragten Intranetbenutzern vielfach kritisiert wurden ausserdem fehlende/unvollständige oder veraltete Inhalte. Dass nicht alle Inhalte immer auf dem neuesten Stand sind, ist gerade bei sehr grossen Intranets praktisch unvermeidlich. In solchen Fällen sollte es deshalb gewährleistet sein, dass eine verantwortliche Person möglichst einfach kontaktiert werden kann.

Mit dieser Darstellung schliesst der theoretische Teil dieser Arbeit. In diesem Kapitel wurden die zentralen Begriffe dieser Arbeit eingeführt und auf Basis des Erwartungs-mal-Wert-Ansatzes ein theoretisches Gerüst für den Fragebogen erstellt. Für die Itemgenerierung wurden bestehende Messinstrumente herangezogen und wichtige Facetten der Benutzerzufriedenheit näher beleuchtet. Im folgenden Kapitel wird nun der Fragebogen vorgestellt und über methodische Aspekte der Datenerhebung informiert.

3 Methode

Die Entwicklung des vorliegenden Fragebogens kann in mehrere Etappen gegliedert werden. Abbildung 3.1 vermittelt einen schematischen Überblick. Die Sichtung der theoretischen Ansätze und der empirischen Befunde wurden im vorherigen Kapitel dargestellt und bildeten die Basis für die Itemgenerierung. Im folgenden Abschnitt werden nun methodische Überlegungen und Vorgehensweisen der (eigentlichen) Fragebogenkonstruktion erläutert. Im Zentrum stehen Fragen, welche die konkrete Umsetzung des Fragebogens betreffen, wie Antwortformat, Testlänge oder Itemauswahl. Das Kapitel informiert weiter über die verwendeten Stichproben und klärt über die Art und Weise auf, wie die Umfragen durchgeführt wurden. Wie Abbildung 3.1 veranschaulicht – und Gegenstand der nächsten zwei Kapitel sein wird – wurde die erste Testversion nach einer Validierung modifiziert und erneut einer Stichprobe vorgelegt. Die Auswertungsergebnisse dieser zweiten Umfrage führten dann zur vorläufigen Endversion des Fragebogens.

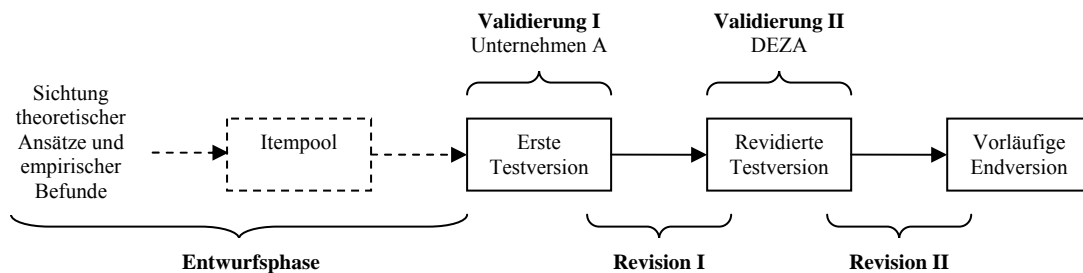


Abbildung 3.1. Schematischer Überblick der vorliegenden Testentwicklung.

Die Umfragen wurden jeweils mit einem webbasierten Fragebogen durchgeführt. Neben den Vorteilen die Online-Umfragen in der Regel mit sich bringen, legte auch der Untersuchungsgegenstand „Intranet“ ein solches Vorgehen nahe. Diese Erhebungsmethode bleibt insbesondere bei der Durchführung nicht ohne methodische Konsequenzen. Folgend wird deshalb auch auf die Erhebungsart eingegangen. Abgeschlossen wird das Kapitel mit Hinweisen zur Auswertung der Rohdaten.

3.1 Fragebogenkonstruktion

3.1.1 Skalierung

Für die Items wurde eine Likert-Skala verwendet. In dieser etablierten Skala zur Messung von Einstellungen werden die Befragten gebeten, über ein Einstellungsobjekt das Ausmass ihrer Zustimmung oder Ablehnung zu einer Reihe von Aussagen einzustufen. Verwendet wurde in diesem Fall eine sechsstufige Ratingskala, wobei die Abstufungen von Eins bis Sechs durchnummeriert waren und die äusseren Punkte schriftlich vorgegeben wurden (1 = Stimme überhaupt nicht zu; 6 = Stimme voll und ganz zu). Eine höhere Zahl drückt damit zugleich ein Mehr an Zustimmung aus. Für die Ratingskala wird Intervallskalenniveau angenommen, was es erlaubt, alle darauf beruhenden statistischen Verfahren zu verwenden. Diese Annahme von Intervallskalenniveau für eine Ratingskala, ohne vorherige empirische und aufwendige Überprüfung der jeweiligen Skalenaxiomatik (Borg & Staufenbiel, 1997), ist in der Forschungspraxis üblich (Bortz, 1999, S. 27f.) und wird hier deshalb nicht weiter ausgeführt.¹⁷

Aus Gründen der Reliabilität und Validität empfiehlt Borg (2001) bei einer Likert-Skala nicht weniger als fünf und nicht mehr als sieben Kategorien zu verwenden. Mit der sechsstufigen Ratingskala standen den Probanden je drei Skalenstufen für eine ablehnende und je drei Skalenstufen für eine zustimmte Einstellung zur Verfügung. Für den Probanden bedeutet eine gerade Anzahl von Kategorien, dass er „gezwungen“ ist, sich bei seinem Urteil für eine Richtung (Ablehnung oder Zustimmung) zu entscheiden. Eine gerade Anzahl Kategorien wurde deshalb gewählt, weil sich die Verwendung einer mittleren bzw. neutralen Kategorie in vielen Untersuchungen als nachteilig erwiesen hat (Rost, 2004, S. 67). So kann ein Proband nach Mummendey (1995, S. 56) eine mittlere Antwortkategorie gewählt haben, weil er tatsächlich eine mittlere Einstellungsposition besitzt, er nicht wusste wie er antworten soll, er die Frage für irrelevant hielt, er eine Antwort verweigert oder weil er so seinen Unmut über eine Frage ausdrücken will (Protest-Antwort). Eine mittlere Antwortkategorie erhebt damit nicht unbedingt die Persönlichkeitseigenschaft, die durch sie gemessen werden soll. Rost (2004, S. 67) weist zudem darauf hin, dass die mittlere Kategorie von motivierten Probanden oft gemieden wird. Aus theoretischer Sicht hat eine solche Vermeidungstendenz zur Folge, dass die mittlere Antwortkategorie mit den anderen Kategorien keine Ratingskala mehr bildet, die aus gleichen (äquidistanten) Kategorienabständen besteht. Eine wichtige

¹⁷ Bortz (1999, S. 27) bezeichnet dieses Vorgehen als „per-fiat“-Messung.

Grundannahme für den Anspruch einer Ratingskala nach Intervallskalenqualität wäre auf diese Weise verletzt.

Der Fragebogen enthält Items, die zur Beantwortung bestimmte Erfahrungen mit dem Intranet voraussetzen. Es kann deshalb nicht ausgeschlossen werden, dass einige Probanden ein Item infolge mangelnder Erfahrung nicht beantworten können. Eine reine Ratingskala ohne Ausweichmöglichkeiten würde in diesen Fällen zu einer Verzerrung führen. Aus diesem Grund wurde den Probanden neben der Ratingskala ein alternatives Antwortfeld angeboten, falls ein Item nicht beantwortet werden konnte.

Zu Recht kann hier kritisiert werden, dass diese zusätzliche Antwortkategorie methodisch zu ähnlichen Problemen führt, wie sie für eine mittlere Antwortkategorie gelten. Auch in diesem Fall können Antworten nicht eindeutig interpretiert werden. Ein Proband könnte dieses Feld gewählt haben, weil er über einen Punkt nicht Bescheid wusste, er könnte sich aber auch dafür entschieden haben, weil er sich nicht zwischen dem Urteil „eher zustimmen“ und „eher nicht zustimmen“ entscheiden konnte.

Trotz dieser methodischen Schwäche, wurde entschieden die Antworten der Probanden mittels sechsstufiger Skala und Ausweichmöglichkeit zu erfassen. Dies erschien als der beste Kompromiss, da das Ausdrücken eines „Nicht-wissens“ möglich sein musste.

3.1.2 Testlänge und Testzeit

Der Begriff der Testlänge kann je nach Gebrauch unterschiedlich verstanden werden. Die folgende Verwendung der Begriffe Testlänge und Testzeit, orientiert sich an Lienert und Raatz (1998, S. 33) und definiert Testlänge als die Anzahl der Testaufgaben und Testzeit als die zur Beantwortung der Testaufgaben benötigte Zeit.

Um eine Schätzung für eine angemessene Testlänge mit guter Reliabilität zu erhalten, wurden ähnliche Messinstrumente als Referenz herangezogen. Dies waren der *IBM Computer Usability Satisfaction Questionnaires* (Lewis, 1995), die *Perceived Usefulness und Perceived Ease of Use Skalen* von Davis (1989) und das Fünfkostenmodell für Endbenutzerzufriedenheit von Doll und Torkzadeh (1988). Die drei Messinstrumente bestehen aus zwischen 12 und 20 likert-skalierten Items. Mit der Spearman-Brown-Formel lässt sich berechnen, um wie viele Items diese Tests theoretisch jeweils verlängert werden müssten, um die gewünschte Reliabilität zu erhalten (Rost, 2004). Auf diese Weise liesse sich eine potentiell „bessere“ Testlänge abschätzen. Da jedoch bereits alle drei Tests über eine

interne Konsistenz von über .90 verfügen, wurde auf diesen Berechnungsschritt verzichtet und eine Testlänge von ± 20 Items als Ausgangsgrösse verwendet.

Mit einer anvisierten Testlänge von ungefähr zwanzig Items, kann mit einer durchschnittlichen Testzeit von 10 bis 15 Minuten gerechnet werden. Dies deckt sich mit einer Empfehlung von Batinic und Bosnjak (1999, S. 307), die für internetbasierte Umfragen eine Testzeit von 6 bis 15 Minuten als angebracht erachten.¹⁸ Eine relativ kurze Testzeit entspricht zudem den Vorgaben die sich aus ökonomischer Sicht ergeben. Der Fragebogen dürfte von den Probanden im Normalfall während der Arbeitszeit ausgefüllt werden, wenn dieser deshalb für das Ausfüllen mehr als 10 bis 15 Minuten in Anspruch nähme, dürfte er sowohl bei den Mitarbeitenden als auch beim Management kaum auf Akzeptanz stossen.

3.1.3 Itemgenerierung und erste Testversion

Aus der Sichtung der theoretischen Ansätze und empirischen Befunde, wie sie im Kapitel zuvor vorgestellt wurden, entstand ein erster Itempool. In einer Überererarbeitung dieses Itempools wurden redundante Items ausgeschlossen und die Items sprachlich überarbeitet¹⁹ Daraus entstand ein erster Vorentwurf des Fragebogens, welcher zwei Intranetmanagern zur Beurteilung vorgelegt wurde. Sie sollten dabei begutachten, ob die Test-Items das Testmerkmal „Benutzerzufriedenheit mit dem Intranet“ inhaltlich gut repräsentieren und ob ihrer Meinung nach ein wichtiger Aspekt durch die Items nicht abgedeckt wurde. Dieser Schritt sollte die inhaltliche Validität erhöhen und führte zur ersten Testversion. Der folgende Kasten führt die Items der ersten Testversion auf.

Überblick der verwendeten Items in der ersten Testversion

- | | |
|---|---|
| 1 | Das Intranet stellt für mich arbeitsrelevante Inhalte zur Verfügung. |
| 2 | Die Informationen auf dem Intranet sind aktuell. |
| 3 | Die Inhalte auf dem Intranet sind so angeordnet, dass es einfach ist, sich zu orientieren und gewünschte Inhalte zu finden. |
| 4 | Die Art und Weise, wie Informationen auf dem Intranet vermittelt werden, ist klar und gut verständlich. |
| 5 | Ich bin genügend vertraut im Umgang mit dem Intranet, so dass ich es optimal für meine Bedürfnisse nutzen kann. |

¹⁸ Zu diesem Schluss kommen sie aufgrund einer Umfrage unter deutschen Internetnutzer (N = 187), die danach befragt wurden, wie viel Zeit sie maximal bereit wären für eine Internetbasierte Umfrage zur Verfügung zu stellen. Rückschlüsse von dieser interviewten Gruppe auf Intranetbenutzer sind daher nur bedingt zulässig. Die Empfehlung von Batinic und Bosnjak kann hier aber trotzdem als Anhaltspunkt für eine angemessene Testzeit dienen.

¹⁹ Für allgemeine Regeln zur Formulierung von Items, siehe Mummendey (1995, S. 63f).

- | | |
|----|--|
| 6 | Das Intranet erleichtert die interne Kommunikation (z.B. durch das Mitarbeiter-Verzeichnis, Mitteilungen des Managements oder Diskussionsforen etc.). |
| 7 | Bei Bedarf, kann ich jederzeit auf das Intranet zugreifen. |
| 8 | Ich bin zufrieden mit der Qualität der Suchmaschine. Sie liefert gute und brauchbare Resultate und eignet sich für das gezielte Auffinden von bestimmten Inhalten oder Dokumenten. |
| 9 | Wenn etwas im Intranet veraltet, falsch oder unvollständig ist, dann ist es einfach, die verantwortliche Person zu kontaktieren. |
| 10 | Das Intranet ermöglicht mir effizientere Arbeitsweisen (z.B. firmeninterne Arbeitsabläufe, Support oder Informationsbeschaffung). |
| 11 | Falls ich eine Mitteilung oder ein Dokument auf das Intranet stellen möchte, ist mir klar, wie ich vorgehen muss. |
| 12 | Mit dem Intranet lässt sich zügig arbeiten (z.B. schneller Seitenaufbau oder Dokumente herunterladen). |
| 13 | Das Intranet ist einfach zu bedienen (z.B. Personalisierung, Bedienen des Mitarbeiter-Verzeichnisses usw.). |
| 14 | Ich bin zufrieden mit der Hilfe bzw. der Unterstützung, die ich bekomme, wenn ich ein Problem oder eine Frage zum Intranet habe (z.B. Helpdesk oder Online-Hilfe). |
| 15 | Das Intranet versorgt mich laufend mit aktuellen Firmen-Neuigkeiten. |
| 16 | Die für meine Arbeit benötigten Informationen finde ich auf dem Intranet so vor, dass ich sie einfach weiterverarbeiten kann. |
| 17 | Ich verlasse mich auf die Informationen im Intranet. |
| 18 | Insgesamt bin ich mit dem Intranet zufrieden. |

Die Items des Vorentwurfs sind im Anhang B1 einsehbar. Ein Vergleich zeigt, dass einige der Items in der Absicht auf eine präzisere bzw. allgemein verständlichere Ausdrucksweise umformuliert wurden. Dies galt für die Items 9, 12, 16 und 18. Zwei Items wurden für die erste Testversion ausgeschlossen. Das Item „Ich finde das Intranet ein wichtiges und nützliches Arbeitsinstrument“ wurde gegenüber Item 1 als zu wenig trennscharf beurteilt und deshalb ausgesondert. Für das Item „Ich weiss, welche Mitteilungen oder Dokumente von mir auf das Intranet gestellt werden sollten“ erwies sich die generelle Einsetzbarkeit als problematisch. So wird die Bereitstellung von Informationen im Intranet im Unternehmen A, aus Gründen der Qualitätssicherung nur von einem relativ kleinen und entsprechend ausgebildeten Personenkreis wahrgenommen. Da dies auch in anderen Firmen der Fall ist, wurde das Item für den Fragebogen als ungeeignet betrachtet und ausgeschlossen.

Die Formulierung des Items 17 im Vorentwurf („Ich kann mich auf die Informationen im Intranet verlassen und darauf vertrauen, dass sie richtig sind“) wurde von einem der Intranetmanager kritisiert, weil die Aussage bei den Probanden Irritationen auslösen könnte, im Sinne von: „Wieso diese Frage? Kann ich mich auf die Informationen etwa nicht verlassen?“. Um das suggestive Potential des Items in diese Richtung zu mildern und einen unerwünschten Effekt zu unterdrücken, wurde das Item umformuliert.

3.2 Stichproben

3.2.1 Stichprobengewinnung

Die Daten der ersten Testvalidierung stammen aus dem Umfrageergebnis einer Unternehmung aus dem Finanzsektor. Die grosse und international tätige Versicherungsgesellschaft meldete ihr Interesse an einer Teilnahme selber an. Aufmerksam auf diese Lizentiatsarbeit wurde der Intranetmanager aufgrund der Beschreibung des Projekts auf der Website des Schwerpunktes Mensch-Maschine Interaktion des Psychologischen Instituts der Universität Basel. Auf die Website stiess er durch die Suchmaschine Google. Da dem Unternehmen zugesichert wurde, dass die veröffentlichte Studie keine Rückschlüsse auf die Qualität ihres Intranets zu lassen wird, wird sie in dieser Arbeit nicht namentlich aufgeführt und folgend als Unternehmen A bezeichnet.

Für die Validierung der überarbeiteten Testversion konnte die Direktion für Entwicklung und Zusammenarbeit (DEZA) gewonnen werden. Die Intranetmanagerin der DEZA sagte nach Anfrage einer Teilnahme zu.

Die Zusammenarbeit sah vor, dass die Unternehmen ihr Intranet für die Umfrage zur Verfügung stellten und im Gegenzug dafür die erhobenen Rohdaten für eigene Auswertungen erhalten. Es stand ihnen zudem zu, am Ende des Fragebogens noch eigene Fragen (demografischer oder intranetspezifischer Art) anzufügen. Des Weiteren bekamen sie Anrecht darauf, die Endversion des Fragebogens später frei für ihre Unternehmung nutzen zu dürfen.

Lienert und Raatz (1998, S. 60) empfehlen für Itemanalysen eine Stichprobengrösse von 400 Probanden. Bei günstigen Bedingungen, wenn der Test bspw. nur für eine relativ homogene Grundgesamtheit angewendet wird, genügen auch 200 Probanden. Für die erste Untersuchung wurde mit 881 Probanden diese Empfehlung mehr als erfüllt. Die zweite Untersuchung konnte mit 127 Probanden diese Vorgabe nicht ganz erfüllen. Weitere Beschreibungen zu den untersuchten Stichproben wie Grösse und Verteilung werden in den Kapiteln 4. resp. 5 beschrieben.

3.2.2 Stichprobenrepräsentativität

Online-Befragungen entwickelten sich im Fahrtwind der rasanten Internetentwicklung relativ schnell zu einer beliebten und breit eingesetzten Untersuchungsmethode bei Markt- und Sozialforschern (Dillman et al., 1999; Granello & Wheaton, 2004). Werden sie jedoch für Umfragen eingesetzt, in denen der Nutzerkreis nicht genau abgrenzbar ist, wie bspw. bei

Bevölkerungsumfragen, weisen Umfragen unter Internetbenutzern erhebliche methodische Probleme auf (Batinic & Bosnjak, 2000).

Kann die Grundgesamtheit jedoch klar umrissen und vollständig über das Internet erreicht werden, sind die Kriterien für Repräsentativität grundsätzlich erfüllbar. Bei Intranet-Umfragen sind solche Bedingungen gegeben, wenn die Mitarbeiter am Arbeitsplatz über vernetzte Computer verfügen (Deutschmann, 1999). Dies war in beiden Untersuchungen dieser Studie der Fall. Neben dieser infrastrukturellen Bedingung (PC, Internetverbindung, Email-Adresse), bedarf es von den Probanden einer gewissen technischen Fertigkeit, damit sie den Online-Fragebogen ausfüllen können (Dillman et al., 1999; Thompson et al., 2003). In den getesteten Stichproben gehört der Umgang mit dem Computer zum täglichen Handwerk, weshalb ausgeschlossen werden kann, dass mangelnde Computerkenntnisse für einige der Probanden eine Barriere zur Teilnahme hätte bedeuten können.

In Bezug auf die Generalisierbarkeit der Resultate gilt sowohl für das Unternehmen A wie auch für die DEZA, dass sie mit dem Zielpublikum des zukünftigen Tests übereinstimmt. Wie repräsentativ die untersuchten Intranets allerdings für die Gruppe der Intranets aller Unternehmen sind, lässt sich nur schwer beurteilen. Zwischen einzelnen Intranets können als Ganzes oder in einzelnen Bereichen erhebliche Unterschiede bestehen. Dieser Punkt mindert die Generalisierbarkeit der Ergebnisse, liess sich aber nicht verhindern und gehört zu einem Problem, wie es sich oft bei psychologischen Untersuchungen stellt.

3.3 Erhebungsmethode und Durchführung

3.3.1 Allgemeine Bemerkungen zu webbasierten Umfragen

Online- bzw. webbasierte Umfragen sind bei vielen Forschern beliebt, weil sie es ermöglichen, innert einer relativ kurzen Zeit eine grosse Menge von Daten zu generieren. Darüber hinaus müssen im Vergleich mit anderen Erhebungsarten im Normalfall auch weniger materielle, finanzielle und personelle Ressourcen eingesetzt werden (Dillman et al., 1999; Granello & Weahnton, 2004). Ein grosser Vorteil webbasierter Umfragen ergibt sich so durch die Möglichkeit der automatischen Dateneingabe. Damit kann sowohl Zeit als auch Arbeitskraft eingespart werden. Thompson und Mitarbeiter (2003) weisen zudem darauf hin, dass eine automatische Dateneingabe zu genaueren Daten führt, da fehlerhafte Codierungen unwahrscheinlicher sind. Hinzu kommt, dass sich der Ablauf bei webbasierten Umfragen in

hohem Grad standardisieren lässt, was eine hohe Durchführungs- und Auswertungsobjektivität ermöglicht (Batinic, 2004, S. 227).

Ein grundsätzliches Problem, welches sich beim Einsatz mit webbasierten-Untersuchungen zwangsläufig ergibt, ist, dass sich der Fragebogen nicht für alle Probanden in gleicher Weise darstellen lässt. Je nach Monitor- und Browsereinstellungen der Probanden kann das Aussehen der Fragebögen variieren und je nach Computer und Netzverbindungen können unterschiedlich lange Ladezeiten bestehen. Reips (2000, S. 331) spricht in diesem Zusammenhang von technischer Varianz. Solche Unterschiede sind zwar nie auszuschliessen, können durch den Verzicht von übertrieben neuen Funktionen und einem simplen Fragebogendesign, welches wenig Arbeitsspeicher beansprucht, klein gehalten werden (Dillman et al., 1999). Schliessen technische Unterschiede keine Probanden systematisch aus, bspw. weil aufgrund von unterschiedlich langen Ladezeiten eine Gruppe die Umfrage häufiger abbricht oder erst gar nicht teilnimmt, kann technische Varianz als Vorteil betrachtet werden, da sie die externe Validität verbessert und damit die Generalisierbarkeit erhöht.²⁰

Zwei andere Probleme, die sich üblicherweise bei Online-Umfragen stellen, betreffen die Kontrolle zum Zugang einer Umfrage und die Möglichkeit, dass ein Proband den Fragebogen mehrfach ausfüllt (Rogelberg, Church, Waclawski & Stanton, 2002). Ein Hyperlink, der zu einer Umfrage führt, kann ohne Probleme an andere Personen weitergeleitet werden. Dies kann dazu führen, dass Personen an der Untersuchung teilnehmen, die nicht aus der anvisierten Grundgesamtheit stammen. Für einen Probanden ist es auch möglich, einen Hyperlink mehr als nur einmal aufzurufen und so einen Fragebogen wiederholt auszufüllen. Beides kann durch technische Vorkehrungen (bspw. Gebrauch von Passwörtern) verhindert werden. Neben dem daraus entstehenden Mehraufwand bei der Umsetzung, gibt Stanton (1998) zu bedenken, dass der Einsatz von Passwörtern bei den Probanden negative Auswirkungen auf die wahrgenommene Anonymität haben kann.

Für beide Untersuchungen kann nicht ausgeschlossen werden, dass der Hyperlink zur Umfrage an Dritte verschickt wurde, oder ein Proband mehrfach an der Umfrage

²⁰ Für die meisten heutigen Unternehmen dürften aufgrund relativ hoher technischer Standards solche Probleme ausgeschlossen werden. Das Beispiel der DEZA zeigt aber, dass solche Fragen je nach Unternehmen noch zu beachten bleiben. So haben die Koordinationsbüros, welche die DEZA auf der ganzen Welt betreibt und ebenfalls Zugriff auf das Intranet haben, zum Teil erhebliche Netzprobleme. Für diese Untersuchung konnte dieser Aspekt allerdings ausser Acht gelassen werden, da sich die Umfrage auf die Zentrale in der Schweiz beschränkte.

teilgenommen hat.²¹ Unter dem inhaltlichen Aspekt dieser Fragebogenerhebung, scheinen aber beide Fälle eher unwahrscheinlich. Problematischer wäre in diesem Zusammenhang sicherlich eine Umfrage zur Zufriedenheit mit den Vorgesetzten. Damit kein Proband mehrfach an der Umfrage teilnahm, weil er unsicher war, ob seine Daten übermittelt wurden, erhielten die Probanden nach dem Abschicken der Umfrage eine klare Rückmeldung mit der Bestätigung der erfolgreichen Datenübermittlung.

3.3.2 Durchführung und Instruktion

In beiden vorliegenden Untersuchungen wurden die Probanden jeweils per E-Mail um eine Teilnahme an der Umfrage gebeten. Die Betreffzeilen wurden mehrsprachig formuliert und lauteten für das Unternehmen A: „Sind Sie mit Ihrem Intranet zufrieden? Êtes-vous satisfait/e de votre Intranet? È contento del Suo Intranet?“. Im Fall der DEZA wurde die Betreffzeile etwas prägnanter formuliert: „Für ein besseres IntraWeb – Pour améliorer IntraWeb.“

Im Inhaltsteil dieser Mails wurde der wertvolle Beitrag hervorgehoben, den eine Teilnahme an der Umfrage zur Verbesserung des Intranets leiste, der zeitliche Aufwand für das Ausfüllen mitgeteilt sowie die Gewährleistung der Anonymität betont. Für den genauen Wortlaut siehe Anhang B2. Über einen Hyperlink konnten die Probanden zur Umfrage gelangen. Auf der aufgerufenen Seite wurden sie dann gebeten, die Aussagen zu lesen und anzugeben, wie stark sie diesen zustimmen. Da sich ein Intranet ständig weiterentwickelt, wurden die Probanden angehalten, sich bei den Beurteilungen der Aussagen so gut es gehe auf den momentanen Zustand zu beziehen. Konnte ein Proband eine Aussage nicht beurteilen, war er ersucht das Feld „weiss nicht“ bzw. in der DEZA-Umfrage das Feld „kann ich nicht beantworten“ anzukreuzen. Am Schluss der Umfrage wurde den Probanden herzlich für ihre Teilnahme gedankt und darauf hingewiesen, dass die Resultate der Umfrage auf dem Intranet publiziert würden.

Bei jeder Aussage hatte ein Proband die Möglichkeit einen Kommentar abzugeben. Vor allem für die Interpretation der Daten können Kommentare eine wichtige zusätzliche Quelle bedeuten. Hier kommt einer Online-Umfrage zugute, dass heute viele Personen schneller mit der Tastatur als von Hand schreiben. Thompson und Mitarbeiter (2003) nehmen deshalb an, dass Probanden in Online-Umfragen motivierter sind, Bemerkungen oder Kommentare abzugeben. Wie später noch näher beschrieben wird, waren beide Fragebögen so programmiert, dass es nicht möglich war ein Item offen (unbeantwortet) zu lassen.

²¹ Bei der Umfrage im Unternehmen A war eine mehrfache Teilnahme über den gleichen Computer nur dann möglich, wenn der Proband zuvor ein Cookie, welches bei der ersten Teilnahme automatisch gesetzt wurde, wieder entfernt hätte.

Die Fenster für die Umfragen waren beide Male für zweieinhalb Wochen geöffnet. Ein paar Tage vor der Schliessung des Fensters wurde an alle Probanden ein Reminder verschickt. Inhalt war neben dem Hinweis auf das baldige Ende der Umfrage ein Aufruf an all jene, die sich zum Thema noch äussern wollten, dies bis zur Schliessung zu tun. Der genaue Textlaut ist im Anhang B2 wiedergegeben. Reminder können als Mittel eingesetzt werden, um die Rücklaufquote zu erhöhen. So fanden Batinic und Moser (2005) in ihrer Untersuchung zu Determinanten der Rücklaufquote in Online-Panels einen positiven Zusammenhang zwischen Remindern und Rücklaufquote. Allerdings handle es sich dabei nicht um eine lineare Beziehung, d.h. mehr als zwei Erinnerungsschreiben scheinen keinen weiteren Zuwachs zu bringen. Für den Zeitraum von zweieinhalb Wochen erschien das einmalige Verschicken eines Reminders als angemessen, weitere hätten wohlmöglich einzelne Probanden verärgert.²²

Frick, Bächtiger und Reips (1999) empfehlen, persönliche Daten der Probanden zu Beginn und nicht am Ende einer Online-Umfrage zu erheben. In ihre Studie fanden sie Hinweise dafür, dass die Drop-Out-Rate auf diese Weise tiefer gehalten werden könne. Weiter vermuten sie, der demografische Teil würde auf diese Weise von den Probanden besser komplettiert. In den hier durchgeführten Umfragen wurden die demoskopischen Fragen dennoch ans Ende gestellt. Dies sollte verhindern, dass Probanden durch die vorangesellten Fragen, in der Art ihrer Beantwortung beeinflusst werden. Jemand, der sich bspw. vor der eigentlichen Umfrage als Wenig-User kategorisiert, könnte bei der Beantwortung gewisser Items eventuell dazu neigen, sein Nutzverhalten bei der Beurteilung miteinzubeziehen. Dies könnte dazu führen, dass er bei gewissen Items zu mildereren bzw. positiveren Urteilen tendiert, da er nicht nur das Intranet, sondern möglicherweise auch sich selber als Quelle für ein schlechtes Abschneiden sieht und sein Urteil entsprechend nach oben „korrigiert“. Durch dieses Vorgehen wurde versucht, Einflüsse die sich aus dem Item-Kontext heraus hätten ergeben können, möglichst gering zu halten (Rogelberg, Church, Waclawski & Stanton, 2002).

3.3.2.1 Beschreibungen der eingesetzten Online-Fragebögen

Konradt und Mitarbeiter (2003) untersuchten die Qualität von computer- und internetbasierten Verfahren in der Berufsdiagnostik. Sie kamen zum Schluss, dass durch Online-Befragungen der Anspruch nach Reliabilität wie auch Validität erfüllt werden kann. Um dies zu erreichen,

²² Vergleiche zwischen Rücklaufquoten von Papier-Bleistift-Umfragen und Online-Umfragen zeigen keine einheitlichen Befunde. So fand man in Studien die diese zwei Methoden bezüglich Rücklaufquote verglichen, sowohl höhere, tiefere als auch gar keine Unterschiede (Thompson et al., 2003).

müssen jedoch Benutzeroberfläche und Dialogführung ergonomischen Kriterien wie z.B. Fehlertoleranz und Selbstbeschreibungsfähigkeit gerecht werden. Folgend werden die eingesetzten Online-Fragebögen kurz beschrieben.

Unternehmen A verfügt für interne Befragungen über ein firmeneigenes webbasiertes Umfrage-Tool. Es lag deshalb aus praktischen Gründen nahe, die Untersuchung über dieses laufen zu lassen. Die Schweizerische Versicherungsgesellschaft ist international tätig, wollte sich jedoch für die Untersuchung auf die Mitarbeitenden in der Schweiz begrenzen. Da sie alle drei grossen Sprachgruppen der Schweiz gleichermassen berücksichtigen wollte, wurde der Fragebogen für die Untersuchung in einer deutschen, französischen und italienischen Version angeboten. Die Übersetzungen für die französische und italienische Version wurden von einem internen Übersetzungsdienst der Firma A erstellt. Abbildung 3.2 zeigt einen Screenshot der Umfrage.

39%

Beurteilen Sie bitte, wie zutreffend die folgenden Aussagen für Sie sind.
(1 = stimme überhaupt nicht zu - 6 = stimme voll und ganz zu)

	1	2	3	4	5	6	weiss nicht
Das Intranet stellt für mich arbeitsrelevante Inhalte zur Verfügung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Informationen auf dem Intranet sind aktuell.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Inhalte auf dem Intranet sind so angeordnet, dass es einfach ist, sich zu orientieren und gewünschte Inhalte zu finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Möchten Sie uns zu diesen Aussagen etwas mitteilen?

[Zurück](#) [Weiter](#)

Abbildung 3.2. Screenshot der Umfrage im Unternehmen A; exemplarisch die Items 1 bis 3 der deutschen Version.

Drückte ein Proband den „Weiter-Button“, ohne zuvor alle Items auf dem Bildschirm beantwortet zu haben, löste dies die Fehlermeldung „Hinweis: Bitte füllen sie die Felder aus, um zur nächsten Frage zu gelangen“ aus. Zusätzlich wurde dabei das unbeantwortete Item farblich markiert. Ein Weiterkommen war nur möglich, wenn alle Items beantwortet wurden. Der Statusbalken oben rechts informierte den Probanden, wieviele Items schon beantwortet wurden. Das Textfeld gab dem Probanden die Möglichkeit, sich zu einem Item zu äussern.

Neben den eigentlichen Fragebogen-Items richtete das Unternehmen A am Ende des Fragebogens noch einige firmenspezifische Fragen an die Probanden. Die Länge des Fragebogens betrug dadurch insgesamt 29 Items.

Für die DEZA-Umfrage wurde im Schwerpunkt Mensch-Maschine Interaktion der Universität Basel ein Online-Fragebogen erstellt. Als Vorlage diente dabei der CSUQ, wie er von der Stimmt AG eingesetzt wird. Die Umfrage selber lief über einen Server der Universität Basel. Während die Umfrage durchgeführt wurde, kam es zu keinerlei technischen Problemen. Abbildung 3.3 gibt einen optischen Eindruck von der Umfrage wider.

The screenshot displays four survey items, each with a 6-point Likert scale and a text input field for comments. The items are:

- Item 5:** "Ich bin genügend vertraut im Umgang mit dem Intranet, so dass ich es optimal für meine Bedürfnisse nutzen kann." (I am sufficiently familiar with the intranet so that I can use it optimally for my needs.)
- Item 6:** "Das Intranet erleichtert die interne Kommunikation (z.B. durch das Mitarbeiter-Verzeichnis, Mitteilungen des Managements oder Diskussionsforen etc.)." (The intranet facilitates internal communication (e.g., through the employee directory, management communications or discussion forums etc.).)
- Item 7:** "Bei Bedarf, kann ich jederzeit auf das Intranet zugreifen." (If needed, I can access the intranet at any time.)
- Item 8:** "Wenn etwas im Intranet veraltet, falsch oder unvollständig ist, dann ist es einfach, die verantwortliche Person zu kontaktieren." (If something in the intranet is outdated, false or incomplete, it is easy to contact the responsible person.)

Each item includes a legend: "1 = Stimme überhaupt nicht zu." (I completely disagree) and "6 = Stimme voll und ganz zu." (I completely agree). Below the legend are six radio buttons labeled 1 through 6, and a "Kann ich nicht beantworten." (I cannot answer) option. A text input field is provided for each item with the prompt "Möchten Sie uns zu dieser Aussage etwas mitteilen?" (Would you like to tell us anything about this statement?). A "weiter" (next) button is located at the bottom right of the survey area.

Abbildung 3.3. Screenshot der Umfrage in der DEZA; exemplarisch die Items 5 bis 8 der deutschen Version.

Wie bereits beim Unternehmen A, konnte ein Proband im Fragebogen nur dann weitergelangen, wenn er zuvor alle Items auf dem Bildschirm beantwortete. Die Probanden konnten ihre Antworten korrigieren, indem sie mittels Cursor einfach auf den „richtigen“ Radiobutton wechselten. Zugunsten einer schlankeren Instruktion und der Annahme, Probanden würden intuitiv von dieser Korrekturmöglichkeit ausgehen, wurde darauf verzichtet, dies explizit zu erwähnen.

Die DEZA wollte dem Fragebogen keine eigenen Items hinzufügen. Der Fragebogen bestand so aus insgesamt 20 Items. Er wurde den Probanden in deutscher und französischer Version angeboten.

3.4 Anmerkungen zu den Auswertungen der Rohdaten

Die Auswertung der Rohdaten basiert auf der klassischen Testtheorie und den daraus abgeleiteten Prinzipien der Testkonstruktion. Da es sich hierbei um ein in der Psychologie allgemein bekanntes Verfahren handelt, wird hier auf eine zusammenfassende Darstellung verzichtet.

Die statistische Auswertung wurde mit Hilfe des Statistikprogramms SPSS 11.5 für Windows ausgeführt. Neben diesem Basismodul wurde auch das Zusatzmodul SPSS Missing Value Analysis verwendet. Dieses Modul ermöglicht die Analyse von fehlenden Werten und bietet Lösungsmöglichkeiten an, wie fehlende Werte ersetzt werden können. Wie bereits erwähnt, wurden die Antworten der Probanden automatisch erfasst, die Daten mussten daher lediglich noch in SPSS importiert werden.

4 Testvalidierung I

Kern dieses Kapitels bilden die Itemanalyse der Umfrageergebnisse vom Unternehmen A und die Selektion bzw. die Überarbeitung der Items aufgrund dieser Analyse. Vor diesen zwei Hauptblöcken wird die Stichprobe beschrieben und die Rohdaten der Umfrage einer ausführlichen Untersuchung unterzogen.

4.1 Stichprobenbeschreibung

Insgesamt wurden 2359 Personen per E-Mail um eine Teilnahme an der Umfrage gebeten. Diesem Aufruf kamen 1104 Personen nach, womit eine Rücklaufquote von 47% erreicht wurde. Die Grössen der drei Stichproben fielen sehr unterschiedlich aus. So zählt die deutschsprachige Stichprobe 908, die französischsprachige 152 und die italienischsprachige 44 Probanden. Diese Unterschiede widerspiegeln die Mitarbeiterstruktur beim Unternehmen A und wurden daher in dieser Form erwartet. Bereits im Vorfeld der Umfrage hatte der Intranetmanager der Unternehmung A darauf hingewiesen, dass die Stichproben unterschiedlich gross ausfallen dürften. Es gibt keine Angaben darüber, wie sich die Stichprobe von 2359 Probanden auf die drei Sprachgruppen aufteilte. Von den 1104 Probanden, die den Fragebogen ausgefüllt haben, wählten 82% die deutsche, 14% die französische und 4% die italienische Version. Für die Itemanalyse wurde nur die deutschsprachige Stichprobe berücksichtigt.

Nach der Datenexploration (siehe Kapitel 4.1.1) wurden 27 der 908 Probanden aus der Stichprobe ausgeschlossen. Der Stichprobenumfang beträgt damit 881 Probanden. Die Altersstruktur der Stichprobe präsentiert sich relativ ausgeglichen. So gaben 221 der 881 Probanden an, unter 30 Jahren alt zu sein, 276 Probanden zwischen 31 und 40 Jahren, 246 zwischen 41 und 50 Jahren und 138 gaben an, über 51 Jahre alt zu sein. Die meisten Probanden arbeiten seit mehr als 10 Jahren beim Unternehmen A (367 Probanden = 41.7%). Etwa gleich viele stehen seit 1 bis 5 Jahren (236 Probanden = 26.8%) bzw. seit 6 bis 10 Jahren (205 Probanden = 23.2%) in einem Angestelltenverhältnis. Die kleinste Gruppe mit 73 Probanden (= 8.3%) arbeitet seit weniger als einem Jahr im Unternehmen. 44 Probanden kreuzten an, Mitglied der Direktion zu sein, 262 zählten sich zum Kader, 527 bezeichneten sich als Mitarbeiter(in) und 48 Probanden wählten die Kategorie Lernende(r). Im Anhang C1 liegen diese Angaben zur Stichprobe in tabellarischer Form vor. Aufgrund eines Missverständnisses wurde das Geschlecht nicht erhoben.

4.1.1 Datenexploration der Rohwerte

Die Datenexploration dient dem Aufdecken von Eingabefehlern und der Überprüfung auf Normalverteilung. Zudem setzt sich dieser Abschnitt mit Ausreißern von der Mittelwertskala und den fehlenden Werten (Missing Data) der Stichprobe auseinander.

4.1.1.1 Verteilung der Rohdaten

Durch die automatische Datenerfassung können Eingabefehler bei der Codierung ausgeschlossen werden. Der (Roh-)Mittelwert der Probanden liegt mit 4.29 und einer Standardabweichung von .79 im oberen Bereich der Sechser-Skala. Der Median beträgt 4.37. Der tiefste Wert liegt bei 1.56 (Minimum), der höchste Wert bei 6 (Maximum). Dies entspricht einer Spannweite von 4.44.

Die Verteilungsform der Mittelwertsskala präsentiert sich linksschief bzw. rechtssteil (Schiefenkennwert = $-.70$, $s_S = .08$) und spitzer, als bei einer Normalverteilung zu erwarten wäre (Exzessivitätskennwert = $.75$, $s_E = .16$). Die gegebene Verteilung ist nicht normalverteilt (Kolmogorov-Smirnov-Test; $p < 0.01$). Das Normalverteilungsdiagramm (Q-Q-Diagramm) und das trendbereinigte Normalverteilungsdiagramm belegen dieses Ergebnis optisch. So liegen im Q-Q-Diagramm die Werte nicht auf einer Geraden, sondern weichen am oberen und besonders am unteren Ende von dieser ab. Auch im trendbereinigten Normalverteilungsdiagramm zeigt sich ein klares Abweichungsmuster von der durch den Nullpunkt verlaufenden Horizontalen.²³

Die durch das Stängel-Blatt-Diagramm (*Stem-and-Leaf-Diagramm*) erzeugte Verteilung weist 28 Extremwerte (= 3.1%) im unteren Mittelskalenbereich (< 2.5) aus. Vier Probanden beantworteten alle Items mit dem höchsten Skalenwert. 509 oder 56% der Probanden zeigen einen Mittelwert zwischen 4.00 und 4.99. Im Anhang C2 sind die Abbildungen der Häufigkeitsverteilung der (Roh-)Mittelwerte, die Normalverteilungsdiagramme sowie das Stängel-Blatt-Diagramm aufgeführt.

4.1.1.2 Ausreisser

Ausreisser können die Ergebnisse multivariater Analysen stark beeinflussen. Folgend wird deshalb der Umgang mit den aufgetretenen Ausreißern dokumentiert.

²³ Für nähere Angaben zum Normalverteilungsdiagramm und dem trendbereinigten Normalverteilungsdiagramm siehe Bühl und Zöfel (2005) auf Seite 229f.

Die 28 ausgewiesenen „Extremes“ im Stängel-Blatt-Diagramm erweisen sich im Boxplot alle als Ausreisser (*Outliers*) und nicht als Extremwerte. Der Abstand sämtlicher Ausreisser liegt im Boxplot damit zwischen anderthalb und drei Kastenlängen. Im Histogramm verlaufen die (Roh-)Mittelskalenwerte im unteren Bereich relativ normal und weisen keine Lücke auf, was auf eine grössere Ungereimtheit deuten würde. Ein kleiner Ausschlag lässt sich zwischen 1.75 und 2.25 beobachten.



Abbildung 4.1. Boxplot der Mittelwertsskala Item 1 bis 18 der Rohwerte aus dem Unternehmen A (N = 908).

Bei einer Normalverteilung könnte bei 908 Probanden zufallsbedingt von einem Probanden mit einem z-Wert kleiner -3.00 ausgegangen werden. Ungefähr bei drei weiteren Probanden könnte man einen z-Wert zwischen -2.58 und -3.00 erwarten. In der vorliegenden Stichprobe werden diese Schätzungen klar überschritten. So gibt es neun Fälle mit z-Werten kleiner -3.00 (statt einem Fall), und zwölf (statt drei) Fälle weisen z-Werte zwischen -2.58 und -3.00 auf.

Die Untersuchung der Outliers zeigte, dass mehrere Probanden mit grosser Wahrscheinlichkeit die Antwortskala nicht richtig benutzten. Es gibt zwei Arten von Indizien für diesen Rückschluss. Ersterer gründet in einem widersprüchlichen Antwort- und Ratingverhalten, wie bspw. folgender Kommentar eines Probanden verdeutlicht, der einen Mittelskalenwert von 1.88 (sechsttiefster Wert der ganzen Stichprobe) aufwies: „Ich finde das System genial, in jeder Beziehung...“. Das zweite Indiz für eine unsachgemässe Nutzung der Antwortskala erschliesst sich aus einem anormalen bzw. in Bezug auf die Stichprobe entgegen gesetztem Antwortmuster. Das heisst Items mit den höchsten Mittelwerten (bspw.

Item 7 und 17, siehe Itemanalyse) wurden von diesen Probanden am tiefsten bewertet, während gleichzeitig die Items mit den tiefsten Mittelwerten (bspw. Item 3 oder 8) am höchsten bewertet wurden. Outliers die über mehrere Items hinweg ein abweichendes Antwortverhalten aufwiesen und/oder deren Kommentare offensichtlich nicht mit dem Rating übereinstimmten, wurden aus der Stichprobe eliminiert. Dies war bei 17 der 28 Outliers der Fall. Zehn Outliers wurden in der Stichprobe belassen, weil sich für diese keine eindeutigen Hinweise für einen fälschlichen Gebrauch der Skala fanden. Für zwei Outliers liess sich die Richtigkeit der relativ tiefen Werturteile, mit den von ihnen verfassten Kommentaren belegen. Bei diesen zehn Outliers wird also davon ausgegangen, dass sie repräsentativ/valide Urteile der Stichprobe repräsentieren. Ein Outlier wurde ausgeschlossen, weil er weniger als zwei Drittel der Items beantwortete, was als ein weiteres Ausschlusskriterium gewählt wurde. Er gehörte damit zu einer Gruppe von vier weiteren Probanden, die weniger als 12 der 18 Items beantworteten und deshalb ausgeschlossen wurden.

Da es nicht auszuschliessen war, dass neben den Outliers weitere Probanden die Skala falsch benutzten, wurde die Stichprobe daraufhin untersucht. Wiederum wurde dabei kontrolliert, ob die abgegebenen Kommentare mit dem Ratingverhalten übereinstimmten. Zudem wurden speziell jene Probanden näher betrachtet, die Item 7 oder Item 17 mit dem Skalenwert Eins oder Zwei werteten. Lediglich für einen Probanden liess sich mit hoher Wahrscheinlichkeit eine missverstandene Skalennutzung belegen, weshalb dieser Proband ausgeschlossen wurde.

Ebenfalls von der Stichprobe ausgeschlossen wurden vier Probanden, die alle Items mit dem Maximalwert 6 beantworteten. Für diese Fälle kann nur bedingt davon ausgegangen werden, dass der Fragebogen seriös beantwortet wurde. Möglich wäre bspw., dass einer oder mehrere Probanden aus dieser Gruppe in irgendeiner Form für das Intranet Mitverantwortung trugen und deshalb an einem möglichst guten Abschneiden des Intranets interessiert waren.

Insgesamt führte die Datenexploration zum Ausschluss von 27 Probanden (= 3.0%) aus der Stichprobe. Die Stichprobe für die Itemanalyse umfasst damit 881 Probanden, womit für die Validierung eine gute Basis gewährleistet war.

4.1.2 Missing Data

Der Fragebogen wurde von keinem der Probanden frühzeitig abgebrochen. Da es für Probanden nicht möglich war, ein Item auszulassen (siehe Methodenteil), sind fehlende Werte deshalb als „weiss-nicht“-Antworten zu interpretieren. Die meisten Items verzeichnen

angesichts der grossen Stichprobe relativ wenig fehlende Werte. Tabelle 4.1 gibt darüber eine Übersicht.

Tabelle 4.1

Übersicht der fehlenden Werte pro Item

Item		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
N		878	858	881	877	877	857	880	814	572	851	753	842	845	642	871	839	875	880
Missing	Count	3	23	0	4	4	24	1	67	309	30	128	39	36	239	10	42	6	1
	%	0.3	2.6	0.0	0.5	0.5	2.7	0.1	7.6	35.1	3.4	14.5	4.4	4.1	27.1	1.1	4.8	0.7	0.1

Die Missing-Quote liegt für sieben Items unter 1% und für weitere sieben Items unter 5%. Item 8 verzeichnet 67 fehlende Werte (= 7.6%). Die meisten fehlenden Werte sind für Item 9 mit 309 Fällen (= 35,1%) und Item 14 mit 239 Fällen (= 27,1%) zu beobachten. Bei Item 11 wählten 128 Probanden (= 14.5%) die Antwortkategorie „weiss nicht“. Es überrascht nicht, dass die meisten fehlenden Werte bei Item 9 und 14 auftreten. So kann Item 9 nur beantworten, wer schon einmal eine Person kontaktieren wollte. Item 14 kann für Probanden ebenfalls schwierig zu beantworten sein, wenn sie zuvor noch nie Hilfe benötigten. Im Zusammenhang mit Item 11 erweist sich der Ausdruck „weiss nicht“ im Nachhinein als problematisch. So ist es vorstellbar, dass einige Probanden anstelle eines tiefen Ratingwerts, die Kategorie „weiss nicht“ wählten. Die verhältnismässig hohe Anzahl von fehlenden Werten bei Item 8 könnte darauf hindeuten, dass einige Probanden im Unternehmen A die Suchmaschine für ihre Tätigkeiten im Intranet nicht benötigen.

Ein in Studien oft verwendetes Verfahren im Umgang mit fehlenden Werten sind der paarweise (*pairwise deletion*) und der listenweise Fallausschluss (*listwise deletion*). Im Statistikprogramm SPSS ist oftmals der listenweise Fallausschluss als Standardeinstellung vorgegeben. Aus statistischer Sicht sind diese Vorgehensweisen relativ unbefriedigend (Allison, 2001; Schafer & Graham, 2002). Beim paarweisen Fallausschluss gehen in die Berechnung jeder Statistik andere Personen ein und würde deshalb voraussetzen, dass die „Sub“-Stichproben grundsätzlich gleich sind. Der grösste Nachteil des listenweisen Fallausschlusses veranschaulicht deutlich die vorliegende Stichprobe. So führt die Elimination aller Probanden, die eines der Items mit „weiss nicht“ beantworteten, zu einer Senkung des Stichprobenumfangs von 881 Probanden zu nur noch 405 Probanden. Dies entspricht einer Reduktion von über 50%. Die Methode erweist sich in diesem Fall nicht nur anfällig für Verzerrungen, sie führt auch zu einem erheblichen Verlust an statistischer Power.

Schafer und Graham (2002) empfehlen im Umgang mit fehlenden Werten deshalb, Maximum-Likelihood-basierte Ersetzungen, wie bspw. durch den Expectation-Maximation-Algorithmus (EM). Die Grundidee hinter dem EM-Algorithmus ist, dass in einem ersten Schritt (Expectation-Schritt) auf Basis der Beziehungen im Datensatz Schätzwerte für die fehlenden Werte generiert werden. Im zweiten Schritt (Maximation-Schritt) werden Kovarianzmatrix und Mittelwerte dieser neuen Matrix berechnet. Diese Schleife wird nun solange durchlaufen, bis sich das Modell substantiell nicht mehr verändert (Allison, 2001, S. 19f.). Die Ersetzung von fehlenden Werten durch EM erwies sich als ein valides und reliables Verfahren und dem listenweisen bzw. paarweisen Fallausschluss in vielen Belangen als überlegen (Schafer und Graham, 2002; Allison, 2001).

Die Güte des EM-Algorithmus lässt sich in einem Vergleich der Mittelwerte und der Standardabweichungen illustrieren. Tabelle 4.2 und 4.3 listen die Mittelwerte und Standardabweichungen nach listenweisem Ausschluss, nach Berücksichtigung aller (vorhanden) Werte und nach dem Expectation-Maximation-Algorithmus auf. Der Vergleich zwischen den All Values und den EM Werten zeigt praktisch identische Werte. Kleinere Differenzen finden sich im Vergleich zu den Mittelwerten und Standardabweichungen, die auf Basis des listenweisen Fallausschlusses errechnet wurden. Diese Differenzen dürften auf die verursachten Verzerrungen aufgrund des Ausschlusses von mehr als der Hälfte der Probanden zurückzuführen sein.

Tabelle 4.2

Überblick der Mittelwerte nach Listwise Deletion, All Values und EM

Item	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
Listwise	4.91	4.72	3.44	4.37	4.47	4.36	5.12	3.33	3.82	4.20	3.49	4.19	4.38	4.57	4.84	4.27	4.77	4.43
All Values	4.82	4.77	3.37	4.41	4.35	4.34	5.15	3.28	3.88	4.21	3.25	4.25	4.38	4.64	4.90	4.27	4.90	4.49
EM	4.82	4.77	3.37	4.41	4.35	4.35	5.15	3.30	3.93	4.21	3.27	4.25	4.38	4.66	4.89	4.27	4.90	4.49

Tabelle 4.3

Überblick der Standardabweichungen nach Listwise Deletion, All Values und EM

Item	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
Listwise	1.08	0.92	1.25	1.00	1.07	1.15	0.97	1.37	1.26	1.06	1.69	1.13	1.02	1.07	0.96	1.03	1.11	0.99
All Values	1.17	0.94	1.27	0.99	1.11	1.17	1.06	1.43	1.28	1.11	1.77	1.15	1.08	1.08	0.96	1.07	1.11	0.99
EM	1.17	0.94	1.27	0.99	1.11	1.17	1.06	1.43	1.28	1.11	1.76	1.15	1.08	1.07	0.96	1.07	1.11	1.00

Aufgrund der reliableren und valideren Resultate, die durch den EM-Algorithmus zu erwarten sind sowie dem deutlichen Gewinn an statistischer Power, wird die folgende Itemanalyse auf Basis der EM-Datenmatrix durchgeführt.

4.3 Itemanalyse

4.3.1 Statistische Kennwerte

Der Skalenmittelwert liegt bei 4.32 mit einer Standardabweichung von .69. Der Median ist mit 4.35 knapp höher. Während der Skalenmittelwert nach dem Ausschluss der 27 Items praktisch gleich bleibt, nimmt die Streuung erwartungsgemäss ab. Der tiefste Mittelskalenwert ist 1.94, der höchstgemessene 5.94. Abbildung 4.2 gibt das Histogramm der Mittelwertskala wider.

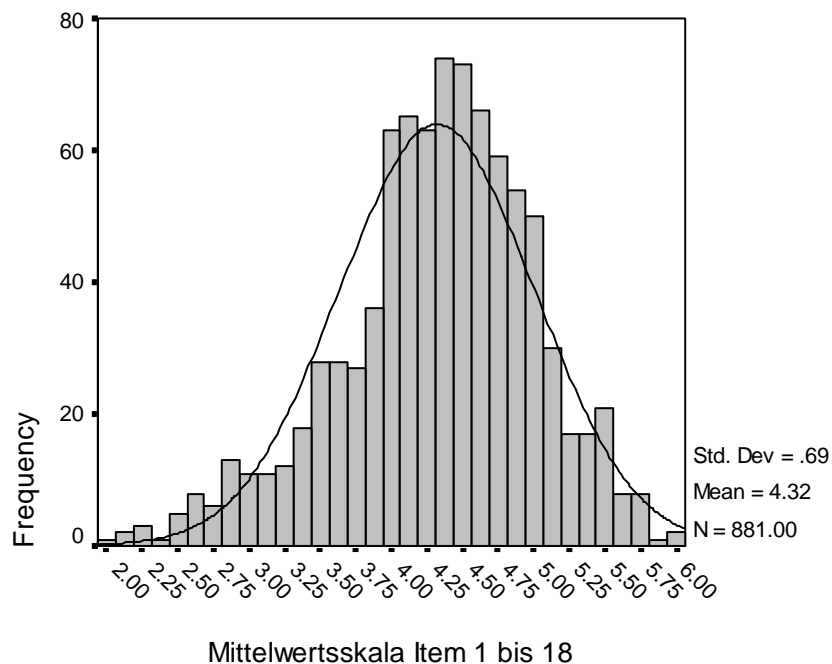


Abbildung 4.2. Häufigkeitsverteilung der Mittelwertskala der Items 1 bis 18.

Die Werteverteilung der Mittelwertsskala ist rechtsgipflig (Schiefenkennwert = $-.51$, $s_s = .08$) und zeigt einen leicht schmalen Verteilungsgipfel (Exzessivitätskennwert²⁴ = $.38$,

²⁴ SPSS berechnet den Exzess (Kurtosis) über das vierte zentrale Moment mit einer Korrekturformel. Nach dieser Berechnung gilt so nicht der Wert 3 für einen normalen Exzess, sondern der Wert Null. Für die

$s_E = .17$). Nach Lienert und Raatz (1998, S. 148) können bei Stichproben von über 400 Probanden Schiefenkennwerte zwischen $-.50$ und $+.50$ trotz statistisch signifikantem Schiefenkennwert als praktisch unbedeutend interpretiert werden. Diese Faustregel wird hier nur sehr knapp verpasst. Der Kolmogorov-Smirnov-Test²⁵ und der Shapiro-Wilk-Test attestieren der Stichprobe keine Normalverteilung ($p < .01$).

Die Mittelwerte der Items bewegen sich zwischen dem tiefsten Wert des Items 11 mit 3.27 und dem höchsten Wert des Items 7 mit 5.15. Dies ergibt eine Spannweite von 1.88. Die Streuung der Items liegt zwischen .93 (Item 15) und 1.64 (Item 11). Die durchschnittliche Streuung beträgt 1.12. Vier Items (2, 4, 14 und 15) besitzen eine Streuung kleiner Eins. Tabelle 4.4 gibt einen Überblick über die Mittelwerte, Standardabweichungen und andere Kennwerte der getesteten Items.

Tabelle 4.4

Item-Übersicht: Anzahl Probanden (N), Mittelwerte (Mean) und Standardabweichungen (SD), Median- und Modalwert sowie die Schiefenkennwerte (S), Exzessivitätskennwerte (E) und varianzberücksichtigte Schwierigkeitsindizes (p_{mv}) der einzelnen Items

	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Median</i>	<i>Mode</i>	<i>S</i>	<i>E</i>	<i>p_{mv}</i>
1 Arbeitsrelevante Inhalte	881	4.82	1.17	5.00	5	-1.14	0.98	.683
2 Aktuelle Informationen	881	4.77	0.93	5.00	5	-0.84	1.04	.657
3 Gestaltung der Informationsarchitektur	881	3.37	1.27	4.00	4	-0.07	-0.73	.360
4 Verständlichkeit der Informationen	881	4.41	0.99	5.00	5	-0.86	0.93	.567
5 Sicherheit im Umgang	881	4.35	1.11	5.00	5	-0.72	0.18	.558
6 Erleichterung der Kommunikation	881	4.35	1.16	4.66	5	-0.62	-0.08	.560
7 Zugangsmöglichkeiten	881	5.15	1.06	5.00	6	-1.72	3.41	.767
8 Qualität Suchmaschine	881	3.30	1.38	3.61	4	-0.14	-0.90	.356
9 Nachfrage und Feedbackmöglichkeiten	881	3.93	1.06	4.00	4	-0.41	0.41	.463
10 Effizientere Arbeitsweisen	881	4.21	1.10	4.00	4	-0.54	0.06	.526
11 Wissen Dokumente aufschalten	881	3.27	1.64	3.16	1	0.11	-1.09	.380
12 Geschwindigkeit	881	4.25	1.13	4.29	5	-0.80	0.39	.538
13 Einfache Bedienbarkeit	881	4.38	1.07	5.00	5	-0.81	0.65	.564
14 Hilfe bei Problemen	881	4.66	0.95	4.92	5	-1.01	1.54	.629
15 Aktuelle Firmenneuigkeiten	881	4.89	0.96	5.00	5	-1.11	1.67	.690
16 Angemessene Informationspräsentation	881	4.27	1.05	4.00	5	-0.67	0.44	.537
17 Vertrauen in Informationen	881	4.90	1.11	5.00	5	-1.21	1.36	.700
18 Globalzufriedenheit	881	4.49	1.00	5.00	5	-0.80	0.66	.586

Anmerkung. Missing values = EM; Standardfehler der Schiefe $s_S = .082$; Standardfehler des Exzesses $s_E = .165$.

Interpretation gleich bleibt sich, dass kleinere Werte auf eine breitgipflige Verteilung und grössere Werte auf eine schmalgipflige Verteilung hin deuten (Leonhart, 2004, S. 53).

²⁵ Leonhart (2004, S. 180) weist darauf hin, dass der Kolmogorov-Smirnov-Test bei grossen Stichproben von über 500 Personen die Gefahr birgt, zu sensitiv zu reagieren und bereits unbedeutende Abweichungen signifikant werden können.

Der Modalwert liegt für 12 der 18 Items beim Skalenwert 5, viermal liegt er beim Skalenwert 4 und je einmal ist der Skalenwert 1 bzw. 6 derjenige Wert, der am häufigsten vorkommt.

Die Häufigkeitsverteilungen der Items weisen für die meisten Items eine rechtsasymmetrische Verteilung auf. Ausser Item 11 besitzen alle Items einen negativen Schiefenkennwert, wobei einige allerdings praktisch Null sind. Bei Item 7 kommt es zu einem Deckeneffekt. Die Items 1 und 17 nähern sich einem Deckeneffekt an. Die Mehrheit der Items haben eine eingipflige und in der Tendenz schmalgipflige (positive Exzessivitätskennwerte) Verteilungsform. Bei den Items 3, 8 und 11 wurde die Skala am besten ausgenutzt, was sich auch in den negativen Exzessivitätskennwerten ausdrückt. Die Histogramme der Items sind im Anhang C3 nachschlagbar.

4.3.2 Schwierigkeitsindizes

Der Schwierigkeitsindex einer Aufgabe (Popularitätsindex einer Frage) ist definiert durch die prozentuale Häufigkeit, mit der sie von einer repräsentativen Stichprobe von Probanden richtig bzw. in Schlüsselrichtung beantwortet wird (Lienert & Raatz, 1998, S. 58).

Für mehrfach abgestufte Itemantworten kann der Schwierigkeitsindex durch den Quotienten der erreichten Wertepunkte zu den erreichbaren Wertepunkten ausgedrückt werden. Diese Art der Berechnung hat den Nachteil, dass nur die Mittelwerte in die Berechnung einfließen. Zwei Items können so dieselben Schwierigkeitsindizes aufweisen, jedoch sehr unterschiedliche Varianzen besitzen. Dies ist deshalb nachteilig, weil bei zwei Items mit gleichem Schwierigkeitsindex jenes Item besser differenzieren würde, welches über eine grössere Varianz verfügt. Fisseni (1997, S. 45) empfiehlt deshalb in Anlehnung an Dahl, Wagner und Baumgärtel²⁶ eine Berechnungsweise des Schwierigkeitsindex, welche die Varianz berücksichtigt. Sie sieht vor, den Schwierigkeitsindex über den Quotienten der Summe von den quadrierten Item-Scores zur Summe der quadrierten erreichbaren Wertepunkte zu bestimmen. Für diese Itemanalyse werden aus diesem Grund die varianzberücksichtigten Schwierigkeitsindizes zur Beurteilung der Items herangezogen. Es sei jedoch darauf hingewiesen, dass diese Berechnungsweise zur Folge hat, dass die Schwierigkeitsindizes der getesteten Items im Durchschnitt um .16 tiefer ausfallen. Abbildung 4.3 veranschaulicht diesen Effekt.

²⁶ Fisseni (1997, S. 45) macht keine Quellenangaben dazu.

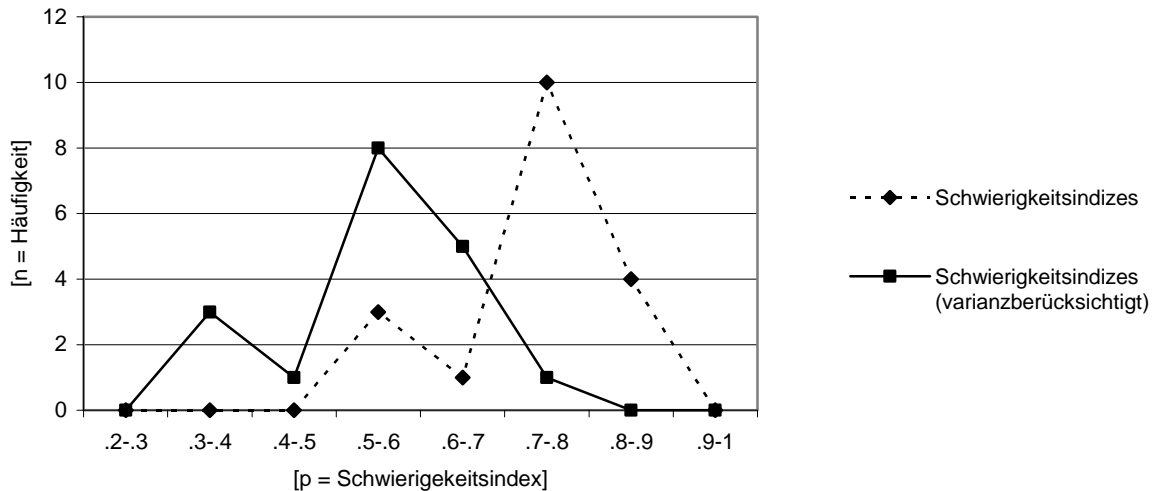


Abbildung 4.3. Häufigkeitsverteilung der Schwierigkeitsindizes der ersten Testversion mit und ohne Berücksichtigung der Varianz.

Wie aufgrund der gehäuft vorkommenden rechtsasymmetrischen Verteilungen bereits absehbar, fallen die Schwierigkeitsindizes, insbesondere ohne Berücksichtigung der Varianz, relativ hoch aus. In Abbildung 4.3 zeigt sich dies in einem Ausschlag im Bereich .70 bis .80. Varianzberücksichtigt streuen die Schwierigkeitsindizes zwischen .36 (Item 8) und .77 (Item 7). 13 von 18 Items weisen Schwierigkeitsindizes im Bereich zwischen .50 und .70 auf. Ein Item besitzt einen Schwierigkeitsindex grösser .70 (Item 7), 4 Items liegen zwischen .30 und .50 (Items 3, 8, 9 und 11). Die durchschnittliche Schwierigkeit über alle Items hinweg beträgt .56 ($SD = .12$). Die varianzberücksichtigten Schwierigkeitsindizes sind in Tabelle 4.4 aufgelistet.

Aus theoriegeleiteten Blickwinkeln werden Schwierigkeitsindizes von .50 favorisiert, weil solche Items in einer Stichprobe mehr Unterscheidungen zwischen je zwei Probanden treffen (Lienert & Raatz, 1998, S. 105) und sie hohe Trennschärfen ermöglichen (Fisseni, 1997, S. 46). Fisseni (1997, S. 46) wendet allerdings ein, dass ein Test, der nur aus Items mit $p = .50$ konstruiert wäre und jeweils dieselben Personen die Löser bzw. die Nichtlöser wären, die Stichprobe immer in zwei Klassen (Löser und Nichtlöser) teilen würde. Dies würde der Absicht widersprechen, eine Skala zu erstellen, die mehr als zwei Ausprägungen misst. Für die Praxis schlägt Fisseni (1997, S. 47) deshalb vor, sowohl Items mit $p = .50$ als auch Items mit $.50 > p < .50$ in einem Test aufzunehmen. Im Normalfall sollten die Schwierigkeitsindizes jedoch innerhalb der Grenzen zwischen .20 und .80 bleiben (Bühl & Zöfel, 2005, S. 457). Dies ist, zumindest was die varianzberücksichtigten Schwierigkeitsindizes betrifft, bei den

vorliegenden Items der Fall.

Die Verteilung der Schwierigkeitsindizes sollte sich auf der Schwierigkeitsskala in dem Bereich häufen, in welchem vom Test eine besonders gute Differenzierung erwünscht wird (Lienert & Raatz, 1998, S. 115). In Abbildung 4.3 lässt sich eine Häufung der Schwierigkeitsindizes im Bereich .50 bis .70 beobachten. Wie in der Diskussion noch näher besprochen, erscheint dies für den Fragebogen durchaus sinnvoll.

Grundsätzlich lässt sich die Frage über Eignung oder Nichteignung eines Items auf Basis von Schwierigkeitsindizes nur im Zusammenhang mit Trennschärfe und Homogenität beantworten, wobei ein solcher Entscheid letztendlich immer auch vom Zweck eines Tests abhängt.

4.3.3 Trennschärfe

Fisseni (1997) hält die Trennschärfe einer Aufgabe, charakterisiert durch ihren Trennschärfekoeffizienten, für den wichtigsten Itemkennwert. Die Trennschärfe „klärt die Position eines Items im Verband der anderen Items, indem sie einen Index liefert, der angibt, wie weit die ‚Menge der Löser‘ über alle Items hinweg identisch bleibt.“ (zit. nach Fisseni, 1997, S. 47). In der Regel wird die Trennschärfe anhand des Test-Scores (inneres Kriterium) ermittelt. Vereinfacht definiert sich die Trennschärfe somit als die (biserielle) Korrelation zwischen Item- und Test-Score.

Bei mehrfach abgestuften Itemantworten empfiehlt sich die Berechnung der Trennschärfe als Produkt-Moment-Korrelation. Tabelle 4.5 listet die Trennschärfen und Cronbach's Alpha der einzelnen Items auf.

Tabelle 4.5

Trennschärfen und Cronbach's Alpha der ersten Testversion (N = 881; Missing values = EM)

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
1	68.4541	119.0654	.4977	.8818
2	68.5123	120.6635	.5651	.8799
3	69.9073	115.9674	.5659	.8792
4	68.869	118.3131	.6396	.8772
5	68.9323	118.783	.5399	.8802
6	68.9332	115.8394	.6367	.8765
7	68.1298	121.6752	.4403	.8837
8	69.9798	120.0139	.3683	.8882
9	69.3476	121.3826	.4553	.8832
10	69.065	115.5694	.6931	.8747
11	70.0081	119.6951	.2994	.8945
12	69.0252	117.7445	.5724	.8790
13	68.9015	117.5206	.6240	.8774
14	68.6174	119.8015	.5960	.8789
15	68.3847	119.9221	.5837	.8792
16	69.0059	116.7764	.6700	.8758
17	68.3818	119.5799	.5082	.8814

$\alpha_{\text{item1-17}} = .8869$

Das Item 18 wurde im Fragebogen als Overall-Item eingesetzt und später zur Messung der Validität bei gezogen. Um ein „künstliches“ Anheben der Werte zu vermeiden, wurde Item 18 deshalb von den Trennschärferechnungen ausgeschlossen. Ein Vergleich der Trennschärfen mit und ohne Item 18 zeigt für die Trennschärfen (inkl. Item 18) tendenziell leicht höhere Werte. Analog verhält es sich für das Cronbach Alpha: $\alpha_{\text{item1-18}} = .90$, $\alpha_{\text{item1-17}} = .89$.

Ein tiefer Trennschärfekoeffizient bedeutet, dass sich ein Item nicht eignet, um zwischen Probanden mit hoher Benutzerzufriedenheit und Probanden mit tiefer Benutzerzufriedenheit zu unterscheiden (Linert & Raatz, 1998, S. 78). Üblicherweise wird die Trennschärfe als Gütekriterium so interpretiert, dass ein Item umso besser ist, je grösser seine Trennschärfe ausfällt (Rost, 2004, S. 370).²⁷ Die Trennschärfekoeffizienten bewegen sich zwischen .30 (Item 11) und .70 (Item 10). Die durchschnittliche Trennschärfe beträgt .54 ($SD = .11$). Fünf Items (1, 7, 8, 9 und 11) besitzen eine Trennschärfe von kleiner .50. Die meisten

²⁷ Rost (2004, S. 370) verweist aber auf den problematischen Aspekt, dass bei mehrkategorialen Itemantworten ein Item umso trennschärfer werden kann, je seltener die mittleren Antwortkategorien verwendet werden. Grafisch würde sich dies durch eine relativ steile bzw. flache Itemfunktion ausdrücken, wobei Itemfunktion hier verstanden wird als Funktion des Erwartungswertes der Antwortvariable in Abhängigkeit von der Personeneigenschaft. Siehe hierzu Rost (2004, Seite 215–225).

Items sind im Bereich zwischen .50 und .70 angesiedelt. Kein Item verfügt über eine Trennschärfe von über .70. Nach Borg und Staufenbiel (1997) gilt eine Trennschärfe von .30 als unterste akzeptable Grenze für eine Variable. Item 11 fällt mit einem Trennschärfekoeffizienten von .30 auf diese Schranke. Ansonsten lassen sich die Trennschärfen in einen mittleren und genügenden Bereich einordnen.

Für die Itemselektion eines homogenen Tests schlagen Fisseni (1997) sowie Lienert und Raatz (1998) die Berechnung von Selektionskennwerten vor. Diese können auf Basis der Trennschärfe und des Schwierigkeitsindizes berechnet werden und gelten als umso besser, je höher sie ausfallen.²⁸ Die Grundidee dahinter ist, Items mit niedrigen Trennschärfen zu eliminieren, auch wenn sie günstige Schwierigkeitsindizes besitzen bzw. Items mit hohen Trennschärfen zu behalten, auch wenn sie relativ ungünstige Schwierigkeitsindizes aufweisen (positive Selektion). Wenn zwei Items dieselbe Trennschärfe besitzen, empfiehlt der Selektionskennwert entweder schwierigere oder leichtere Items einzubeziehen.²⁹ Von den getesteten Items weisen die Items 8 (mit einem Selektionskennwert von .39), 11 (.31) und 9 (.46) die tiefsten Werte auf. Alle anderen Items haben Selektionskennwerte von über .50 und unter .70.

4.3.4 Homogenität

Die Items des *Intranet Satisfaction Questionnaire* sollen alle dasselbe Konstrukt ‚Benutzerzufriedenheit‘ erfassen. Die im Fragebogen eingesetzten Items erfassen jedoch zwangsläufig unterschiedliche Merkmalsfacetten. Die Frage nach der Homogenität eines Tests soll klären, inwieweit sich diese Facetten überlappen. Die Homogenität ist also ein Mass dafür, inwiefern ein Test dieselbe Eigenschaft erfasst. Es gibt mehrere Konzepte zur Untersuchung der Homogenität. Folgend wird sie anhand der Interkorrelation und der Faktorenanalyse besprochen.

²⁸ Für weitere Angaben zum Selektionskennwert oder dessen Berechnungsformel siehe Fisseni (1997, S. 60–62) oder Lienert und Raatz (1998, S. 129f.).

²⁹ Dies lässt sich bspw. beim Vergleich zwischen Item 1 und Item 9 mit relativ ähnlichen Trennschärfen, aber unterschiedlicher Schwierigkeit beobachten. Nach dem Selektionskennwert wird Item 1, welches einen relativ hohen Schwierigkeitsindex aufweist, gegenüber Item 9 mit mittlerer Schwierigkeit, leicht favorisiert.

4.3.4.1 Interkorrelation

Im Sinne der Interkorrelation gilt ein Test dann als homogen, wenn seine Items relativ hoch miteinander korrelieren. Inhaltlich bedeutet dies, dass die Items, wenn auch nicht identische, so doch wenigstens ähnliche Merkmalsfacetten repräsentieren.

Tabelle 4.6

Interkorrelationsmatrize und Homogenitätsindizes der Items 1 bis 18 der ersten Testversion

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1																	
2	.430	1																
3	.241	.293	1															
4	.320	.469	.482	1														
5	.436	.336	.368	.406	1													
6	.425	.372	.427	.456	.434	1												
7	.310	.396	.165	.338	.250	.288	1											
8	.122	.235	.414	.247	.110	.264	.160	1										
9	.125	.302	.319	.332	.195	.288	.207	.307	1									
10	.502	.412	.431	.464	.445	.543	.331	.312	.279	1								
11	.103	.085	.251	.159	.309	.243	.056	.127	.253	.236	1							
12	.231	.314	.374	.407	.298	.392	.358	.218	.297	.446	.241	1						
13	.267	.303	.485	.492	.382	.458	.293	.257	.303	.474	.252	.487	1					
14	.339	.389	.317	.411	.310	.386	.340	.232	.369	.425	.203	.440	.456	1				
15	.389	.491	.297	.454	.347	.440	.378	.164	.285	.440	.127	.345	.413	.440	1			
16	.403	.363	.439	.509	.345	.455	.286	.340	.302	.568	.252	.465	.474	.444	.416	1		
17	.378	.461	.186	.353	.308	.334	.338	.161	.308	.413	.085	.317	.285	.382	.410	.433	1	
18	.420	.498	.584	.565	.468	.528	.371	.436	.386	.614	.221	.518	.600	.533	.503	.619	.487	1
H	.320	.362	.357	.404	.338	.396	.286	.241	.286	.431	.188	.362	.393	.377	.373	.418	.332	.491

Anmerkung. Alle Korrelationen signifikant ($p < 0.01$); H = Homogenitätskoeffizient
N = 881; Missing values = EM

Die Korrelationsmatrix weist für alle Items signifikante Korrelationen ($p < 0.01$) auf. Weiter kommen keine negativen Korrelationen zum Vorschein, was einer homogenen Skala widersprechen würde. Für die Stärke des Zusammenhangs sind die Beträge der Korrelationskoeffizienten massgebend, welche hier relativ moderat ausfallen.

Die Items 10, 13 und 16 weisen mit Korrelationen um .60 die höchsten Zusammenhänge mit Global-Item 18 auf. Etwas geringere, aber immer noch mittlere Korrelationen mit Global-Item 18 finden sich für die Items 3, 4, 6, 12, 14 und 15. Mit Abstand am tiefsten korreliert das Item 11 mit dem Global-Item ($r = .22$). Ebenfalls relativ geringe Korrelationen mit Item 18 zeigen die Items 7 ($r = .37$) und 9 ($r = .39$).

Die Homogenitätsindizes³⁰ liegen zwischen .19 (Item 11) und .49 (Item 18). Beim Global-Item 18 handelt es sich damit um das homogenste Items, welches im Mittel mit allen anderen Items am höchsten korreliert. Diese Charakteristik wurde von Item 18 in der Funktion als Global-Item erwartet. Item 11 hat nicht nur den tiefsten Homogenitätsindex, es weist auch nur eine Korrelation von unter .30 mit dem Global-Item auf. Dies deutet stark darauf hin, dass Item 11 nur wenige Überlappungen mit den anderen Items hat und es sich um ein heterogenes Item handelt, welches in der untersuchten Stichprobe wenig zur Aufklärung von Benutzerzufriedenheit beiträgt. Ebenfalls einen relativ tiefen Homogenitätskoeffizienten hat Item 8, welches ausser mit Item 3 und Item 18 nur sehr geringe Zusammenhänge aufweist. Item 7 und Item 9 zeigen keine Korrelationen über .40. Dies drückt sich auch in verhältnismässig tiefen Homogenitätskoeffizienten aus. Der Homogenitätsindex für die Skala³¹ beträgt .35.

Ein Grund für die relativ moderaten Interkorrelationen könnte in der teilweisen Unvereinbarkeit von Reliabilität und Validität liegen (Verdünnungsparadoxon). Die Items wurden für die Skala so ausgewählt, dass sie das komplexe Konstrukt Benutzerzufriedenheit möglichst valide und entsprechend breit erfassen. Dies steht zum Teil im Widerspruch, möglichst homogene Items zu verwenden, die zu einer hohen Reliabilität führen und entsprechend hohe Interkorrelationen ermöglichen.

4.3.4.2 Faktorenanalyse

Der Kaiser-Meyer-Olkin-Koeffizient (KMO) gibt Auskunft darüber, ob die Durchführung einer Faktorenanalyse aufgrund der vorliegenden Korrelationen zu rechtfertigen ist (Bühner, 2004, S. 160). Er beträgt für die vorliegende Stichprobe .93 und kann damit als sehr gut betrachtet werden. Auch der Bartlett-Test auf Sphärizität fällt signifikant ($p < 0.01$) aus, was angesichts der grossen Stichprobe nicht sehr überrascht. Ebenso attestieren die MSA-Koeffizienten der Anti-Image-Matrix den Testkennwerten eine gute Eignung für eine Faktorenanalyse. Wie bereits beschrieben, sind die verwendeten Variablen jedoch nicht normalverteilt. Da unterschiedliche Verteilungen die Korrelationshöhe der Items beeinflussen können, sind die Ergebnisse von Faktorenanalysen unter diesen Voraussetzungen deshalb mit Vorsicht zu interpretieren (Bühner, 2004, S. 138).

³⁰ Der Homogenitätsindex wird hier nach dem Vorschlag von Lienert und Raatz (1997, S. 99) operationalisiert. Er entspricht für eine Aufgabe in deren mittleren Interkorrelation mit den (n - 1) übrigen Aufgaben.

³¹ Der Homogenitätsindex der Skala entspricht dem Mittelwert der Homogenitätskoeffizienten der einzelnen Items.

Im Sinne der Faktorenanalyse sind Items dann homogen, wenn sie gemeinsam auf demselben Faktor hoch laden. Bei Homogenität ist eine Ein-Faktor-Lösung oder zumindest ein varianzstarker erster Faktor zu erwarten (Lienert & Raatz, 1997, S. 113). Die Hauptkomponentenanalyse extrahierte nach der Varimax-Methode für die 17 Items drei Faktoren mit einem Eigenwert > 1.0 (Kaiser-Guttman Kriterium). Die drei Faktoren erklären zusammen rund 53% der Varianz, wobei vom ersten Faktor ca. 38,5%, vom zweiten Faktor ca. 8% und vom dritten Faktor noch ca. 6,5% der Varianz erklärt werden. Der Eigenwertverlauf in Abbildung 4.4 veranschaulicht den varianzstarken ersten Faktor. Eine Tabelle mit der erklärten Gesamtvarianz ist im Anhang C nachschlagbar.

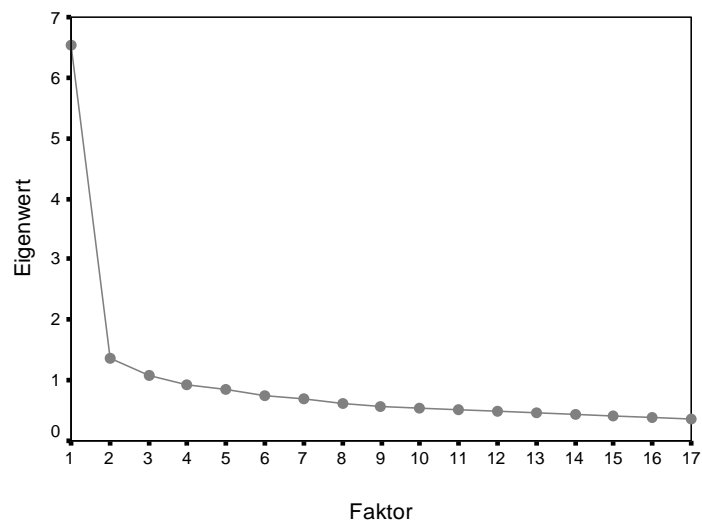


Abbildung 4.4. Screeplot der Faktorenanalyse von Items 1 bis 17 der ersten Testversion.

Nach dem Scree-Test von Cattell sind jene Faktoren als bedeutsam zu betrachten, deren Eigenwerte vor dem Knick liegen (Bortz, 1999, S. 528). In diesem Fall könnte somit für eine Ein-Faktoren-Lösung argumentiert werden, was die Annahme einer homogenen Skala prinzipiell unterstützt.

Für die Bildung einer homogenen Skala kann die Faktorenanalyse Argumente für die Itemselektion liefern, indem Items bevorzugt werden, die auf dem Hauptfaktor hoch laden. Bei einer Mehr-Faktoren-Lösung sollte jedoch in Betracht gezogen werden, anstelle eines womöglich inhomogenen Tests, homogene Subtests zu konstruieren. Die Faktorenanalyse kann dann als Grundlage zur Aufgabenselektion für die Subtests dienen (Lienert & Raatz, 1997, S. 113).

Die rotierte Komponentenmatrix in Tabelle 4.7 gibt Aufschluss darüber, wie hoch die einzelnen Items auf den drei Faktoren laden.

Tabelle 4.7

Rotierte Komponentenmatrix der Items 1 bis 17 der ersten Testversion (N = 881, Missings values = EM)

Item	Component		
	1	2	3
1 Arbeitsrelevante Inhalte	.62	.44	-.17
2 Aktuelle Informationen	.72	.11	.21
3 Gestaltung der Informationsarchitektur	.10	.51	.56
4 Verständlichkeit der Informationen	.48	.37	.39
5 Sicherheit im Umgang	.33	.72	-.06
6 Erleichterung der Kommunikation	.40	.57	.23
7 Zugangsmöglichkeiten	.65	.00	.17
8 Qualität Suchmaschine	.07	.06	.72
9 Nachfrage und Feedbackmöglichkeiten	.22	.07	.65
10 Effizientere Arbeitsweisen	.49	.54	.25
11 Wissen Dokumente aufschalten	-.19	.65	.21
12 Geschwindigkeit	.35	.35	.44
13 Einfache Bedienbarkeit	.29	.50	.44
14 Hilfe bei Problemen	.50	.25	.40
15 Aktuelle Firmenneuigkeiten	.66	.24	.19
16 Angemessene Informationspräsentation	.43	.44	.41
17 Vertrauen in Informationen	.69	.09	.16

Die Items, deren Ladungen in Tabelle 4.7 fett markiert sind, lassen sich relativ eindeutig einem der drei Faktoren zuordnen. Für mehr als die Hälfte der Items ist jedoch eine klare Zuteilung nicht möglich, da die Items auf zwei oder gar drei Faktoren vergleichsweise hoch laden. Nach Fürntratt (1969, zit. nach Bortz, 1999, S. 335f.) können Variablen mit mittleren Ladungen auf mehreren Faktoren dann einem Faktor zugewiesen werden, wenn der Quotient aus quadrierter Ladung und Kommunalität den Wert 0.5 nicht unterschreitet. Die Items 1, 3, 6 und 14 könnten auf diese Weise zu dem Faktor, auf welchem sie die höchste Ladung aufweisen, gezählt werden (siehe *kursive* Formatierungen). Ohne Zuteilung blieben damit aber immer noch die Items 4, 10, 12, 13 und 16.

Da mehrere Items mittlere Ladungen auf zwei oder drei Faktoren aufweisen, scheint es denkbar, dass die Faktoren nicht völlig unabhängig voneinander sind. Es wurde deshalb auch eine schiefwinklige Rotation (Oblimin direkt; Eigenwert > 1; Delta-Wert = 0) getestet. Die Ergebnisse verbesserten die Interpretierbarkeit der Faktorenstrukturen jedoch nicht. Die Drei-Faktoren-Struktur scheint somit nur sehr bedingt geeignet, die Daten sinnvoll zu reduzieren.

Eine andere Methode der Faktorenextraktion schlägt Horn mit der Parallelanalyse vor (Horn, zit. nach Bortz, 1999, S. 529).³² Grundidee ist der Vergleich des Eigenwertverlaufs der empirisch ermittelten Korrelationsmatrix mit dem Eigenwertverlauf einer aus normalverteilten Zufallsvariablen ermittelten Korrelationsmatrix. Als relevante Faktoren gelten jene, deren Eigenwerte über dem „zufälligen“ Eigenwertverlauf liegen (O’Connor, 2000).³³ Von den drei Faktoren mit einem Eigenwert grösser Eins ist dies für zwei der Faktoren der Fall. Der Output der Parallelanalyse ist im Anhang C4 abgebildet.

Aufgrund dieses Resultats wurde eine Faktorenanalyse mit erzwungener Zwei-Faktoren-Lösung (Hauptkomponentenanalyse, Varimax) getestet. Die rotierte Komponentenmatrix ist in der Tabelle 4.8 aufgeführt.

Tabelle 4.8

Rotierte Komponentenmatrix der Hauptkomponentenanalyse von Item 1 bis 17 mit erzwungener Zwei-Faktoren-Lösung (N = 881, Varimax, Missings values = EM)

Item	Component	
	1	2
1 Arbeitsrelevante Inhalte	.70	.08
2 Aktuelle Informationen	.72	.15
3 Gestaltung der Informationsarchitektur	.18	.74
4 Verständlichkeit der Informationen	.53	.48
5 Sicherheit im Umgang	.48	.36
6 Erleichterung der Kommunikation	.51	.49
7 Zugangsmöglichkeiten	.63	.07
8 Qualität Suchmaschine	.06	.59
9 Nachfrage und Feedbackmöglichkeiten	.21	.53
10 Effizientere Arbeitsweisen	.58	.48
11 Wissen Dokumente aufschalten	-.05	.59
12 Geschwindigkeit	.39	.52
13 Einfache Bedienbarkeit	.37	.62
14 Hilfe bei Problemen	.53	.41
15 Aktuelle Firmenneuigkeiten	.69	.22
16 Angemessene Informationspräsentation	.49	.55
17 Vertrauen in Informationen	.69	.11

Versucht man die Faktoren anhand der Items mit den höchsten Ladungen zu interpretieren, könnte eine mögliche Interpretation der Faktoren dahin gehen, dass der erste

³²Nach Bühner (2004, S. 143) wird die Parallelanalyse als Extraktionsmethode von verschiedenen Autoren favorisiert.

³³Das Statistikprogramm SPSS 11.5 bietet die Berechnung von Parallelanalysen standardmässig nicht an. O’Conner (2000) bietet jedoch eine SPSS-Syntax an, mit der Parallelanalysen berechnet werden können. Für die hier berechnete Parallelanalyse wurde diese Syntax verwendet. Für weitere Angaben zur Syntax, zur Berechnung der zufälligen Eigenwerte sowie deren Interpretation, sei auf den Artikel von O’Connor (2000) verwiesen. Für einen schnellen Überblick bietet sich auch Bühner (2004) an.

Faktor die subjektiv wahrgenommene Qualität der Informationen und der zweite Faktor die subjektiv wahrgenommene Nutzbarkeit/Tauglichkeit des Intranets erfasst. Der erste Faktor umfasst in diesem Sinne, inwieweit die abrufbaren Informationen im Intranet die Erwartungen der Benutzer erfüllen, wie z.B. in Bezug auf Arbeitsrelevanz, Aktualität, Informationsbedürfnis (News), Verständlichkeit oder Vertrauen. Der zweite Faktor umschließt die Erwartungen der Benutzer in Bezug darauf, wie sie ihre Ziele/Absichten mit dem Intranet erreichen können wie bspw. durch einfache Bedienbarkeit, schnelles Auffinden oder leichte Orientierung. Diese Auffassung der Faktoren ähnelt der Usability-Definition der ISO-Norm 9241, die zwischen Effektivität und Effizienz unterscheidet (siehe Theorieteil). Zumindest theoretisch scheint eine solche Zwei-Faktoren-Struktur plausibel, aufgrund der vielen Items mit mittleren Ladungen auf beiden Faktoren erweist sich aber auch diese Faktorenstruktur als unbefriedigend. Das Komponentendiagramm illustriert, wie keine hinreichende Abbildung der Punktwolke durch zwei (unabhängige) Faktoren gelingt.

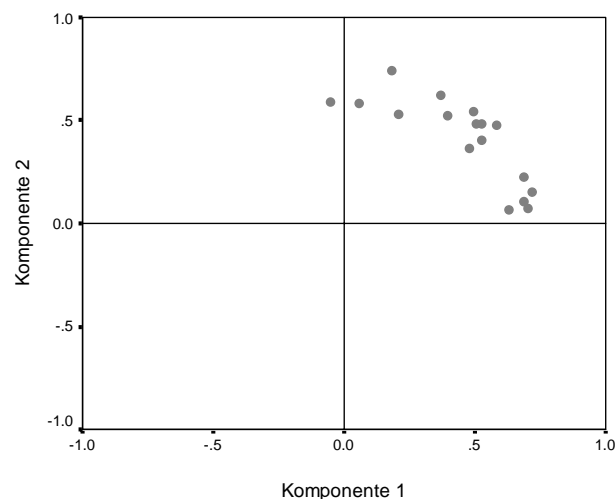


Abbildung 4.5. Komponentendiagramm im rotierten Raum nach erzwungener Zwei-Faktoren-Struktur (Varimax, Item 1 bis 17 der ersten Testversion, N = 881, Missing Values = EM).

Für eine Interpretation der Faktoren optimal wären Punkte, die eng an den Achsen und möglichst weit weg vom Nullpunkt liegen würden. Im Punkteschwarm des Komponentendiagramms lassen sich unten rechts fünf Items erkennen, die relativ eindeutig auf dem ersten Faktor laden. Es handelt sich dabei um die Items 1, 2, 7, 15 und 17. Die vier Items, die etwas klarer auf dem zweiten Faktor laden, lassen sich oben links des Punkteschwarms identifizieren (Item 3, 8, 9 und 11). Zwischen diesen beiden Grüppchen bildet sich eine Art „Knäuel“, der sich von beiden Achsen etwa gleich weit entfernt befindet. Bei den erwähnten Items, die relativ klar auf einem der Faktoren laden, fällt auf, dass die

Fünfer-Gruppe aus den Items mit den höchsten Schwierigkeitsindizes besteht, während sich die Vierer-Gruppe aus den Items mit den tiefsten Schwierigkeitsindizes zusammensetzt. Lienert und Raatz (1997, S. 113) weisen darauf hin, dass bei Items mit unregelmässig verteilten Schwierigkeitsindizes, Schwierigkeitsfaktoren auftreten können, die die Interpretation der Faktorenstruktur erschweren können.

Auch hier führte eine schiefwinklige Rotation (Oblimin direkt, Delta-Wert = 0) zu keiner Verbesserung im Sinne einer klareren Struktur.

Theoretisch plausibel bleibt grundsätzlich eine Ein-Faktoren-Struktur, welche versucht, das Konstrukt „Benutzerzufriedenheit“ relativ breit zu erfassen. Der Scree-Test würde diese Annahme stützen. Tabelle 4.9 listet in der Komponentenmatrix die Ladungen der Items bei einer Ein-Faktoren-Struktur auf.

Tabelle 4.9

Komponentenmatrix nach erzwungener Ein-Faktoren-Lösung (Varimax, Item 1 bis 17 der ersten Testversion, N = 881, Missing values = EM)

Item	Component
1 Arbeitsrelevante Inhalte	.58
2 Aktuelle Informationen	.64
3 Gestaltung der Informationsarchitektur	.62
4 Verständlichkeit der Informationen	.71
5 Sicherheit im Umgang	.60
6 Erleichterung der Kommunikation	.70
7 Zugangsmöglichkeiten	.52
8 Qualität Suchmaschine	.42
9 Nachfrage und Feedbackmöglichkeiten	.50
10 Effizientere Arbeitsweisen	.75
11 Wissen Dokumente aufschalten	.34
12 Geschwindigkeit	.64
13 Einfache Bedienbarkeit	.69
14 Hilfe bei Problemen	.67
15 Aktuelle Firmenneuigkeiten	.67
16 Angemessene Informationspräsentation	.73
17 Vertrauen in Informationen	.59

Die Items laden zwischen .34 (Item 11) und .76 (Item 10) auf dem Faktor. Das Quadrat der Ladung eines Items beschreibt den gemeinsamen Varianzanteil zwischen Item und Faktor. Relativ niedrig und von unter .30 fällt dieser Anteil für die Items 7, 8, 9 und 11 aus. Im Anhang C4 ist eine Tabelle abgebildet, die die Kommunalitäten der einzelnen Items bei unterschiedlicher Faktorenstruktur aufführt. Eine tiefe Kommunalität schliesst allerdings nicht aus, dass ein Item reliabel sein kann. Wird Benutzerzufriedenheit durch einen einzigen Faktor erklärt und das Merkmal damit sehr breit gemessen, kann es sein, dass ein Item „nur mit

einem Teil seiner systematischen Varianz zur Aufklärung dieses Faktors beiträgt“ (zit. nach Bühner, 2004, S. 143).

Für die 17 Items konnte keine der durchgeführten Faktorenanalysen eine Einfachstruktur³⁴ hervorbringen. Die Frage nach der Anzahl Faktoren fällt wie gezeigt uneindeutig aus. Nach dem Scree-Test wird ein Faktor, nach der Parallelanalyse werden zwei Faktoren und nach dem Kaiser-Guttman-Kriterium drei Faktoren angezeigt. Da explorative Faktorenanalysen über keine aussagekräftigen Fit-Masse verfügen, lässt sich ihre Güte auf diese Weise nicht beurteilen. Theoretisch plausibel erscheinen sowohl eine Ein- wie auch eine Zwei-Faktoren-Struktur. Möglicherweise lässt sich die ‚Faktorenfrage‘ nach der Umfrage in der DEZA besser klären und wird deshalb vorerst stengelassen.

4.3.5 Reliabilität

Unter Reliabilität wird die Zuverlässigkeit eines Messinstrumentes verstanden. Lienert und Ratz (1998, S. 9) definieren die Reliabilität eines Tests als „Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht.“ Eine Form der Messung der Reliabilität ist die interne Konsistenz, ausgedrückt durch den Reliabilitätskoeffizienten (Cronbach Alpha).

Das Cronbach Alpha fällt mit .89 für die Skala relativ hoch aus. Die Tabelle 4.5 gibt in der Spalte ‚Cronbach Alpha if Item Deleted‘ Auskunft darüber, ob durch den Ausschluss eines Items eine höhere Reliabilität erzielt werden könnte. Dies wäre der Fall durch die Elimination des Items 8 oder 11, wobei der Ausschluss von Item 8 nur eine sehr geringe Verbesserung bedeuten würde. Ohne Item 11 stiege das Cronbach Alpha um beinahe .01.

4.3.4 Kriteriumsbezogene Validität

Kriteriumsbezogene Validität kann durch einen Vergleich des Testscores mit einem Kriterien-Score ermittelt werden. Im Normalfall wird dabei der Grad der Übereinstimmung mittels eines Korrelationskoeffizienten ausgedrückt (Fisseni, 1997, S. 98). Als Kriterium stand lediglich das gleichzeitig erhobene Global-Item 18 zur Verfügung.

Die Korrelation der Mittelwertskala (Item 1 bis 17) mit Global-Item 18 fiel signifikant ($p < .01$) und mit $r = .80$ relativ hoch aus. Dies spricht für das Messinstrument. Da das

³⁴ Unter dem Begriff Einfachstruktur versteht man, dass die Ladungen der Items auf einem Faktor sehr hoch und auf den anderen Faktoren sehr gering sind (Leonhart, 2004, S. 381).

Kriterium nur durch ein gleichzeitig erhobenes Item repräsentiert wird, sollte die Aussagekraft jedoch nicht überbewertet werden.

4.4 Diskussion

4.4.1 Skala

Die Probanden bewerteten die Items tendenziell im oberen Bereich der Sechser-Skala. Angesichts der Stichprobe, in der die Items getestet wurden, kommt dieser Befund nicht ganz überraschend. Bei einer Firma, die in der Versicherungsbranche zu den Grössten gehört, kann ein relativ gutes Intranet erwartet werden. Dieser Befund deckt sich zudem mit Umfrageergebnissen, welche die Firma Stimmt AG mit dem CSUQ (Lewis, 1995) gemacht hat. In ihren Umfragen bei 16 Grossunternehmen, die hauptsächlich aus dem Dienstleistungssektor stammten, fielen die Messwerte jeweils in die oberen Skalenbereiche. So lag der durchschnittliche Wert für *System Usability Overall* auf der 7-Skala bei 4.9 (Stimmt AG, 2004, S. 58).³⁵ Es lässt sich damit zwar nicht schlüssig belegen, scheint aber plausibel, dass die meisten grösseren Unternehmungen aus Benutzersicht tendenziell eher ein gutes, als ein schlechtes Intranet führen. Für den Test bedeutet dies, dass er vor allem im oberen Skalenbereich gut differenzieren wird. Mit der Häufung der (varianzberücksichtigten) Schwierigkeitsindizes im Bereich zwischen .50 bis .70 erweist sich der Test damit als relativ gut ‚justiert‘.

4.4.2 Items

Item- und Testkennwerte sind immer abhängig von der Stichprobe, an der sie gewonnen werden (Kranz, 1979, S. 55). Dies bedeutet, dass ein Item welches sich hier zur Trennung von zufriedenen und unzufriedenen Benutzern als schlecht geeignet erwies, in einer anderen Stichprobe womöglich besser abgeschnitten hätte. Da nach dieser Itemrevision eine weitere Testung der Items bei der DEZA anstand, wurde bei der Itemrevision eher ein moderater Massstab gewählt, ein Item im Zweifelsfall also eher beibehalten als ausgemustert. Grundsätzlich kann der Mehrheit der Items gute Kennwerte bescheinigt werden. Folgend werden jene Items besprochen, die sich für die Skala als relativ „schwach“ herausstellten.

³⁵ Leider werden im Report zu den 16 Umfragen nur Durchschnittswerte und keine Angaben zu Standardabweichungen gemacht, der Aussagewert bleibt damit für die Interpretation beschränkt.

Die Häufigkeitsverteilung der Werte von *Item 1* zeigt einen Deckeneffekt. Sein Homogenitätsindex und sein Trennschärfekoeffizient sind verhältnismässig tief. Ein Ausschluss könnte daher in Betracht gezogen werden. Es wird aber aus dem oben genannten Grund nochmals getestet.

Item 7 verfügt über eine hinreichende, aber relativ tiefe Trennschärfe. Zudem verfügt das Item über einen hohen Schwierigkeitsindex, welcher sich auch im beobachteten Deckeneffekt bei der Punkteverteilung widerspiegelt. Zu hohe oder zu tiefe Schwierigkeitsindizes widersprechen dem differentialpsychologischen Ansatz der klassischen Testtheorie, Probanden mit hoher Merkmalsausprägung von Probanden mit tiefer Merkmalsausprägung zu trennen. Aus dieser Perspektive könnte für einen Ausschluss des Items argumentiert werden. Inhaltliche Überlegungen sprechen allerdings für die Beibehaltung des Items im Fragebogen. Ohne zuverlässigen und stabilen Zugang ist ein Intranet als Arbeitsinstrument nicht denkbar. Schon auf kleinere Störungen dürften Mitarbeitende im Normalfall empfindlich reagieren, weil Arbeitsschritte plötzlich verzögert werden (bspw. kann ein Telefonat nicht gemacht werden, weil nicht auf das Mitarbeiter-Verzeichnis zugegriffen werden kann), oder einige Geschäfte unvorhergesehen nicht mehr erledigt werden können (bspw. kann ein wichtiges Dokument für die anstehende Sitzung nicht herunter geladen werden). Weiter ist es möglich, dass sich die Erwartungen betreff Zugang aufgrund technischer Fortschritte ständig erhöhen (bspw. möchte ein Aussendienstmitarbeiter über sein Mobiltelefon eine Information im Intranet abrufen können). *Item 7* wird aus diesen Gründen trotz eines relativ hohen Schwierigkeitsindizes und mittlerer Trennschärfe für den zweiten Testlauf im Fragebogen beibehalten.

Bei *Item 8* sind die Trennschärfe und der Homogenitätsindex verhältnismässig tief. Das Cronbach-Alpha liesse sich durch einen Ausschluss des Items leicht erhöhen. Positiv gewertet werden kann die vergleichsweise grosse Streuung des Items. Die grössere Itemvarianz führt dazu, dass das Item Personen besser differenziert (Kranz, 1979, S. 54). *Item 8* scheint einen Aspekt zu erfassen, der sich nur wenig mit den andern Items überschneidet. Der vergleichsweise tiefe Mittelwert weist auf eine grössere Anzahl von unzufriedenen Probanden hin. Aus den Intranet Reports der Stimmt AG (2003, 2004) ist bekannt, dass eine gute Suchmaschine neben einer guten Informationsarchitektur als eines der schwierigsten zu erreichenden Ziele ist. Für die nächste Umfrage wird das Item nochmals getestet. Allerdings wird das Item in eine kürzere Version umformuliert. Neu lautet es: „Wenn ich etwas mit der Suchmaschine suche, finde ich die gewünschte Information innert nützlicher Frist.“

Item 9 fällt vor allem durch seine hohe Anzahl von Probanden, die das Item mit „weiss nicht“ beantworteten, negativ auf. Das Item besitzt zudem eine mittlere Trennschärfe und einen verhältnismässig tiefen Selektionskennwert und Homogenitätsindex. Positiv ist seine mittlere Schwierigkeit. Es ist das einzige Item mit einem Schwierigkeitsindex zwischen .40 und .50. Interessant bei diesem Item ist, dass es zur kleinen Gruppe gehört, welches auf den zweiten Faktor lädt. Das Item wird trotz relativ schlechter Itemkennwerte im Fragebogen belassen, um zu untersuchen, wie Personen einer anderen Stichprobe auf das Item reagieren.

Die Itemkennwerte von *Item 11* sprechen klar für einen Ausschluss des Items. Es handelt sich um ein wenig homogenes, kaum trennscharfes und nicht reliables Item. Das Item wird deshalb aus dem Fragebogen eliminiert. Eine erneute Testung scheint wenig sinnvoll.

Item 17 besitzt ähnliche Itemkennwerte wie *Item 1* und zeigt in der Tendenz einen Deckeneffekt. *Believability* zählt nach Wang und Strong (1994) zum wichtigsten Datenattribut für Benutzer (siehe Kapitel 2.3). Es scheint daher sinnvoll, das Item nochmals zu testen.

Neben der „klassischen“ Itemanalyse wurden auch die Kommentarfelder ausgewertet, um Rückschlüsse über die Qualität der Items zu gewinnen. Den beiden Intranetreports 2003 und 2004 (Stimmt AG, 2003, 2004) ist zu entnehmen, dass bei Umfragen mit dem CSUQ (Lewis, 1995) das Sujet der Informationsarchitektur die meisten Reaktionen in Form von Rückmeldungen bzw. Kommentaren hervorrief. Diese Beobachtung trifft auch für diese Umfrage zu. In der deutschsprachigen Stichprobe verfassten 113 Probanden einen Kommentar zu Item 3. Dies entspricht einem Anteil von 13% (N = 908³⁶). Die Analyse zeigt, dass von den 113 abgegebenen Kommentaren sich knapp ein Drittel (N = 36) direkt auf das Thema Suchmaschine/-funktionen bezogen. Diese Vermischung der Ebenen Informationsarchitektur und Suchmaschine/-funktion in Bezug auf Item 3 war nicht beabsichtigt. Möglicherweise führte der Satzteil „... und gewünschte Inhalte zu finden“ bei vielen Probanden zu einer missverständlichen Annahme, dass Item 3 sich ebenfalls auf die Suchmaschine bezog. Um diese unerwünschte Vermischung ausschliessen zu können, wurde Item 3 deshalb trotz relativ guter Itemkennwerte umformuliert. Die Neuformulierung ist kürzer und präziser gewählt: „Das Intranet ist übersichtlich gestaltet und die Struktur nachvollziehbar.“ Gegenüber dem neu formulierten Item 3 erschien die Formulierung von Item 4 nicht mehr genug trennscharf. Während sich Item 3 auf die Verständlichkeit der Oberfläche bezieht, will Item 4 die Verständlichkeit der Inhalte erfassen. Damit diese beiden

³⁶ Für die Kommentarfeldanalyse wurden alle Probanden berücksichtigt, d.h. auch jene, die bei der Datenexploration rausgenommen wurden.

Ebenen genauer zum Ausdruck kommen, wurde Item 4 ebenfalls umformuliert bzw. präzisiert. Für die zweite Fragebogenversion lautet Item 4 deshalb: „Wenn ich auf dem Intranet etwas lese, sind die Informationen so verfasst, dass sie klar und verständlich sind.“

Da auch die DEZA den Fragebogen in Französisch anbieten wollte, wurden abschliessend noch die Item-Mittelwerte der verschiedenen Sprachgruppen verglichen.³⁷ Die Varianzanalyse (One-Way ANOVA) zeigt signifikante Mittelwertsunterschiede für die Items 3, 4, 5, 6 und 14 an (siehe Anhang C5). Der Scheffé-Test weist dabei jeweils signifikante Mittelwertsunterschiede zwischen der deutsch- und französischsprachigen Stichprobe aus. Die Gründe für diese Unterschiede können unterschiedliche Ursachen haben. Aufgrund der beachtlichen Differenz bei der Grösse der Stichproben sind Stichprobeneffekte nicht unwahrscheinlich. Ein anderer Grund für die Unterschiede könnte in einer „fehlerhaften“ Übersetzung der ursprünglich in Deutsch verfassten Items liegen. Items könnten aufgrund der Sprache von den Probanden unterschiedlich interpretiert worden sein und sich dadurch auf die Art der Beantwortung niederschlagen haben. Für die weitere Testung wurden deshalb bei allen Items mit signifikanten Mittelwertsunterschieden die Übersetzungen kontrolliert. Die zwei neu hinzugezogenen Personen, welche nicht in die Übersetzung des ersten Fragebogens involviert waren, attestierten den Items dabei eine treffende Übersetzung. Eine dritte mögliche Erklärung könnte in einer tatsächlich vorhandenen unterschiedlichen Zufriedenheit liegen. Oft werden aus Kosten- und Effizienzgründen nicht alle Inhalte in alle Sprachen übersetzt. Item 4 erfasst so bspw. die Zufriedenheit im Zusammenhang mit der Art und Weise, wie Informationen auf dem Intranet vermittelt werden. Der französischen Sprachminderheit liegen eventuell weniger Inhalte in ihrer Muttersprache vor, was sich negativ auf ihre Zufriedenheit auswirken könnte. Insbesondere auch bei den Items 3 und 14 kann der Sprache eine kritische Rolle zukommen.

³⁷ Es sei darauf hingewiesen, dass es nicht Ziel dieser Arbeit war, den Fragebogen in verschiedenen Sprachen zu validieren. Die Methoden wurden dementsprechend auch nicht darauf ausgerichtet.

5 Testvalidierung II

Drei Monate vor Start der vorliegenden Umfrage führte die DEZA bereits bei einem Teil seiner MitarbeiterInnen eine Befragung zum Thema Intranet durch. Die Intranetverantwortliche der DEZA wollte deshalb für diese Umfrage verständlicherweise nicht nochmals dieselben Personen anfragen. Auch aus methodischer Sicht schien es besser, diese Personen bereits im Vorfeld auszuschliessen, um mögliche Verzerrungen zu verhindern. Insgesamt konnte so eine Stichprobengrösse von nur 127 Probanden gewonnen werden, was weit unter der von Lienert und Ratz (1998) empfohlenen Stichprobengrösse von mindestens 200 Probanden liegt. Als zweites Handicap erwies sich, dass das „Suchmaschinen-Item“ nicht validiert werden konnte. Zum Zeitpunkt der Umfrage plante die DEZA erst die Einführung einer solchen. Beim Relaunch des Intranets ein Jahr vor der Umfrage wurde es vorgezogen, mit der Einführung einer Suchmaschine zuzuwarten, da man der Ansicht war, es sei besser keine Suchmaschine zu haben, als eine womöglich schlecht umgesetzte. Es machte daher keinen Sinn, das „Suchmaschinen-Item“ im Fragebogen einzusetzen. Die DEZA bot somit nur suboptimale Bedingungen für die zweite Validierung. Da im Rahmen der vorliegenden Arbeit in der verbleibenden Zeit keine neue Stichprobe bzw. Firma gefunden werden konnte, wurde entschieden, die Umfrage trotz benannten Schwächen durchzuführen.

Die Auswertung der Daten verlief nach demselben Muster wie bei Unternehmung A. Die Ergebnisse werden deshalb in strafferer Form präsentiert und sich wiederholende methodische Überlegungen ausgelassen.

5.1 Stichprobenbeschreibung

Die Umfrage in der DEZA beschränkte sich auf die Mitarbeitenden der Zentrale in Bern (N = 472). Die Gruppe, welche nicht bereits in die vorherige Intranet-Umfrage involviert war, betrug noch 370 Personen.³⁸ Von diesen 370 Personen reagierten 173 auf die E-Mail und beantworteten den Fragebogen. Damit konnte eine Rücklaufquote von nicht ganz 47% erreicht werden. 134 (= 77,5%) Personen füllten die deutschsprachige und 39 Personen (= 22,5%) die französischsprachige Version der Umfrage aus. Aufgrund der Datenexploration wurden 7 Personen (= 5,2%) aus der deutschsprachigen Stichprobe wegen einer falschen Skalenbenutzung ausgeschlossen. Die Stichprobe für die Itemanalyse umfasste damit 127

³⁸ Der Zufallsgenerator eines Computerprogramms erlaubt es, aus dem Pool aller Mitarbeitenden der DEZA beliebig grosse Stichproben zu bilden. Bei den 370 Personen handelte es sich also um jene Personen, die zuvor nicht vom Zufallsgenerator ausgewählt wurden.

Probanden. Die französischsprachige Stichprobe wurde für die weiteren Auswertungen nicht berücksichtigt

Neben den Items des Fragebogens wurden von den Probanden zusätzlich Geschlecht, Alter, durchschnittliche Benutzung des Intranets sowie Länge ihres Arbeitsverhältnisses bei der DEZA erhoben. Die deutschsprachige Stichprobe setzt sich aus 72 Frauen und 54 Männern zusammen. Die erhobenen Altersklassen präsentieren sich ausgeglichen. 24 Personen gaben an, 29 Jahre und jünger zu sein, 33 zwischen 30 und 39 Jahren, 40 zwischen 40 und 49 Jahren und 29 Personen gaben an, 50 Jahre und älter zu sein. Die meisten der Probanden arbeiten seit 3 bis 5 Jahren bei der DEZA ($N = 45$). 17 Personen sind seit weniger als einem Jahr angestellt und 10 Personen stehen seit 1 bis 2 Jahren in einem Angestelltenverhältnis. 25 Personen leisten seit 6 bis 10 Jahren und 29 Probanden seit mehr als 11 Jahren ihren Dienst bei der DEZA. Fast zwei Drittel der Probanden benutzen das Intranet mehrmals täglich ($N = 79$; 62,2%). 41 der Probanden oder 32,3% nutzen das Intranet einmal pro Tag. Eine kleine Anzahl der Probanden gibt an, das Intranet wöchentlich zu nutzen ($N = 6$; 4,7%). Damit nutzen fast 95% der Probanden mindestens einmal pro Tag das Intranet. Ein Proband machte keine Angaben zu seiner Person. Im Anhang D1 finden sich diese Angaben in tabellarischer Form.

5.2 Itemanalyse

Die fehlenden Werte wurden erneut mit dem EM-Algorithmus ersetzt und die Itemanalyse auf Basis der EM-Datenmatrix durchgeführt. Die Untersuchung der Ausreißer führte wie bereits beschrieben zum Ausschluss von 7 Probanden.

5.2.1 Statistische Kennwerte

Der Mittelwert der Skala (Item 1 bis Item 16) liegt bei 4.60 und einer Standardabweichung von .56. Der Median ist mit 4.68 leicht höher als der Mittelwert. Das Minimum beträgt 3.06 und das Maximum 5.81 Punkte, was einer Spanne von 2.75 Punkte entspricht. Die Werteverteilung ist in der Tendenz rechtsgipflig (Schiefenkennwert = -.45, $s_S = .22$) und die Kurtosis nahe Null (Exzessivitätskennwert = .04, $s_E = .43$). Im Normalverteilungsdiagramm (Q-Q-Diagramm) reihen sich die Werte relativ perlenartig der Geraden entlang, fransen unten jedoch etwas aus. Das trendbereinigte Normalverteilungsdiagramm zeigt im unteren und oberen Skalenbereich ein etwas unruhiges Abweichungsmuster. Der von Lillefors leicht modifizierte Kolmogorov-Smirnov Test und der Shapiro-Wilk-Test zur Überprüfung auf

Normalverteilung fallen nicht signifikant aus ($p > 0.05$). Es kann somit von einer Normalverteilung ausgegangen werden. Abbildung 5.1 gibt Aufschluss über die Wertverteilung.

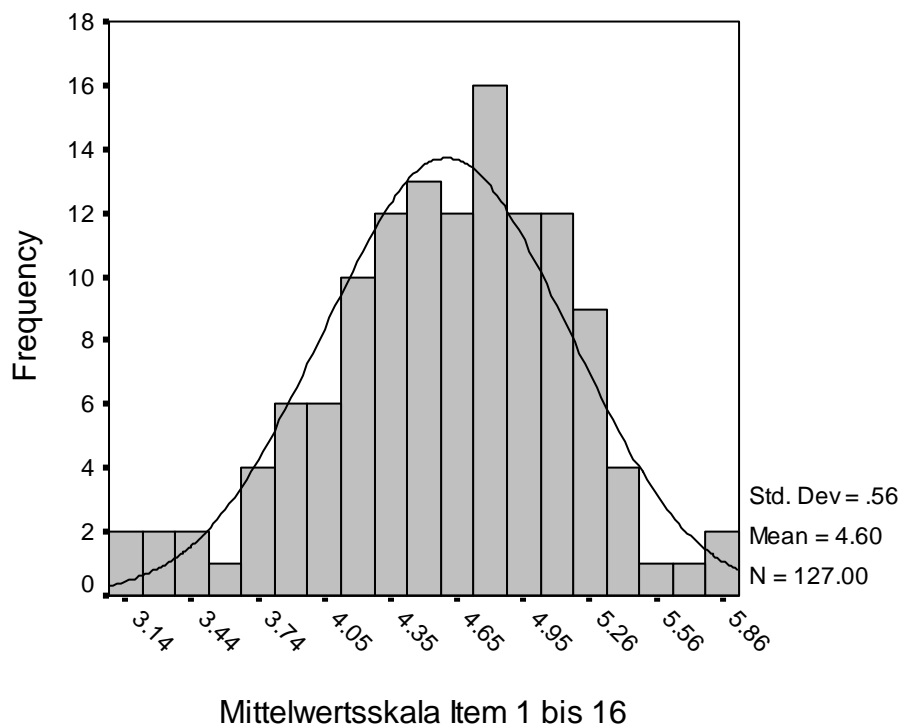


Abbildung 5.1. Häufigkeitsverteilung der Mittelwertsskala Item 1 bis 16 der zweiten Testversion.

Die Mittelwerte der Items bewegen sich zwischen 3.19 (Item 3) und 5.26 (Item 1). Ihre Standardabweichungen variieren von .75 (Item 2) bis 1.25 (Item 3). Neun Items (1, 2, 4, 7, 8, 12, 13, 14 und 15) streuen weniger als Eins. Vier Items (1, 7, 13 und 15) besitzen einen Mittelwert von über 5.0 und nur zwei Items (3 und 10) haben im Mittel einen Wert von weniger als 4.0. Median- und Modalwerte sind generell hoch. Nur ein einziges Mal ist der Median kleiner als 4 (Item 3).

Die Histogramme der Items sind im Anhang D2 nachschlagbar. Die Werteverteilungen weisen tendenziell rechtsasymmetrische und schmalgipflige Formen auf. Bei mehreren Items wurde die Skala im unteren Bereich von den Probanden nicht oder selten genutzt. So wählte keiner der 127 Probanden die Kategorie „stimme überhaupt nicht zu“ für die Items 1, 2, 4, 6, 7, 8, 12 und 15. Die Belegung der ersten drei Kategorien liegt mehrmals unter 10% der gesamten Stichprobe (Item 1, 2, 4, 7, 8, 12, 13, 15). Tabelle 5.1 bietet einen Überblick einiger Item-Kennwerte.

Tabelle 5.1

Item-Übersicht: Anzahl Probanden (N), Mittelwerte ($Mean$) und Standardabweichungen (SD), Median- und Modalwert sowie die Schiefenkennwerte (S), Exzessivitätskennwerte (E) und Schwierigkeitsindizes mit Berücksichtigung der Varianz (p_{mv}) und ohne Berücksichtigung der Varianz (p_m)

	N	$Mean$	SD	$Median$	$Mode$	S	E	p_{mv}	p_m
1 Arbeitsrelevante Informationen	127	5.26	0.83	5	6	-1.03	1.01	.787	.877
2 Aktuelle Informationen	127	4.81	0.75	5	5	-0.23	-0.20	.658	.801
3 Gestaltung der Informationsarchitektur	127	3.19	1.25	3	4	-0.02	-0.97	.326	.531
4 Verständlichkeit der Informationen	127	4.85	0.76	5	5	-0.74	1.32	.669	.808
5 Sicherheit im Umgang	127	4.33	1.01	5	5	-0.89	0.48	.550	.722
6 Erleichterung der Kommunikation	127	4.87	1.00	5	5	-0.93	0.62	.685	.811
7 Zugangsmöglichkeiten	127	5.13	0.83	5	5	-0.91	1.02	.750	.855
8 Nachfrage und Feedbackmöglichkeiten	127	4.77	0.85	5	5	-0.87	1.50	.651	.794
9 Effiziente Arbeitsweise möglich	127	4.33	1.05	4	4	-0.44	0.26	.551	.722
10 Geschwindigkeit	127	3.78	1.21	4	4	-0.27	-0.69	.437	.630
11 Einfache Bedienbarkeit	127	4.21	1.17	4	5	-0.62	-0.15	.529	.701
12 Hilfe bei Problemen	127	4.90	0.92	5	5	-1.21	1.82	.691	.817
13 Aktuelle Firmenneuigkeiten	127	5.16	0.76	5	5	-1.50	6.14	.756	.860
14 Angemessene Informationspräsentation	127	4.42	0.99	5	5	-0.83	0.95	.570	.737
15 Vertrauen in Informationen	127	5.25	0.77	5	5	-1.22	2.48	.782	.875
16 Globalzufriedenheit	127	4.39	1.08	5	5	-0.93	0.88	.567	.731

Anmerkung. Missing values = EM; Standardfehler der Schiefe $s_s = .215$; Standardfehler des Exzesses $s_e = .427$

5.2.2 Schwierigkeitsindizes

Aufgrund der beobachteten Verteilungen konnten hohe Schwierigkeitsindizes erwartet werden. Abbildung 5.2 stellt die Häufigkeitsverteilungen der Schwierigkeitsindizes mit und ohne Berücksichtigung der Varianz gegenüber.

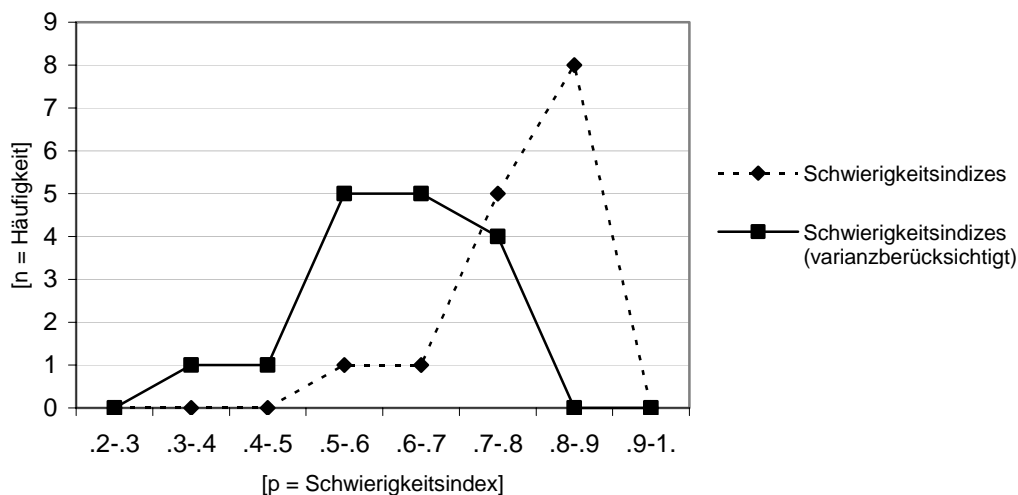


Abbildung 5.2. Häufigkeitsverteilung der Schwierigkeitsindizes der zweiten Testversion mit und ohne Berücksichtigung der Varianz.

Die Schwierigkeitsindizes streuen mit Berücksichtigung der Varianz hauptsächlich zwischen .50 und .80. Ohne deren Einbezug verteilen sie sich überwiegend im Bereich von .70 bis .90. Die Empfehlung von Fisseni (1997, S. 44f.), sich bei der Itemanalyse auf varianzberücksichtigte Schwierigkeitsindizes zu beziehen, stellt sich als heikel heraus. Betrachtet man die Häufigkeitsverteilung der Schwierigkeitsindizes in Abbildung 5.2, so würde man vermutlich aufgrund der Kurve der varianzberücksichtigten Schwierigkeitsindizes ‚bessere‘ Häufigkeitsverteilungen bei den Items erwarten als die beobachteten (siehe Anhang D2). Um dieser Verzerrung entgegenzuwirken, scheint es daher angezeigt, beide Berechnungsweisen im Auge zu behalten. Tabelle 5.1 listet die Schwierigkeitsindizes deshalb mit und ohne Berücksichtigung der Varianz auf.

5.2.3 Trennschärfe

Das Globalitem wurde wiederum aus den Trennschärfenberechnungen ausgeschlossen. Die Tabelle 5.2 listet die Ergebnisse auf.

Tabelle 5.2

Trennschärfen und Cronbach's Alpha der Items 1 bis 15 der zweiten Testversion (N = 127, Missing values = EM)

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
1 Arbeitsrelevante Informationen	63.9948	59.6367	.3320	.8410
2 Aktuelle Informationen	64.4464	61.4347	.2175	.8458
3 Gestaltung der Informationsarchitektur	66.0657	52.6037	.5726	.8275
4 Verständlichkeit der Informationen	64.4059	59.7238	.3653	.8393
5 Sicherheit im Umgang	64.9212	56.2002	.4836	.8331
6 Erleichterung der Kommunikation	64.3885	56.0493	.5004	.8320
7 Zugangsmöglichkeiten	64.1259	58.7759	.4015	.8376
8 Nachfrage und Feedbackmöglichkeiten	64.4889	58.7956	.3858	.8384
9 Effiziente Arbeitsweise möglich	64.9247	53.9966	.6165	.8245
10 Geschwindigkeit	65.4776	53.3431	.5534	.8288
11 Einfache Bedienbarkeit	65.0475	52.6383	.6231	.8235
12 Hilfe bei Problemen	64.3505	55.8708	.5722	.8281
13 Aktuelle Firmenneuigkeiten	64.0927	60.1988	.3215	.8413
14 Angemessene Informationspräsentation	64.8322	54.662	.6082	.8254
15 Vertrauen in Informationen	64.0027	59.2174	.4041	.8375

$$\alpha_{\text{item1-15}} = .8433$$

Die Trennschärfekoeffizienten variieren zwischen .22 (Item 2) und .62 (Item 11). Im Mittel betragen die Trennschärfen .46 ($SD = .13$). Mit den Items 3, 9, 10, 11, 12 und 14 gelingt es am besten, zwischen Probanden mit hoher und tiefer Ausprägung zu unterscheiden. Item 2

erweist sich in dieser Stichprobe als ungeeignet (Borg & Staufenbiel, 1997). Verhältnismässig tiefe Werte von .30 bis .50 besitzen die Items 1, 4, 7, 8, 13 und 15. Die tiefsten Selektionskennwerte weisen die Items 2, 4 und 13 auf.

5.2.4 Homogenität

Die Homogenität der Skala wurde wiederum anhand der Interkorrelation und der Faktorenanalyse untersucht.

5.2.4.1 Interkorrelation

Wie die Tabelle 5.3 aufzeigt, sind im Gegensatz zur Interkorrelationsmatrix des Unternehmens A nicht mehr alle Korrelationen signifikant. Dies dürfte auch mit der deutlich geringeren Stichprobengrösse zusammenhängen. Item 1 und 2 weisen mehrere nicht signifikante Korrelationen auf. Darunter auch einen nicht bedeutsamen Zusammenhang zwischen Item 2 und Global-Item 16.

Tabelle 5.3

Interkorrelationsmatrize und Homogenitätsindizes der Items 1 bis 16 der zweiten Testversion (N = 127, Missing values = EM)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1															
2	.158	1														
3	.266**	-.034	1													
4	.091	.272**	.119	1												
5	.292**	-.086	.524**	.145	1											
6	.080	.152	.286**	.214*	.349**	1										
7	.142	.095	.265**	.102	.213*	.234**	1									
8	.145	.254**	.206*	.120	.125	.220*	.342**	1								
9	.356**	-.031	.563**	.212*	.547**	.394**	.205*	.189*	1							
10	.337**	.250**	.336**	.230**	.260**	.263**	.372**	.243**	.367**	1						
11	.119	.076	.539	.243**	.335**	.451**	.334**	.299**	.394**	.451**	1					
12	.177*	.124	.491**	.243**	.312**	.365**	.205*	.275**	.411**	.290**	.521**	1				
13	.123	.181*	.124	.280**	-.019	.134	.215*	.237**	.205*	.288**	.157	.224*	1			
14	.246**	.329**	.381**	.326**	.285**	.332**	.229**	.247**	.451**	.417**	.470**	.356**	.248**	1		
15	.009	.148	.140	.395**	.281**	.375**	.135	.195*	.301**	.165	.210*	.313**	.256**	.339**	1	
16	.329**	.098	.683**	.227*	.520**	.464**	.375**	.309**	.629**	.521**	.605**	.601**	.198*	.538**	.255**	1
H	.191	.132	.326	.215	.272	.287	.231	.227	.346	.319	.347	.327	.190	.346	.234	.423

Anmerkung. *p < .05, **p < .01, H = Homogenitätskoeffizient.

Dreimal kommt es zu einem negativen Vorzeichen, der Zusammenhang ist in diesen Fällen aber jeweils praktisch Null. Im Schnitt fallen die Korrelationen erneut relativ tief aus, was sich im vergleichsweise tiefen Homogenitätskoeffizienten der Skala von .28 ausdrückt.

Nur Global-Item 16 hat einen Homogenitätskoeffizienten von mehr als .40 und ist erneut das homogenste Item. Die höchsten Zusammenhänge mit Item 16 finden sich für die Items 3, 9, 11 und 12.

5.2.4.2 Faktorenanalyse

Die Stichprobengröße von $N = 127$ ist nicht optimal für eine Faktorenanalyse jedoch ausreichend (Bühner, 2004, S. 151). Der Kaiser-Meyer-Olkin-Koeffizient beträgt .83 und attestiert den Variablen damit eine gute Eignung für eine Faktorenanalyse. Ebenso fällt Bartlett's Test auf Sphärizität signifikant ($p < 0.01$) aus. Die MSA-Koeffizienten in der Anti-Image-Matrix sind hoch und meist über .80. Für Item 2 fällt er mit .60 relativ tief aus. Nach Bühner (2004) ist aber erst ab einem $MSA < .50$ ³⁹ eine Eignung nicht mehr gegeben. Abgesehen davon, dass die berücksichtigten Variablen erneut nicht der Normalverteilung folgen, sind die Bedingungen für eine Faktorenanalyse also gegeben.

Nach dem Kaiser-Guttman-Kriterium extrahierte die Faktorenanalyse (Varimax, Hauptkomponentenanalyse) vier Faktoren mit einem Eigenwert grösser Eins. Die vier Faktoren erklären zusammen rund 58% der Varianz. Der erste Faktor erklärt ca. 32%, der zweite 11%, der dritte 8% und der vierte Faktor noch weitere 7% der Varianz. Eine Tabelle mit der erklärten Gesamtvarianz findet sich im Anhang D3. Die Ladungen der Items in der rotierten Komponentenmatrix der Vier-Faktoren-Struktur erfüllen jedoch weder die statistischen Bedingungen für eine Interpretation (Guadagnoli & Velicer, 1988, zit. nach Bortz, 1999, S. 507), noch ist eine inhaltliche Interpretation sinnvoll.⁴⁰ Der Screeplot in Abbildung 5.3 veranschaulicht erneut einen varianzstarken ersten Faktor.

³⁹ Bühner (2004, S. 160) spricht in seinem Buch fälschlicherweise von KMO-Koeffizienten statt von MSA-Koeffizienten. So schreibt er, eine Eignung sei bei „KMO $< .50$ “ nicht gegeben, dürfte dabei aber KMO mit MSA verwechselt haben.

⁴⁰ Das Gleiche gilt auch für weitere getestete Faktoren-Strukturen, wie einer Hauptkomponentenanalyse mit Varimax-Rotation und erzwungener Drei-Faktoren-Lösung oder für getestete oblique Rotationen (Oblimin direct, Delta = 0).

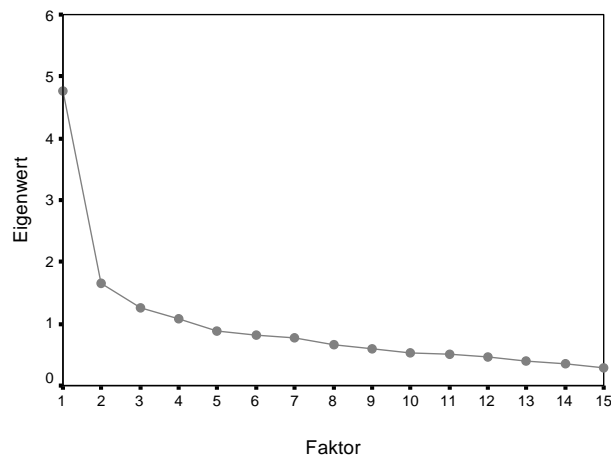


Abbildung 5.3. Screeplot der Faktorenanalyse der Items 1 bis 15 der zweiten Testversion (Hauptkomponentenanalyse, Varimax, $N = 127$, Missing values = EM).

Die Parallelanalyse nach Horn ermittelt zwei Faktoren, deren Eigenwerte über zufällig erzeugten Eigenwerten liegen. Der Output der Parallelanalyse auf Basis der SPSS-Syntax von O’Conner (2000) ist im Anhang D3 abgedruckt. Aufgrund dieses Ergebnisses wurde eine Hauptkomponentenanalyse mit Varimax-Rotation und erzwungener Zwei-Faktoren-Struktur durchgeführt. Das Ergebnis der rotierten Komponentenmatrix ist in Tabelle 5.4 dargestellt. Wie bereits in der Komponentenmatrix aus der Umfrage beim Unternehmen A, lassen sich mehrere Items nicht klar einem Faktor zuordnen.

Tabelle 5.4

Rotierte Komponentenmatrix der Hauptkomponentenanalyse mit erzwungener Zwei-Faktoren-Lösung. ($N = 127$, Varimax, Missings values = EM)

Item	Component	
	1	2
1 Arbeitsrelevante Informationen	.40	.13
2 Aktuelle Informationen	-.16	.70
3 Gestaltung der Informationsarchitektur	.81	.01
4 Verständlichkeit der Informationen	.13	.60
5 Sicherheit im Umgang	.77	-.09
6 Erleichterung der Kommunikation	.51	.32
7 Zugangsmöglichkeiten	.35	.33
8 Nachfrage und Feedbackmöglichkeiten	.22	.49
9 Effizientere Arbeitsweisen	.78	.12
10 Geschwindigkeit	.45	.46
11 Einfache Bedienbarkeit	.66	.30
12 Hilfe bei Problemen	.59	.31
13 Aktuelle Firmenneuigkeiten	.06	.60
14 Angemessene Informationspräsentation	.47	.53
15 Vertrauen in Informationen	.27	.49

Die Umformulierung von Item 4 nach der ersten Itemanalyse mit dem Ziel einer klaren Abgrenzung zu Item 3 scheint, zumindest was ihr Ladeverhalten auf den zwei Faktoren betrifft, gelungen. Die Items 3, 5, 9 und 11 lassen sich relativ eindeutig dem ersten, die Items 2, 4 und 13 verhältnismässig klar dem zweiten Faktor zuordnen. Item 1 und 7 weisen auf beiden Faktoren relativ tiefe Ladungen auf. Die restlichen Items zeigen zwei mittlere oder eine mittlere und eine niedrige Ladung. Ein Vergleich mit der Komponentenmatrix des Unternehmens A ist nur sehr bedingt zulässig, da es sich nicht mehr um die exakt gleichen Items handelt. Es ist aber interessant, dass Item 1 und 15 nicht mehr klar auf einem Faktor laden wie zuvor. Wie beim Unternehmen A, kann die Zwei-Faktoren-Struktur erneut nicht überzeugen. Es wurde deshalb wiederum eine Ein-Faktoren-Struktur getestet. Die Komponentenmatrix ist in Tabelle 5.5 abgebildet.

Tabelle 5.5

Komponentenmatrix nach erzwungener Ein-Faktoren-Lösung (N = 127, Varimax, Missing values = EM)

Item	Component
1 Arbeitsrelevante Informationen	.40
2 Aktuelle Informationen	.27
3 Gestaltung der Informationsarchitektur	.68
4 Verständlichkeit der Informationen	.44
5 Sicherheit im Umgang	.59
6 Erleichterung der Kommunikation	.60
7 Zugangsmöglichkeiten	.48
8 Nachfrage und Feedbackmöglichkeiten	.46
9 Effizientere Arbeitsweisen	.71
10 Geschwindigkeit	.63
11 Einfache Bedienbarkeit	.72
12 Hilfe bei Problemen	.67
13 Aktuelle Firmenneuigkeiten	.39
14 Angemessene Informationspräsentation	.69
15 Vertrauen in Informationen	.50

Vergleichsweise hohe Ladungen von $> .60$ können für die Items 3, 9, 10, 11, 12 und 14 beobachtet werden. Der gemeinsame Varianzanteil dieser Items mit dem Faktor beträgt damit mindestens 36%. Knapp darunter liegen Item 5 und 6. Eine Übersicht der Kommunalitäten der Items bei unterschiedlicher Faktorenlösung ist im Anhang D3 abgedruckt. Die Faktorendiskussion wird hier nach der Darstellung der Resultate kurz unterbrochen, um die Itemanalyse abzuschliessen. Danach wird sie in der Diskussion nochmals aufgegriffen.

5.2.5 Reliabilität

Zur Kontrolle der Zuverlässigkeit der Skala wurde wiederum die interne Konsistenz bestimmt. Der Reliabilitätskoeffizient Cronbach Alpha fällt mit .84 relativ gut aus. Durch den Ausschluss von Item 2 liesse sich die interne Konsistenz leicht verbessern. Eine Übersicht der Cronbach's Alpha ist in Tabelle 5.2 aufgelistet.

5.2.6 Kriteriumsbezogene Validität

Da kein anderes Kriterium zur Verfügung stand, wurde wiederum das Gopal-Item als Kriterium benutzt und mit der Mittelwertskala (Item 1 bis 15) korreliert. Der Zusammenhang fällt signifikant ($p < 0.01$) und mit einer Korrelation von .79 relativ hoch aus ($N = 127$, Missing values = EM). Die gemeinsame Varianz beträgt damit rund 63%.

5.3 Diskussion

Bevor die einzelnen Items besprochen und revidiert werden, wird zuerst die Frage der Faktorenstruktur wieder aufgegriffen. Abgeschlossen wird das Kapitel mit der vorläufigen Endversion des Fragebogens.

5.3.1 Faktorenstruktur

Ein-Faktor-Struktur oder Zwei-Faktoren-Struktur?

Aufgrund der Resultate im Unternehmen A und in der DEZA sind für die Bildung einer angestrebten homogenen Skala theoretisch zwei Wege denkbar, die es folgend gegeneinander abzuwägen gilt. Die Frage nach einer Ein-Faktor- oder Zwei-Faktoren-Struktur ist deshalb von grosser Bedeutung, da sie Auswirkungen auf die Itemselektion hat.

Die Parallelanalyse nach Horn extrahierte beide Male zwei Faktoren. Eine replizierte Faktorenstruktur ist an sich jedoch noch kein schlagendes Argument für deren Richtigkeit, da auch konkurrierende Faktorlösungen in gleicher Weise replizierbar sein können (Bühner, 2004, S. 164). Wie bereits beschrieben, liesse sich aus theoretischer Sicht für eine Zwei-Faktoren-Struktur mit den Faktoren „subjektiv wahrgenommene Qualität der Informationen“ und „subjektiv wahrgenommene Nutzbarkeit/Tauglichkeit des Intranets“ argumentieren. Für die Itemselektion würde dies bedeuten, dass versucht werden sollte, zwei homogene Subtests zu bilden. Es müsste deshalb überlegt werden, Items mit Ladungen auf beiden Faktoren auszuschliessen. Dies birgt in Bezug auf die Itemselektion grosse Schwächen, da Items

ausgeschlossen werden müssten, die aufgrund der Itemanalyse zu den ‚Besten‘ gehören wie bspw. das Item „Angemessene Informationspräsentation“, das Item „Erleichterung der Kommunikation“ oder auch das Item „Geschwindigkeit“. Zugleich wird hier eine inhaltliche Schwäche dieser Zwei-Faktoren-Annahme deutlich. Item 12 „Geschwindigkeit“ kann so im Sinne des ersten Faktors verstanden werden als Qualität einer Information, schnell zur Verfügung zu stehen. Im Sinne des zweiten Faktors als Tauglichkeitsbeweis, weil ein schnelles Navigieren zu einer Vielzahl von Informationen möglich ist. Das Item „Angemessene Informationspräsentation“ kann ebenfalls als Attribut für Qualität einer Information, als auch im Sinne einer guten Tauglichkeit für den User verstanden werden. In ähnlicher Weise könnte auch für das „Hilfe bei Problemen“-Item argumentiert werden. Theoretisch erweist es sich daher als problematisch, von zwei unabhängigen Faktoren auszugehen. Vielmehr handelt es sich um eine Frage des Standpunktes. Das Problem lässt sich auch dadurch veranschaulichen, wenn man bedenkt, dass das Intranet als Medium für Informationen benutzt wird und die zwei Ebenen von der Sache her zwangsläufig eng verknüpft sind. Wenn Benutzerzufriedenheit erfasst werden will, erscheint eine Trennung damit nicht sinnvoll.

Für die Skala lässt sich damit schlussfolgern, dass eine Ein-Faktoren-Struktur mit dem Faktor „Benutzerzufriedenheit“ am plausibelsten erscheint. Unterstützt wird diese Annahme durch den jeweils varianzstarken ersten Faktor, der sich besonders beim Unternehmen A markant veranschaulichen lässt. Betrachtet man zudem die Kommunalitäten der Items (siehe Anhang) bei unterschiedlichen Faktorenlösungen fällt auf, dass bei vielen Items die Kommunalität von der Zwei- zur Ein-Faktoren-Struktur nicht markant abnimmt.

Für die Itemselektion beutet dies, dass weniger homogene Items eliminiert werden sollten. Nachdem in der Itemrevision des Unternehmens A ein relativ moderater Massstab angesetzt wurde, werden die Kriterien für einen Verbleib in der Skala nun strenger angesetzt. Folgend werden die Items kurz durchgegangen und ihre Tauglichkeit diskutiert. Die Itemrevision bezieht sich dabei auf beide Untersuchungen. Dies geschah im Bewusstsein, dass es methodisch überaus heikel ist, Itemkennwerte von Items zu vergleichen, die nicht mit der exakt gleichen Zusammensetzung von Items erhoben und analysiert wurden. Für dieses Vorgehen sprachen jedoch der deutliche Stichprobenunterschied von fast 7:1 und der Umstand, dass die Stichprobe der DEZA nicht die empfohlene Grösse von mindestens 200 Probanden für eine Itemanalyse erreichte (Lienert & Raatz, 1998). Es sei dabei vorausgeschickt, dass diese Überarbeitung der Erstellung einer Zwischenversion dient, welche

nochmals getestet werden muss. Dies konnte jedoch nicht mehr im Rahmen der vorliegenden Arbeit geschehen.

5.3.2 Itemrevision

Folgende Items haben sich in beiden Itemanalysen als gut herausgestellt und werden deshalb im Fragebogen belassen, ohne weiter besprochen zu werden: *Item 6* „Erleichterung der Kommunikation“, *Item 10* „Geschwindigkeit“, *Item 11* „Einfache Bedienbarkeit“ und *Item 12* „Hilfe bei Problemen“.

Item 1 „Arbeitsrelevante Inhalte“ erweist sich in beiden Untersuchungen als mässig homogen und verfügt über vergleichsweise tiefe Trennschärfen. Die Mittelwerte fallen mit 4.82 beim Unternehmen A und 5.26 bei der DEZA relativ hoch aus. Bei der DEZA kommt es zu einem Deckeneffekt. Sowohl den Verantwortlichen des Unternehmens A, wie den Verantwortlichen der DEZA scheint es gelungen, diesbezüglich die Erwartungen der Mitarbeitenden zu erfüllen. Bei guten Intranets scheint das Item zu Deckeneffekten zu neigen, es ist jedoch anzunehmen, dass das Item bei weniger guten Intranets an Bedeutung gewinnt. Für Entwickler deckt dieses Item zudem um ein zentrales Ziel ab. Das Item sollte deshalb nochmals getestet werden und wird im Fragebogen belassen.

Item 2 „Aktuelle Informationen“ zeigt zwischen den Itemkennwerten der beiden Umfragen eine enorme Diskrepanz. So korreliert bspw. das Item mit dem Global-Item im Unternehmen A signifikant ($p < 0.01$) und mit fast .50, während sich in der DEZA kein Zusammenhang findet. Hier könnte es zu einer Verzerrung gekommen sein. Das Intranet der DEZA wurde erst vor knapp einem Jahr komplett überarbeitet. Viele ältere Inhalte wurden dabei nicht mehr migriert und bei weiteren Inhalten wurde der Zeitpunkt genutzt, sie wieder auf den neuesten Stand zu bringen. Die Inhalte im Intranet der DEZA waren zum Zeitpunkt der Umfrage daher eventuell „überdurchschnittlich“ aktuell. Wenn die Inhalte jedoch nicht sorgfältig gepflegt werden, könnte sich dies bereits in relativ kurzer Zeit ändern und einige der Mitarbeitenden verärgern. Da die Itemkennwerte bei der Unternehmung A, bei deutlich grösserer Stichprobe, nicht schlecht waren und es hier wohlmöglich aufgrund des verhältnismässig jungen Alters des Intranets der DEZA zu Verzerrungen kam, wird das Item im Fragebogen belassen. Zudem fragt es sich, ob es nicht besser wäre, den Begriff „Inhalte“ dem verwendeten Begriff „Informationen“ vorzuziehen. Das Wortpaar „Informationen-

aktuell“ könnte falsche Assoziationen erzeugen.⁴¹ Ein Proband könnte spontan eher an Mitteilungen bzw. News denken, als generell an Inhalte (*Content*) im Intranet. Da es sich hier nur um einen äusserst kleinen Eingriff handelt, wird das Item deshalb umformuliert zu: „Die Inhalte im Intranet sind aktuell.“

Item 3 „Gestaltung der Informationsarchitektur“ hat sich nach der Umformulierung als ein brauchbares Item erwiesen.

Item 4 „Verständlichkeit der Informationen“ verfügt nach der Umformulierung über vergleichsweise schlechte Itemkennwerte, was für einen Ausschluss sprechen würde. Es ist allerdings heikel, dass Item einzig aufgrund der Werte der zweiten Umfrage auszuschliessen. Das Item sollte nochmals in einer grösseren Stichprobe validiert werden, weshalb es zumindest vorläufig im Fragebogen belassen wird.

Item 5 „Sicherheit im Umgang“ liegt von der Güte der Itemkennwerte her in beiden Untersuchungen im Vergleich mit den anderen Items in einem mittleren Bereich. Es handelt sich also um ein nicht besonders ‚gutes‘, aber auch nicht um ein wirklich ‚schlechtes‘ Item. Damit erscheint sowohl ein Beibehalt als auch ein Ausschluss des Items gerechtfertigt. Die Analyse der Kommentarfelder zeigte, dass viele Probanden dieses Item in Bezug auf die Vertrautheit mit der Informationsarchitektur beantworteten. Diese Beobachtung lässt sich zumindest bei der DEZA-Umfrage in der Tendenz durch die vergleichsweise hohe Korrelation von Item 5 mit Item 3 stützen (siehe Korrelationsmatrix). Diese vermutete Redundanz sowie die Absicht, eine homogenere Skala und damit eine klare Ein-Faktoren-Struktur zu erreichen, gibt den Ausschlag, das Item zu eliminieren.

Item 7 „Zugangsmöglichkeiten“ konnte sich nach der ersten Testung erneut nicht bewähren und wird deshalb aus dem Fragebogen gestrichen. Das Item macht eventuell Sinn vereinzelt in Unternehmen eingesetzt zu werden. Beispielsweise wenn die DEZA den Fragebogen auch in den Koordinationsbüros im Ausland einsetzen möchte oder sonst Zweifel bestehen, ob Mitarbeiter ihren Bedürfnissen entsprechend nicht immer zugreifen können (Bsp. Aussendienstmitarbeiter über Mobiltelefone usw.). Dort wo generell traditionelle Büroarbeitsplätze vorherrschen bringt es wenig Erkenntniszuwachs.

⁴¹ Im Sinne des Tagesschau-Jargons: Hier die aktuellsten Informationen zum Tagesgeschehen. In der kleinen Anzahl von Kommentaren, die zu diesem Item abgegeben wurden, liess sich kein Hinweis auf eine solche Verzerrung erkennen.

Item 8 „Nachfrage und Feedbackmöglichkeiten“ verzeichnet hohe Abstinenzen und verfügt in beiden Durchgängen über vergleichsweise schlechte Itemkennwerte. Das Item wird deshalb nicht mehr berücksichtigt.

Bei *Item 13* „Aktuelle Firmenneuigkeiten“ führen die zwei Ite-manalysen zu unterschiedlichen Ergebnissen. Beim Unternehmen A konnte sich das Item aufgrund seiner Itemkennwerte bewähren. Die Ergebnisse der DEZA-Umfrage sprechen eher für einen Ausschluss. Firmenkommunikation ist im Normalfall ein zentraler Bestandteil des Intranets. Inhaltliche Überlegungen und das Abschneiden des Items beim Unternehmen A sprechen für einen Verbleib des Items.

Die Werteverteilungen von *Item 15* „Vertrauen in Informationen“ tendieren beide Male in Richtung eines Deckeneffekts. Die Itemkennwerte liessen sich in ihrer Güte im Vergleich mit den anderen Items jeweils im unteren Mittel einordnen. In der Absicht, eine höhere Homogenität zu erreichen, wird das Item deshalb ausgeschlossen.

Das Item „Qualität Suchmaschine“ konnte aus erwähnten Gründen nicht ein zweites Mal getestet werden. Nach der ersten Ite-manalyse wies es relativ tiefe Trennschärfe- und Homogenitätskoeffizienten auf. Ein Ausschluss hätte zudem eine höhere interne Konsistenz zur Folge. Positiv zu vermerken sind die vergleichsweise hohe Streuung und die Eigenschaft, im unteren Skalenbereich zu differenzieren. Interessant erwies sich im Zusammenhang mit dem Suchmaschine-Item die Auswertung der Kommentarfelder der DEZA-Umfrage. Ein Viertel der Probanden (43 der 173 Probanden) äusserte sich im „allgemeinen Kommentarfeld“ direkt zum Thema Suchmaschine. Das Fehlen der Suchmaschine im Intranet wurde dabei als eine der grössten Schwachpunkte bezeichnet und mehrere Probanden forderten die Installation einer Suchmaschine. Dies belegt, dass die Suchmaschine als zentrales Element eines Intranets wahrgenommen wird und auch mit der Zufriedenheit zusammenhängen dürfte. Das „Suchmaschine-Item“ sollte deshalb nochmals getestet werden und das Item damit im Fragebogen belassen werden.

5.3.3 Vorläufige Endversion des Fragebogens

Nach der oben beschriebenen Itemrevision setzt sich der Fragebogen noch aus 13 Items zusammen. Im folgenden Kasten sind die Items der vorläufigen Endversion des Fragebogens aufgeführt. Im Anhang D4 sind die französische und italienische Version des *Intranet Satisfaction Questionnaire* abgedruckt. Es handelt sich dabei um Übersetzungen, welche zum Zeitpunkt der vorliegenden Arbeit noch nicht validiert wurden.

Vorläufige Endversion des Intranet Satisfaction Questionnaire

- 1 Das Intranet stellt für mich arbeitsrelevante Inhalte zur Verfügung.
- 2 Die Inhalte im Intranet sind aktuell.
- 3 Das Intranet ist übersichtlich gestaltet und die Struktur nachvollziehbar.
- 4 Wenn ich auf dem Intranet etwas lese, sind die Informationen so verfasst, dass sie klar und verständlich sind.
- 5 Das Intranet erleichtert die interne Kommunikation (z.B. durch das Mitarbeiter-Verzeichnis, Mitteilungen des Managements oder Diskussionsforen etc.).
- 6 Wenn ich etwas mit der Suchmaschine suche, finde ich die gewünschte Information innert nützlicher Frist.
- 7 Das Intranet ermöglicht mir effizientere Arbeitsweisen (z.B. firmeninterne Arbeitsabläufe, Support oder Informationsbeschaffung).
- 8 Mit dem Intranet lässt sich zügig arbeiten (z.B. schneller Seitenaufbau oder Dokumente herunterladen).
- 9 Das Intranet ist einfach zu bedienen (z.B. Personalisierung, Bedienen des Mitarbeiter-Verzeichnisses).
- 10 Ich bin zufrieden mit der Hilfe bzw. Unterstützung, die ich bekomme, wenn ich ein Problem oder eine Frage zum Intranet habe (z.B. Online-Hilfe oder Help-Desk).
- 11 Das Intranet versorgt mich laufend mit aktuellen Firmen-Neuigkeiten.
- 12 Die für meine Arbeit benötigten Informationen finde ich auf dem Intranet so vor, dass ich sie einfach weiterverarbeiten kann.
- 13 Insgesamt bin ich mit dem Intranet zufrieden.

6 Diskussion

Die Ergebnisse der Itemanalysen wurden in den Kapiteln 4 und 5 diskutiert und werden deshalb hier nicht wiederholt. Dieser Teil wird als abschliessende Diskussion der vorliegenden Arbeit verstanden.

Die zwei Testvalidierungen brachten eine Skala von 13 Items hervor, die sich zu eignen scheint, ein zuverlässiges und valides Bild über die Benutzerzufriedenheit der Mitarbeiter mit dem Intranet zu erstellen. Der Fragebogen gibt Auskunft, wo den Mitarbeitern der Schuh drückt und möglicherweise Massnahmen zur Verbesserung der Benutzerzufriedenheit getroffen werden sollten. Der Fragebogen kann helfen, Schwachpunkte in den wichtigsten Bereichen zu identifizieren, er liefert aber keine Informationen darüber, wieso Benutzer mit einem bestimmten Aspekt unzufrieden sind. Gewisse Rückschlüsse lassen sich eventuell aus Kommentaren der Benutzer ziehen, weshalb es sich empfiehlt, bei der Durchführung zu jedem Item ein Kommentarfeld bereitzustellen. In der Umfrage bei der DEZA konnte so bspw. aus den Kommentaren eruiert werden, welche Kategorien der Informationsarchitektur den Benutzern nicht klar sind. Je nach Komplexität eines Problembereichs sind aber weitere Analysen notwendig, um wirkungsvolle Verbesserungsmassnahmen zu bestimmen.

Das Vorgehen zur Testentwicklung war explorativ. Im Vorfeld wurden keine Annahmen über mögliche Faktoren bzw. Faktorenstrukturen getroffen. Es gibt nur eine bescheidene Anzahl validierter Messinstrumente zur Erhebung der Benutzerzufriedenheit. Die zwei bedeutendsten dürften von Doll und Torkzadeh (1988) sowie von Lewis (1995) stammen. Die Autoren gehen in ihren Instrumenten von mehr als einem Faktor zur Erfassung des Konstrukts Benutzerzufriedenheit aus. Die vorliegenden Daten unterstützen keine solche Mehrfaktoren-Struktur, sondern legen eine Ein-Faktoren-Struktur nahe. Wie in Kapitel 5 beschrieben, sollte die Skala jedoch nochmals in der bestehenden Endfassung validiert werden.

Die beiden untersuchten Stichproben können dem Dienstleistungssektor zugerechnet werden. Interessant wäre deshalb eine Validierung bei einem Unternehmen aus der Industrie bzw. aus dem sekundären Sektor. Zudem könnte geprüft werden, ob sich der Test für verschiedene Subgruppen eines Unternehmens wie bspw. Angestellte oder Kader gleichermaßen eignet. In beiden durchgeführten Umfragen zeigte sich, dass sich die Art der Beantwortung einzelner Items zwischen den Sprachgruppen signifikant unterschied. Insbesondere bei mehrsprachigen Intranets scheint es deshalb angezeigt, Umfrageergebnisse

für die Interpretation stets auch auf solche Unterschiede zu überprüfen. Ansonsten besteht die Gefahr, mögliche Unterschiede in der Zufriedenheit gegenüber einzelner Aspekte zu übersehen. Wird bspw. die Online-Hilfe nur in einer Sprache angeboten, könnte dies zur Folge haben, dass die Benutzer anderer Sprachgruppen weniger zufrieden mit der Hilfe sind.

Eine andere Gefahr besteht darin, dass trotz des relativ kleinen Umfangs der Umfrage, Personen, die sich häufig unter Zeitdruck wähnen, weniger bereit sind, die Umfrage auszufüllen. Trifft dies in einer Umfrage zu, fehlt eine Gruppe von Personen, die tendenziell das Intranet womöglich kritischer beurteilt bzw. schneller unzufrieden reagiert, da sie in der Interaktion mit dem Intranet ungeduldiger ist.

Der Fragebogen stützt sich theoretisch auf das Erwartungs-mal-Wert-Prinzip (Fishbein, 1967a, 1967b). Kritisiert werden könnte, dass die Erwartungen der Benutzer nicht separat erhoben wurden, sondern durch die vorhandene Literatur und bestehende Messinstrumente erschlossen wurde. Die Erwartungen der Benutzer zu erheben, wäre sehr aufwändig gewesen und schien aufgrund der bereits bestehenden Literatur nicht zwingend. So ist es fraglich, ob eine Umfrage zu den Erwartungen der Benutzer neue Erkenntnisse erbracht hätte.

Die Länge des Fragebogens erlaubt es, diesen jeweils noch mit firmenspezifischen oder mit Items, die einen bestimmten Aspekt noch vertiefend erfassen sollten, zu ergänzen. Er kann für sich alleine durchgeführt oder bei Bedarf, von Firmen als ein Grundmodul für eine Intranetumfrage genutzt werden. Der Fragebogen erweist sich damit als relativ ‚handlich‘. Wird die Umfrage online durchgeführt, was sich in diesem Fall anbietet, lässt sich die Umfrage ohne grossen Aufwand bewerkstelligen. Untersuchungsleiter sollten aber beachten, dass nach neueren Forschungsergebnissen die Bereitschaft von Mitarbeitenden, an internen Umfragen teilzunehmen, sinkt. Nach Stanton und Rogelberg (2002, S. 276) ist dies möglicherweise auf eine Art *survey fatigue* der Mitarbeitenden zurückzuführen. Besonders durch die einfache Handhabung und den geringen Aufwand webbasierter Umfragen besteht die Gefahr, den Mitarbeitern zu viele Umfragen vorzusetzen. Rogelberg und Mitarbeiter (2002, S. 148) empfehlen Unternehmen deshalb die Einrichtung einer zentrale Stelle, die sämtliche Umfragen koordiniert und verhindert, dass Mitarbeiter *overinterviewed* werden.

Um aussagekräftige Resultate zu bekommen, sollte auch der Erhebungszeitpunkt gut gewählt werden. Falls beispielsweise Änderungen in der Navigation vorgenommen wurden, sollte den Mitarbeitern Zeit gegeben werden, sich an diese zu gewöhnen, bevor man die Umfrage durchführt. Auch wenn im Vorfeld der Umfrage öfters (unübliche) technische Probleme auftraten, könnten die Ergebnisse aufgrund dieser nahe liegenden Erfahrungen

verzerrt werden. Eine Umfrage sollte also zu einem möglichst repräsentativen Zeitpunkt stattfinden.

Der Fragebogen lässt sich für normative Analysen, in der Praxis oft als *Benchmarking* bezeichnet, nutzen. Es geht dabei darum, neu gewonnene Daten mit internen oder externen Daten zu vergleichen, also bspw. die eigenen Antworten den Antworten eines anderen Unternehmens oder den letztjährigen Antworten gegenüberzustellen. Diese Vergleichsmöglichkeiten sollen helfen, beobachtete Daten besser einordnen zu können. Die Stimmt AG (2003, 2005) verglich so die Ergebnisse verschiedener Unternehmen aus den Intranetumfragen mit dem CSUQ (Lewis, 1995). Unternehmen war es damit möglich, ihr Abschneiden mit dem CSUQ als Ganzes oder in einzelnen Fragen mit anderen Unternehmen zu vergleichen. Besonders in der Wirtschaft sind solche *Benchmarkings* populär. Rogelberg und Mitarbeiter (2002) äussern sich aufgrund der mangelnden Äquivalenz zwischen Unternehmen allerdings kritisch zu diesen Vergleichen. Hinter dieser Äquivalenzkritik steht das Grundproblem, dass man ungleiche Gruppen nicht vergleichen sollte. Die Mitarbeiter zweier Unternehmen unterscheiden sich aber zwangsläufig immer (bspw. auf demografischer, geografischer oder sozioökonomischer Ebene). Konsequenterweise lassen sich Unterschiede zwischen beobachteten Daten und Daten anderer Unternehmen nicht allein auf das Intranet des Unternehmens zurückführen. Das Intranet einer Firma schneidet so eventuell nicht deshalb schlechter ab als das Intranet einer anderen Firma weil, es ‚objektiv‘ schlechter ist, sondern weil die Erwartungen seiner Mitarbeiter höher sind, als in der anderen Firma. Rogelberg und Mitarbeiter (2002) weisen zudem auf die Gefahr hin, aufgrund von Vergleichen Probleme zu marginalisieren oder gar zu ignorieren und auf entsprechende Verbesserungsmassnahmen zu verzichten. Wird der Fragebogen als Benchmarkumfrage genutzt, um die Entwicklung der Benutzerzufriedenheit über die Jahre zu dokumentieren, sollte bedacht werden, dass sich die Erwartungen der Benutzer in Bezug auf das Intranet erhöhen können. So ist es möglich, dass die Benutzerzufriedenheit weniger ansteigt, als man vielleicht aufgrund getroffener Verbesserungsmassnahmen erwartet hätte. Vergleiche anhand von Benchmarks können zwar durchaus interessant sein, Schlussfolgerungen sollten jedoch mit der nötigen Vorsicht gezogen werden.

Das Intranet wird sich in den nächsten Jahren sowohl technisch als auch inhaltlich weiterentwickeln. Die Stimmt AG (2005) geht davon aus, dass die Bedeutung der Intranets weiter zunimmt, sei es als Informations- und Kommunikationskanal oder als zentrale Drehscheibe von unterstützenden Aktivitäten. Innerhalb dieser Entwicklungsdynamik

verändern sich auch die Bedürfnisse der Benutzer. Um zufriedene Intranetbenutzer hervorzubringen, ist es deshalb wichtig, solche Bedürfnisse zu kennen bzw. zu erheben, zu priorisieren und entsprechende Lösungen umzusetzen (Stimmt AG, 2005, S. 80).



Abbildung 6.1. Illustration übernommen aus Huber (1997, S. 123).

Benutzerzufriedenheit im Intranet ist eine komplexe Angelegenheit, für deren Erreichung Kenntnisse verschiedener Disziplinen von Nöten sind. Einen „goldenen Weg“ zur Schaffung hoher Benutzerzufriedenheit gibt es nicht. Sowohl für die Wissenschaft wie für die Praxis ist es deshalb wichtig, ein Messinstrument zur Hand zu haben, mit welchem die Benutzerzufriedenheit zuverlässig erhoben werden kann. Die vorliegende Arbeit leistet hierzu mit dem *Intranet Satisfaction Questionnaire* einen Beitrag.

7 Literaturverzeichnis

- Al-Gahtani, S. S. & King, M. (1999). Attitudes, satisfaction and usage: Factors contributing to each in the acceptance of information technology. *Behaviour & Information Technology*, 18 (4), 277–297.
- Allison, P. D. (2001). *Missing Data. Sage University Papers: Series on quantitative applications in the social sciences*, 07–136. Thousands Oaks, CA: Sage.
- Bailey, J. E. & Pearson S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, 29 (5), 530–545.
- Baitsch, C., Katz, C., Spinass, P. & Ulich, E. (1989). *Computerunterstützte Büroarbeit. Ein Leitfaden für Organisation und Gestaltung*. Zürich: vdf.
- Batinic, B. (2004). Netzwerkbasierende Mitarbeiterbefragungen. In G. Hertel & U. Konradt (Hrsg.), *Human Resource Management im Inter- und Intranet. Reihe Internet und Psychologie: Neue Medien in der Psychologie, Bd. 7* (S. 220–234). Göttingen: Hogrefe.
- Batinic, B. & Bosnjak, M. (2000). Fragebogenuntersuchungen im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (2., überarbeitete und erweiterte Auflage, S. 287–317). Göttingen: Hogrefe.
- Batinic, B. & Moser, K. (2005). Determinanten der Rücklaufquote in Online-Panels. *Zeitschrift für Medienpsychologie*, 17 (2), 64–75.
- Borg, I. (2001). Mitarbeiterbefragungen. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 373–396). Göttingen: Hogrefe.
- Borg, I. & Staufenbiel, T. (1997). Theorien und Methoden der Skalierung. Eine Einführung (3., überarbeitete Auflage). In K. Pawlik (Hrsg.), *Methoden der Psychologie, 11*. Bern: Huber.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5., vollständig überarbeitete Auflage). Berlin: Springer.
- Bühl, A. & Zöfel, P. (2005). *SPSS 12. Einführung in die moderne Datenanalyse unter Windows*. München: Pearson Studium.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bundesamt für Informatik und Telekommunikation BIT (2003). *Internet Technische Grundlagen*. o.O: ohne Verlagsangabe.
- DeLone, H. W. & McLean, R. E. (1992). Information system success: The quest for the dependent variable. *Information Systems Research*, 3 (1), 60–95.

- Deutschmann, M. (1999). Mitarbeiterbefragungen via Intranet. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser & A. Werner (Hrsg.), *Current Internet science – trends, techniques, results*. Zürich: Online Press. [WWW document]. Available URL: http://gor.de/gor99/tband99/pdfs/a_h/deutschmann_ad.pdf [6. Jan. 2006].
- Dillman, D. A., Tortora, R. D. & Bowker, D. (1999). *Principles for Constructing Web Surveys*. [WWW document]. Available URL: <http://survey.sesrc.wsu.edu/dillman/papers/websurveyppr.pdf> [5. Jan. 2006].
- Doll, W. J. & Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS Quarterly*, 12 (2), 259–274.
- Doll, W. J., Deng, X., Raghunathan, T. S., Torkzadeh, G. & Xia, W. (2004). The meaning and measurement of user satisfaction: A multigroup invariance analysis of the end-user computing satisfaction instrument. *Journal of Management Information Systems*, 21 (1), 227–262.
- Eagly, E. A. & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 269–322). New York: Mc Graw-Hill.
- Fishbein, M. (1967a). A behavior theory approach to the relations between beliefs about an object and the attitude toward the object. In M. Fishbein (Ed.), *Readings in attitude theory and measurement* (pp. 389–400). New York: Wiley.
- Fishbein, M. (1967b). A consideration of beliefs, and their role in attitude measurement. In M. Fishbein (Ed.), *Readings in attitude theory and measurement* (pp. 257–266). New York: Wiley.
- Fisher, S. L. & Howell, A. W. (2004). Beyond user acceptance: an examination of employee reactions to information technology systems. *Human Resource Management*, 43 (2 & 3), 243–258.
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik* (2., überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.
- Frick, A., Bächtiger, M. T. & Reips, U.-D. (1999). Financial incentives, personal information and drop-out in online studies. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser & A. Werner (Hrsg.), *Current Internet science – trends, techniques, results*. Zürich: Online Press. [WWW document]. Available URL: http://gor.de/gor99/tband99/pdfs/a_h/frick.pdf [6. Jan. 2006].
- Fürntratt, E. (1969). Zur Bestimmung der Anzahl interpretierbarer gemeinsamer Faktoren in Faktorenanalysen psychologischer Daten. *Diagnostika*, 15, 62–75.

- Gizycki, G. von (2002). Usability – Nutzerfreundliches Web-Design. In M. Beier & V. von Gizycki (Hrsg.), *Usability. Nutzerfreundliches Web-Design* (S. 1–17). Berlin: Springer.
- Goldstein, E. B. (1997). *Wahrnehmungspsychologie. Eine Einführung*. Heidelberg: Spektrum.
- Granello, D. H. & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counseling & Development*, 82, 387–393.
- Guadagnoli, E. & Velicer, W.F. (1988). Relation of sample size to the component patterns. *Psychological Bulletin*, 103, 265–275.
- Harrison, A. W. & Rainer, K. R. (1996). A general measure of user computing satisfaction. *Computers in Human Behavior*, 12 (1), 29–92.
- Herczeg, M. (1994). *Software-Ergonomie. Grundlagen der Mensch-Computer-Kommunikation*. Bonn: Addison Wesley.
- Hoffmann, C. (2001). *Das Intranet. Ein Medium der Mitarbeiterkommunikation*. Konstanz: UVK.
- Huang, J.-H., Yang, C, Jin, B.-H. & Chiu, H. (2004). Measuring satisfaction with business-to-employee systems. *Computers in Human Behavior*, 20, 17–35.
- Huber, O. (1997). *Das psychologische Experiment: Eine Einführung* (2. Auflage). Bern: Huber.
- International Standardization Organisation (1998). *Ergonomics requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*. (ISO 9241–11:1998(E)).
- Kaiser, T.M. (2000). *Methode zur Konzeption von Intranets*. Dissertation, Universität St. Gallen. Bamberg: Difo-Druck OHG.
- Katz, D. (1969). *Gestaltpsychologie*. Basel: Schwabe.
- Kohlisch, O. & Kuhmann, W. (1997). System response time and readiness for task execution – the optimum duration of inter-task delays. *Ergonomics*, 40 (3), 265–280.
- Konradt, U., Lehmann, K., Böhm-Rupprecht, J. & Hertel, G. (2003). Computer- und internetbasierte Verfahren der Berufseignungsdiagnostik. Ein empirischer Überblick. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment* (S. 105–124). Göttingen: Hogrefe.
- Kortzfleisch, F. O. von (1997). Kooperieren und Lernen im Intranet. *Information Management*, 2, 28–35.
- Kranz, H. T. (1979). *Einführung in die klassische Testtheorie. Methoden in der Psychologie*, Bd. 8. Frankfurt am Main: Fachbuchhandlung für Psychologie GmbH.
- Leonhart, R. (2004). *Lehrbuch Statistik. Einstieg und Vertiefung*. Bern: Huber.

- Leung, H. K. N. (2001). Quality metrics for intranet applications. *Journal of Information Technology*, 38, 137–152.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires : Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7 (1), 57–78.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Psychologie Verlags Union.
- Lynch, P. J. & Horton, S. (2002). *Web style guide: Basic design principles for creating web sites* (2nd ed.). New Haven: Yale University Press.
- McKeen, J. D., Guimaraes, T. & Wetherbe, J. C. (1994). The relation between user participation and user satisfaction: An investigation of four contingency factors. *MIS Quarterly*, 18 (4), 427–451.
- McKinney, V., Yoon, K. & Zahedi, F. M. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research*, 13 (3), 296–315.
- Meier, P. (2002). *Interne Kommunikation im Unternehmen. Von der Hauszeitung bis zum Intranet*. Zürich: Füssli.
- Melone, N. P. (1990). A theoretical assessment of the user-satisfaction construct in information system research. *Management Science*, 36 (1), 76–91.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63 (2), 81–97.
- Mummendey, H. D. (1995). *Die Fragebogen-Methode* (2., korrigierte Auflage). Göttingen: Hogrefe.
- Nielsen, J. (1997). *The difference between intranet and internet design* [WWW document]. Available URL: <http://www.useit.com/alertbox/9709b.html> [15. Feb. 2006].
- Nielsen, J. (2006). *10 best intranets of 2006* [WWW document]. Available URL: http://www.useit.com/alertbox/intranet_design.html [15. Feb. 2006].
- Nielsen, J. & Tahir, M. (2002). *Homepage Usability. 50 websites deconstructed*. Indianapolis, IN: New Riders Publishing.
- Nievergelt, J. (1982). Errors in dialog design and how to avoid them. *Proceedings of the 1982 international Zurich seminar on digital communications*, 94–99.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*, 32 (3), 396–402.

- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Palvia, P. C. (1996). A model and instrument for measuring small business user satisfaction with information technology. *Information & Management*, 31, 151–163.
- Phleps, R. & Mok, M. (1999). Managing the risks of intranet implementation: An empirical study of user satisfaction. *Journal of Information Technology*, 14, 39–52.
- Pott, O. (1998). *Intranet-Bibel. Das Standardwerk zur Intranet-Praxis*. Kilchberg: SmartBooks Publishing AG.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1999). *Human-Computer Interaction*. Harlow: Addison-Wesley.
- Rauterberg, M., Spinas, P., Strohm, O., Ulich, E. & Waeber, D. (1994). *Benutzerorientierte Software-Entwicklung. Konzepte, Methoden und Vorgehen zur Benutzerbeteiligung. Mensch Technik Organisation, Bd. 3*. Zürich: vdf.
- Reips, U.-D. (2000). Das psychologische Experimentieren im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 319–343). Göttingen: Hogrefe.
- Rogelberg, S. G., Church, A. H., Waclawski, J. & Stanton, J. M. (2002). Organizational survey research. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 141–160). Oxford: Blackwell Publishers.
- Rosenberg, M. J. & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson & J. W. Brehm (Eds.), *Attitude organization and change* (pp. 1–14). New Haven, CT: Yale University Press.
- Rosson, M. B. & Carroll, J. M. (2002). *Usability engineering. Scenario-based development of human-computer interaction*. San Diego, CA: Morgan Kaufmann Publishers.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollständig überarbeitete und erweiterte Auflage). Bern: Huber.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.
- Schmotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Zeitschrift für Experimentelle Psychologie*, 46 (3), 217–236.
- Schneider, W. (2000). *Kundenzufriedenheit. Strategie, Messung, Management*. Lech: Verlag Moderne Industrie.

- Schnotz, W. (2002). Wissenserwerb mit Texten, Bildern und Diagrammen. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia im Internet. Lehrbuch für Studium und Praxis* (3., vollständig überarbeitete Auflage, S. 65–81). Weinheim: Beltz Psychologie Verlags Union.
- Shneiderman, B. (1998). *Designing the user interface* (3rd. ed.). Reading, Mass.: Addison-Wesley.
- Spinas, P. (1987). *Arbeitspsychologische Aspekte der Benutzerfreundlichkeit von Bildschirmsystemen*. Zürich: Administration & Druck AG.
- SPSS GmbH Software (2003). *SPSS Base 12.0 Benutzerhandbuch*. München: SPSS Inc.
- Stanton, J. M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, 51, 709–725.
- Stanton, J. M. & Rogelberg, S. G. (2002). Beyond online survey: Internet research opportunities for industrial-organizational psychology. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 275–294). Oxford: Blackwell Publishers.
- Stimmt AG (2003). *Stimmt Intranet Report 2003. Eine Analyse Schweizer Intranets*. o.O: ohne Verlagsangabe.
- Stimmt AG (2005). *Stimmt Intranet Report 2004*. o.O: ohne Verlagsangabe.
- Thompson, L. F., Surface, E. A., Martin, D. L. & Sanders, M.G. (2003). From paper to pixels: Moving personnel surveys to the web. *Personnel Psychology*, 56, 197–227.
- Töpfer, A. & Mann, A. (1999). Kundenzufriedenheit als Messlatte für den Erfolg. In A. Töpfer (Hrsg.), *Kundenzufriedenheit. Messen und steigern* (2., erweiterte und überarbeitete Auflage, S. 59–110). Neuwied: Luchterhand.
- Tullis, T. & Wood, L. (2004). *How many users are enough for a card sorting study?* Paper presented at the Usability Professionals Association Conference. Minneapolis, MN.
- Ulich, E. (2001). *Arbeitspsychologie* (5., vollständig überarbeitete und erweiterte Auflage). Zürich: vdf.
- Wang, R. Y. & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12 (4), 5–34.
- Wertheimer, M. (1921). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, 1, 47–58.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 4, 301–351.

World Wide Web Consortium (W3C) (1999). *Web Content Accessibility Guidelines 1.0*
[WWW document] Available URL: <http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505/> [20. Feb. 2006].

8 Anhang

Anhang A: Theorieteil	S. I–III
Anhang B: Methode	S. IV–VII
Anhang C: Testvalidierung I	S. VIII–XVII
Anhang D: Testvalidierung II	S. XVIII–XXV

Anhang A: Theorie

Anhang A1

Modell und Items des Messinstruments von Doll und Torkzadeh (1988).

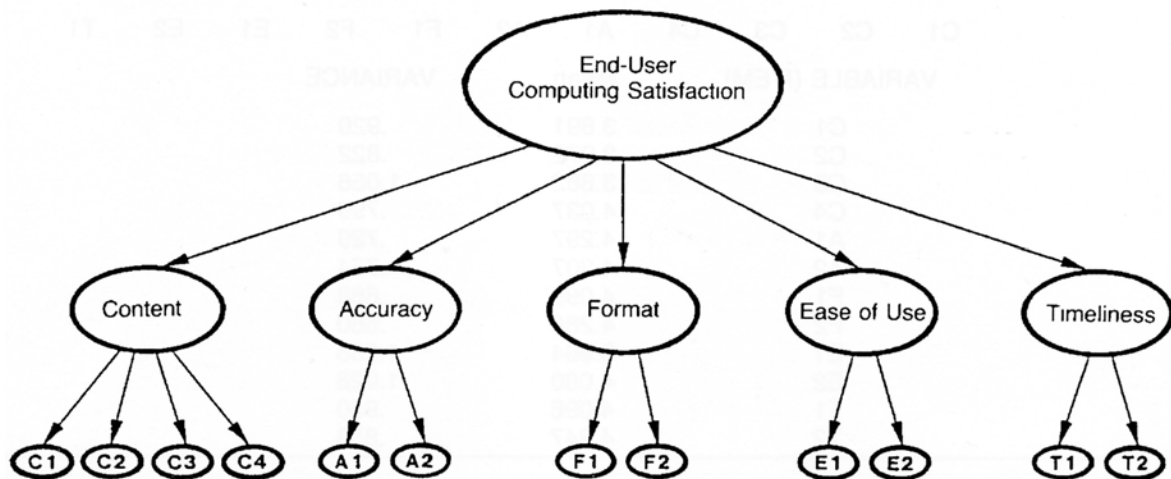


Abbildung A1.1. Model for Measuring End-User Computing Satisfaction von Doll und Torkzadeh (1988, S. 268).

Itemübersicht:

Content

- C1 Does the system provide the precise information you need?
- C2 Does the information content meet your needs?
- C3 Does the system provide reports that seem to be just about exactly what you need?
- C4 Does the system provide sufficient information?

Accuracy

- A1 Is the system accurate?
- A2 Are you satisfied with the accuracy of the system?

Format

- F1 Do you think the output is presented in a useful format?
- F2 Is the information clear?

Ease of Use

- E1 Is the system user friendly?
- E2 Is the system easy to use?

Timeliness

- T1 Do you get the information you need in time?
- T2 Does the system provide up-to-date information?

Skalierung: Likertskalierung mit den fünf Stufen 1 = almost never; 2 = some of the time; 3 = about half of the time; 4 = most of the time; 5 = almost always.

Anhang A2

Itemübersicht des IBM Computer System Usability Questionnaires (CSUQ) von Lewis (1995, S. 77f.).

1. Overall, I am satisfied with how easy it is to use this system.
2. It is simple to use this system.
3. I can effectively complete my work using this system.
4. I am able to complete my work quickly using this system.
5. I am able to efficiently complete my work using this system.
6. I feel comfortable using this system.
7. It was easy to learn to use this system.
8. I believe I became productive quickly using this system.

9. The system gives error messages that clearly tell me how to fix problems.
10. Whenever I make a mistake using the system, I recover easily and quickly.
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system is clear.
12. It is easy to find the information I need.
13. The information provided with the system is easy to understand.
14. The information is effective in helping me complete my work.
15. The organization of information on the system screen is clear.

16. The interface of this system is pleasant.
17. I like using the interface of this system.
18. This system has all the functions and capabilities I expect it to have.

19. Overall, I am satisfied with this system.

Overall Satisfaction = Item 1 bis 19

System Usefulness = Item 1 bis 8

Information Quality = Item 9 bis 15

Interface Quality = Item 16 bis 18

Skalierung: Siebenstufige Likertskala mit den vorgegeben Polen *strongly agree* und *strongly disagree*.

Anhang A3

Itemübersicht der zwei Skalen *Perceived Usefulness* und *Perceived Ease of Use* von Davis (1998, S. 340).

Perceived Usefulness:

1. Using CHART-MASTER* in my job would enable me to accomplish tasks more quickly.
2. Using CHART-MASTER would improve my job performance.
3. Using CHART-MASTER in my job would increase my productivity.
4. Using CHART-MASTER would enhance my effectiveness on the job.
5. Using CHART-MASTER would make it easier to do my job.
6. I would find CHART-MASTER useful in my job.

Perceived Ease of Use:

1. Learning to operate CHART-MASTER would be easy for me.
2. I would find it easy to get CHART-MASTER to do what I want to do.
3. My interaction with CHART-MASTER would be clear and understandable.
4. I would find CHART-MASTER to be flexible to interact with.
5. It would be easy for me to become skillful at using CHART-MASTER.
6. I would find CHART-MASTER easy to use.

**Anmerkung.* Der Begriff CHART-MASTER kann je nach Verwendung des Fragebogens mit einem anderen Begriff ersetzt werden.

Skalierung: Siebenstufige Likertskala mit den Polen *likely* und *unlikely*. Die Stufen werden vorgeben mit *extremely*, *quite*, *slightly*, *neither*, *slightly*, *quite extremely*.

Anhang B: Methode

Anhang B1

Überblick der Items des Testvorentwurfs

1	Das Intranet stellt für mich arbeitsrelevante Inhalte zur Verfügung.
2	Die Informationen auf dem Intranet sind aktuell.
3	Die Inhalte auf dem Intranet sind so angeordnet, dass es einfach ist, sich zu orientieren und gewünschte Inhalte zu finden.
4	Die Art und Weise, wie Informationen auf dem Intranet vermittelt werden, ist klar und gut verständlich.
5	Ich bin genügend vertraut im Umgang mit dem Intranet, so dass ich es optimal für meine Bedürfnisse nutzen kann.
6	Das Intranet erleichtert die interne Kommunikation (z.B. durch das Mitarbeiter-Verzeichnis, Mitteilungen des Managements oder Diskussionsforen etc.).
7	Bei Bedarf, kann ich jederzeit auf das Intranet zugreifen.
8	Ich bin zufrieden mit der Qualität der Suchmaschine. Sie liefert gute und brauchbare Resultate und eignet sich für das gezielte Auffinden von bestimmten Inhalten oder Dokumenten.
9*	Es ist einfach, mit der verantwortlichen Person eines Inhalts Kontakt aufzunehmen.
10	Das Intranet ermöglicht mir effizientere Arbeitsweisen (z.B. firmeninterne Arbeitsabläufe, Support oder Informationsbeschaffung).
11*	Ich weiss, welche Mitteilungen oder Dokumente von mir auf das Intranet gestellt werden sollten.
12	Falls ich eine Mitteilung oder ein Dokument auf das Intranet stellen möchte, ist mir klar, wie ich vorgehen muss.
13*	Mit dem Intranet lässt sich zügig arbeiten (z.B. Navigieren oder Dokumente herunterladen).
14	Das Intranet ist einfach zu bedienen (z.B. Personalisierung, Bedienen des Mitarbeiter-Verzeichnisses usw.).
15	Ich bin zufrieden mit der Hilfe bzw. der Unterstützung, die ich bekomme, wenn ich ein Problem oder eine Frage zum Intranet habe (z.B. Helpdesk oder Online-Hilfe).
16	Das Intranet versorgt mich laufend mit aktuellen Firmen-Neuigkeiten.
17	Informationen, die ich aus dem Intranet beziehe, kann ich ohne Modifikationen für meine Arbeit weiter verwenden.
18*	Ich kann mich auf die Informationen im Intranet verlassen und darauf vertrauen, dass sie richtig sind.
19*	Ich finde das Intranet ein wichtiges und nützliches Arbeitsinstrument.
20*	Im Grossen und Ganzen bin ich mit dem Intranet zufrieden.

* Items die sich von der ersten Testversion unterscheiden, siehe Methodenteil.

Anhang B2:

Wiedergabe des E-Mails zur Rekrutierung der Probanden in der DEZA, original in Deutsch und Französisch, hier exemplarisch in Deutsch.

Sind Sie mit dem IntraWeb zufrieden?

Gerne möchten wir dies von Ihnen erfahren und Sie deshalb bitten, diese Umfrage auszufüllen. Mit Ihrer Teilnahme tragen Sie wesentlich dazu bei, das Intranet zu verbessern.

Das Ausfüllen dieses Fragebogens nimmt ca. 10 Minuten in Anspruch. Ihre Angaben werden selbstverständlich streng vertraulich behandelt.

<http://phpserver.psycho.unibas.ch/deza/>

Vielen Dank für Ihre Mitarbeit.

Sabine Schumacher, Intranet-Verantwortliche, MuK
Jonas Lötscher, Institut für Psychologie der Universität Basel

PS: In Ergänzung zur Umfrage, die im Oktober 2005 stattgefunden hat, dient diese zweite Umfrage nicht nur dazu, weitere Hinweise für Verbesserungsmöglichkeiten zu erhalten, sondern ermöglicht zusätzlich, einen neuen und optimierten Fragebogen zu testen, der an der Universität Basel entwickelt wird.

Wiedergabe der Instruktion bei der DEZA, original in Deutsch und Französisch hier exemplarisch in Deutsch.

Sind Sie mit Ihrem Intranet zufrieden?

Gerne möchten wir dies von Ihnen erfahren und Sie deshalb bitten, diese Umfrage auszufüllen. Mit Ihrer Teilnahme tragen Sie wesentlich dazu bei, das Intranet zu verbessern.

Das Ausfüllen dieses Fragebogens sollte ca. 10 Minuten in Anspruch nehmen. Ihre Angaben werden selbstverständlich streng vertraulich behandelt.

Bitte lesen Sie jede der folgenden Aussagen und geben Sie an, wie stark Sie dieser zustimmen. Können Sie eine Aussage nicht beurteilen, kreuzen Sie bitte "k.A." an. Zu jeder Aussage haben Sie auch die Möglichkeit, einen Kommentar abzugeben.

Ein Intranet entwickelt sich ständig weiter, versuchen Sie deshalb so gut es geht, Ihre Antworten auf den momentanen Zustand zu beziehen.

Vielen Dank für Ihre Mitarbeit.

[Text nach Beendigung der Umfrage]

Herzlichen Dank für Ihre Teilnahme! Die Resultate dieser Umfrage werden ca. Ende... im Intranet publiziert.

Sabine Schumacher, Intranet-Team
Jonas Lötscher, Institut für Psychologie der Universität Basel

Wiedergabe des Erinnerungsschreibens in der DEZA, original in Deutsch und Französisch, hier exemplarisch in Deutsch.

Für ein besseres IntraWeb

Seit zwei Wochen können Sie an unserer Online-Befragung zum Thema Intranet teilnehmen (s. Mail vom 27.1.06). Ein Grossteil der angeschriebenen Personen hat den Fragebogen bereits ausgefüllt. Wir danken allen Teilnehmenden herzlich für die Mitarbeit. Falls Sie noch nicht teilgenommen haben und sich noch zu diesem Thema äussern möchten, dann haben Sie bis am Mittwoch, 15. Februar 2006 die Möglichkeit dazu. Klicken Sie dazu einfach auf den nachfolgenden Link:

<http://phpserver.psych.unibas.ch/deza/>

Für Ihre Teilnahme an dieser Befragung danken wir Ihnen herzlich und stehen Ihnen für weitere Auskünfte gerne zur Verfügung.

Freundliche Grüsse

Intranet-Team
Sabine Schumacher
Institut für Psychologie, Universität Basel
Jonas Loetscher

Wiedergabe der Instruktion im Unternehmen A, original in Deutsch, Französisch und Italienisch, hier exemplarisch in Deutsch.

Sind Sie mit Ihrem Intranet zufrieden?

Gerne möchten wir dies von Ihnen erfahren und Sie deshalb bitten, diese Umfrage auszufüllen. Mit Ihrer Teilnahme tragen Sie wesentlich dazu bei, das Intranet zu verbessern.

Das Ausfüllen dieses Fragebogens sollte ca. 15 Minuten in Anspruch nehmen. Ihre Angaben werden selbstverständlich streng vertraulich behandelt.

Bitte lesen Sie jede der folgenden Aussagen und geben Sie an, wie stark Sie dieser zustimmen. Können Sie eine Aussage nicht beurteilen, kreuzen Sie bitte "k.A." an. Zu jeder Aussage haben Sie auch die Möglichkeit, einen Kommentar abzugeben.

Ein Intranet entwickelt sich ständig weiter, versuchen Sie deshalb so gut es geht, Ihre Antworten auf den momentanen Zustand zu beziehen.

Vielen Dank für Ihre Mitarbeit.

[Text nach Beendigung der Umfrage:]

Herzlichen Dank für Ihre Teilnahme! Die Resultate dieser Umfrage werden ca. Ende... im Intranet publiziert.

Das Intranet Competence Center

Wiedergabe des Erinnerungsschreibens im Unternehmen A, original in Deutsch, Französisch und Italienisch, hier exemplarisch in Deutsch.

Sind Sie mit Ihrem Intranet zufrieden? Danke für die Teilnahme und Erinnerung

Sehr geehrte Mitarbeiterin, sehr geehrter Mitarbeiter

Seit einer Woche können Sie an unserer Online-Befragung zum Thema Intranet teilnehmen. Ein Grossteil der angeschriebenen Personen hat den Fragebogen bereits ausgefüllt. Wir danken allen Teilnehmern herzlich für die Mitarbeit. Falls Sie noch nicht teilgenommen haben und sich noch zu diesem Thema äussern möchten, dann haben Sie bis am Dienstag, 8. November 2005 die Möglichkeit dazu. Klicken Sie dazu einfach auf den nachfolgenden Link:

<http://www.unternehmenA/uc/intranet/>

Für Ihre Teilnahme an dieser Befragung danken wir Ihnen herzlich und stehen Ihnen für weitere Auskünfte gerne zur Verfügung.

Freundliche Grüsse

Das Intranet Competence Center

Anhang C

Anhang C1

Die Tabellen C1.1 bis C1.3 listen die erhobenen Angaben zu den Probanden aus der Stichprobe des Unternehmens A auf.

Tabelle C1.1

Altersstruktur der Probanden aus der Stichprobe des Unternehmens A

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unter 30 Jahre	221	25.1	25.1	25.1
	31 - 40 Jahre	276	31.3	31.3	56.4
	41 - 50 Jahre	246	27.9	27.9	84.3
	Über 51 Jahre	138	15.7	15.7	100.0
	Total	881	100.0	100.0	

Tabelle C1.2

Überblick der Dauer des Angestelltenverhältnisses der Probanden aus der Stichprobe des Unternehmens A

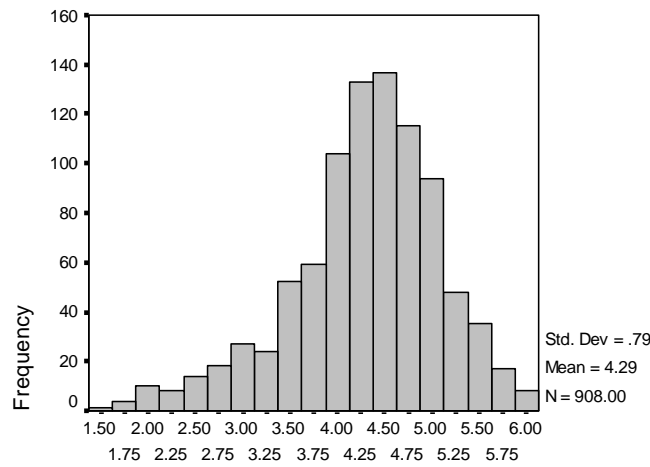
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Weniger als 1 Jahr	73	8.3	8.3	8.3
	1 bis 5 Jahre	236	26.8	26.8	35.1
	6 bis 10 Jahre	205	23.3	23.3	58.3
	Mehr als 10 Jahre	367	41.7	41.7	100.0
	Total	881	100.0	100.0	

Tabelle C1.3

Überblick der Funktionen der Probanden aus der Stichprobe des Unternehmens A

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Lernende(r), Career Starter	48	5.4	5.4	5.4
	Mitarbeiter(in)	527	59.8	59.8	65.3
	Kader	262	29.7	29.7	95.0
	Direktionsmitglied	44	5.0	5.0	100.0
	Total	881	100.0	100.0	

Anhang C2



Mittelwertskala Item 1 bis 18

Abbildung C2.1. Häufigkeitsverteilung der Mittelwertsskala (Item 1 bis 18) der Rohwerte aus dem Unternehmen A (N = 908).

Mittelwertskala Item 1 bis 18 Stem-and-Leaf Plot

Frequency	Stem &	Leaf
28.00	Extremes	(=<2.5)
.00	2 .	.
38.00	2 .	567889
61.00	3 .	00112233444
119.00	3 .	55556667777888889999
270.00	4 .	0000000011111111112222222233333333344444444
239.00	4 .	555555556666666666777777778888888889999
110.00	5 .	000001111222233344
39.00	5 .	55567&
4.00	6 .	0

Stem width: 1.00
Each leaf: 6 case(s)

& denotes fractional leaves.

Abbildung C2.2. SPSS-Output des Stängel-Blatt-Diagramms von der Mittelwertsskala (Item 1 bis 18) der Rohwerte des Unternehmens A (N = 908).

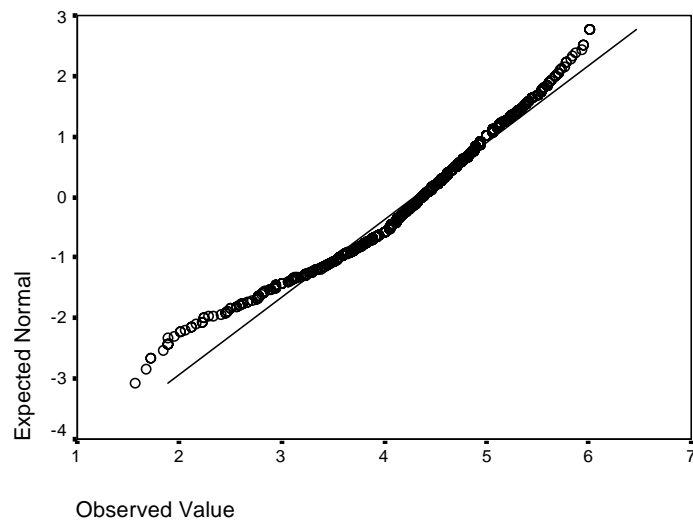


Abbildung C2.3. Normalverteilungsdiagramm (Q-Q-Diagramm) von der Mittelwertskala (Item 1 bis 18) der Rohwerte aus dem Unternehmen A ($N = 908$).

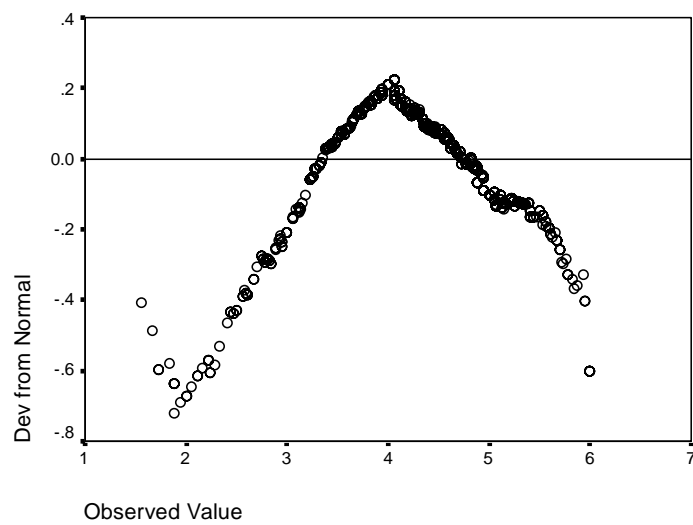
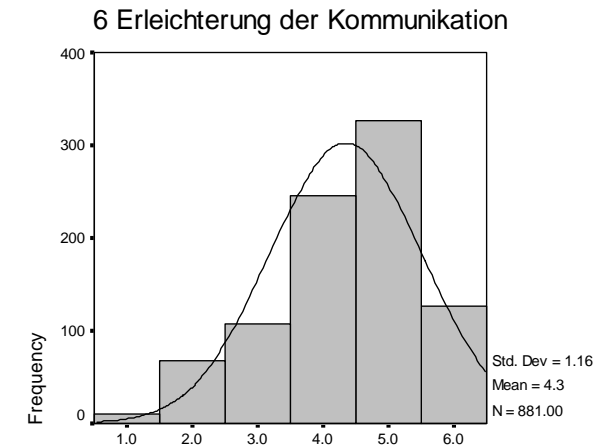
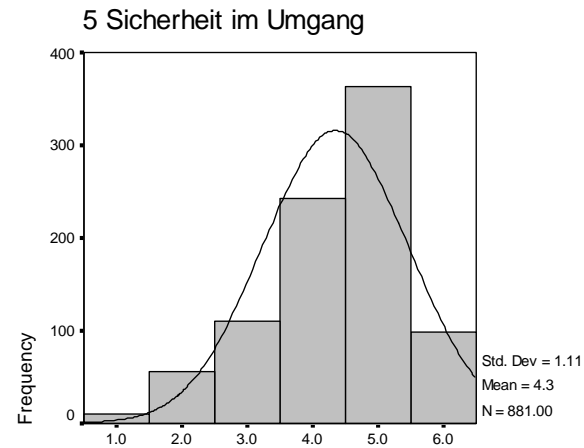
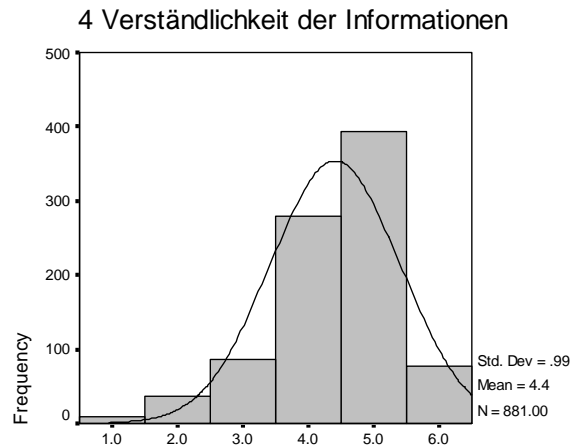
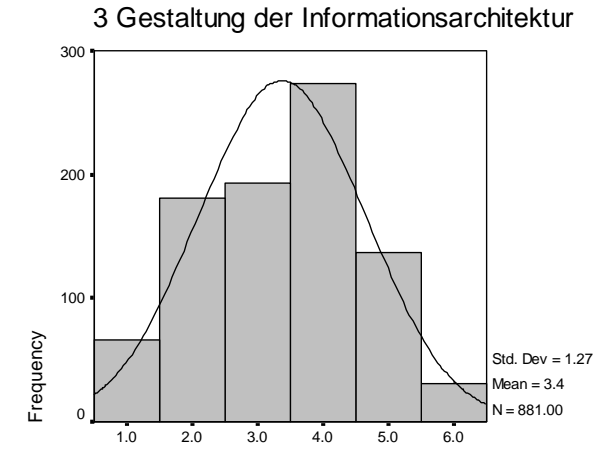
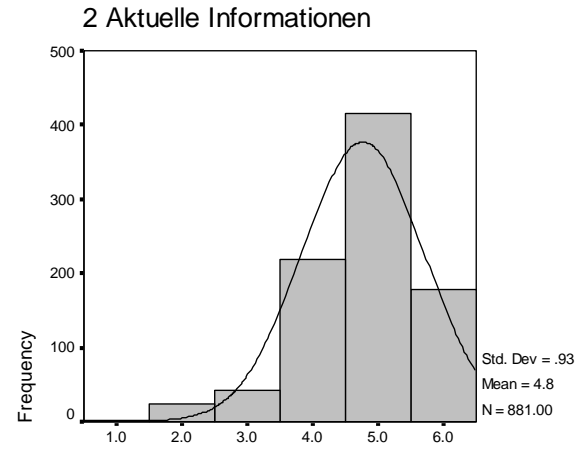
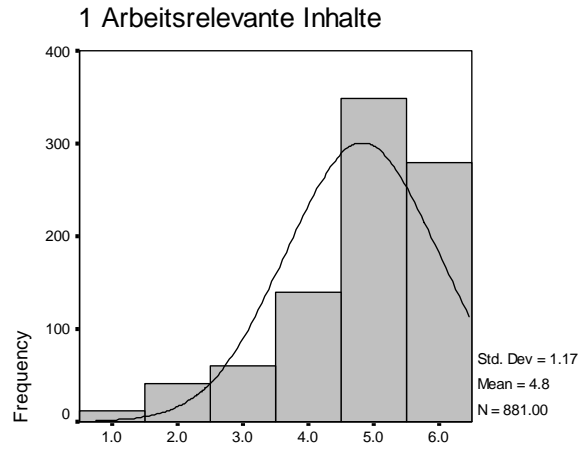
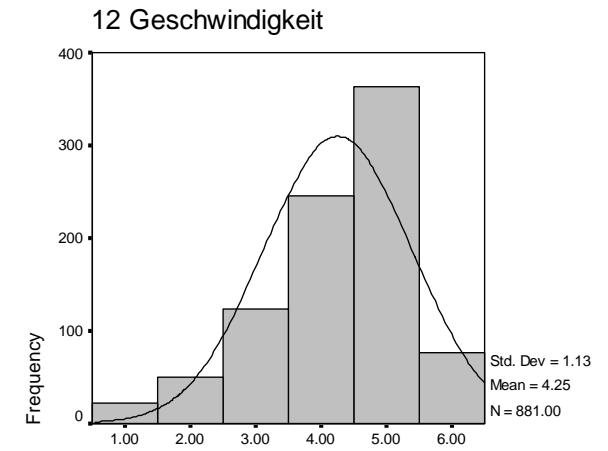
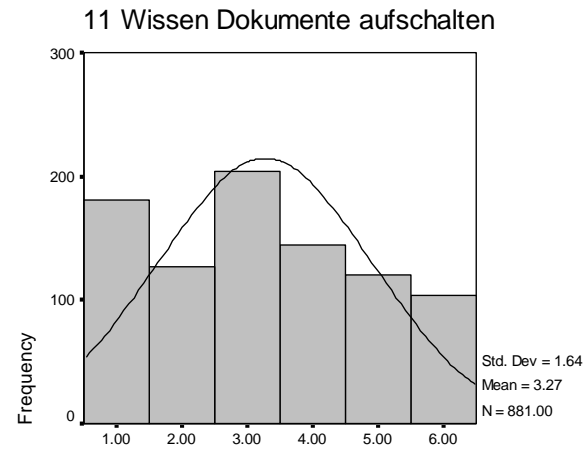
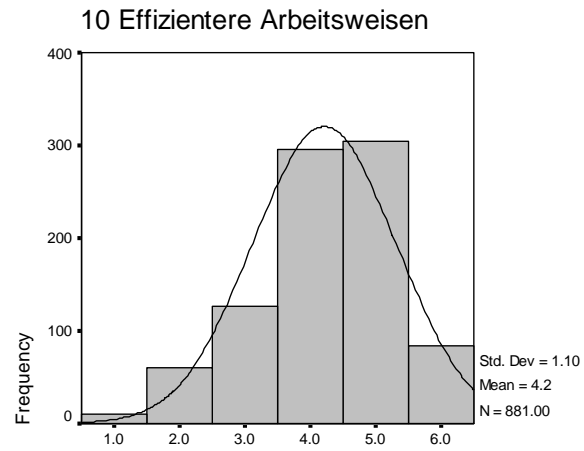
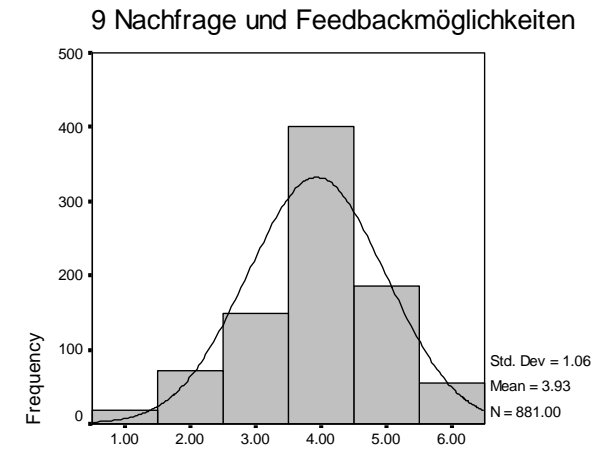
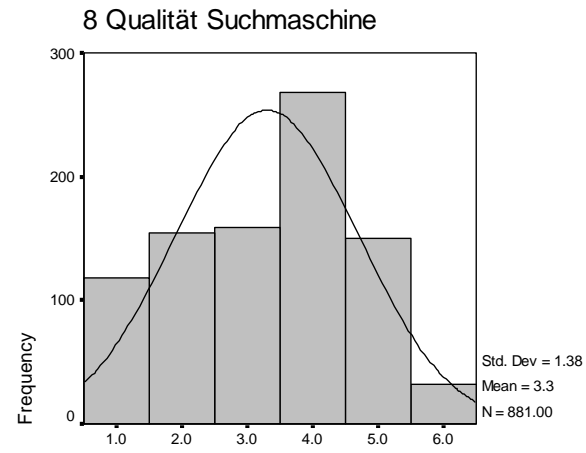
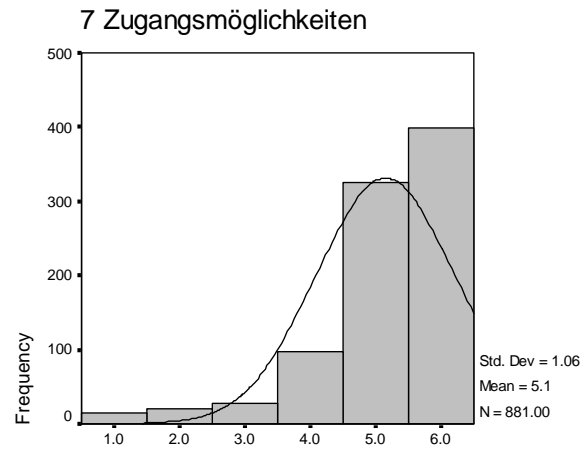
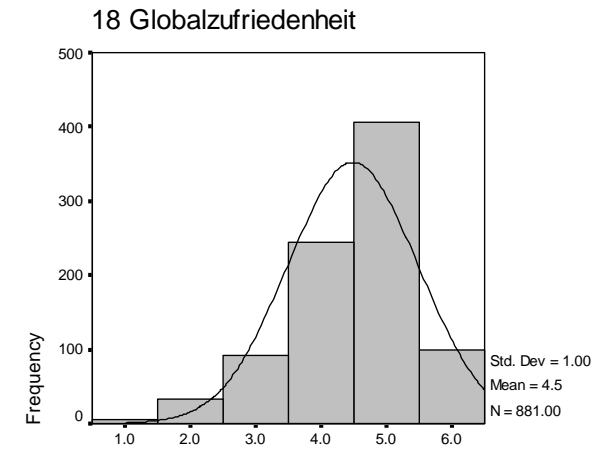
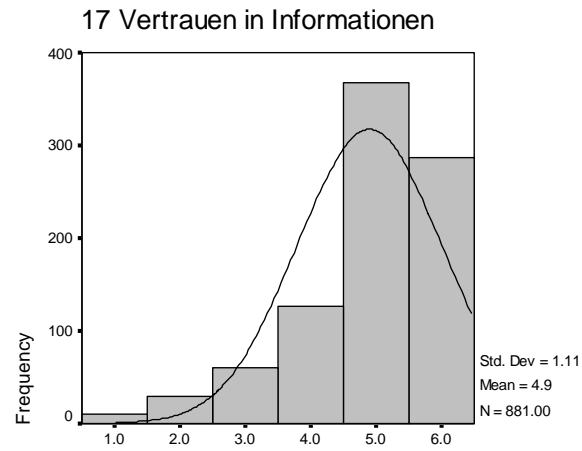
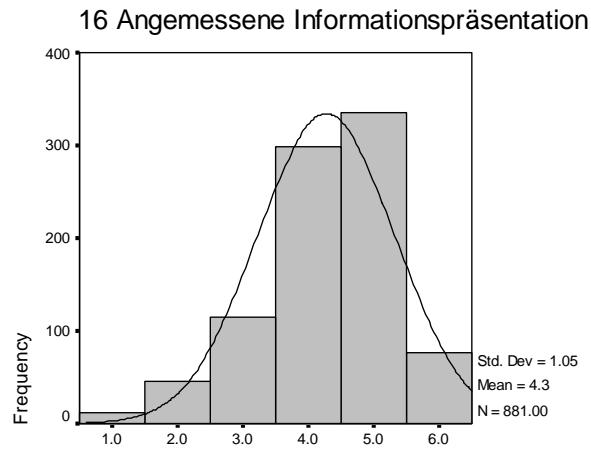
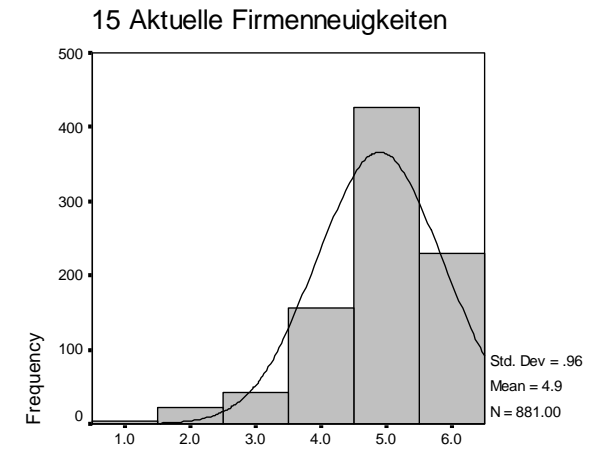
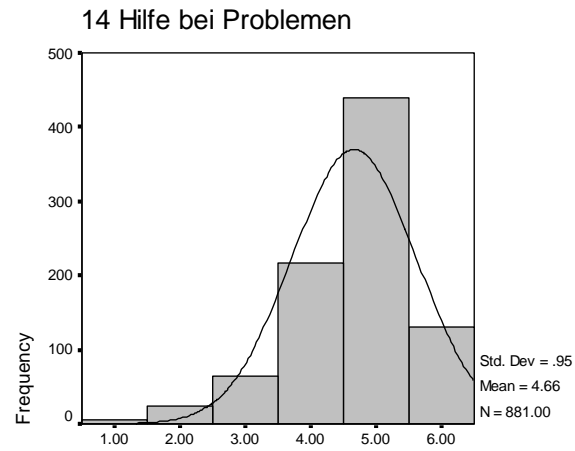
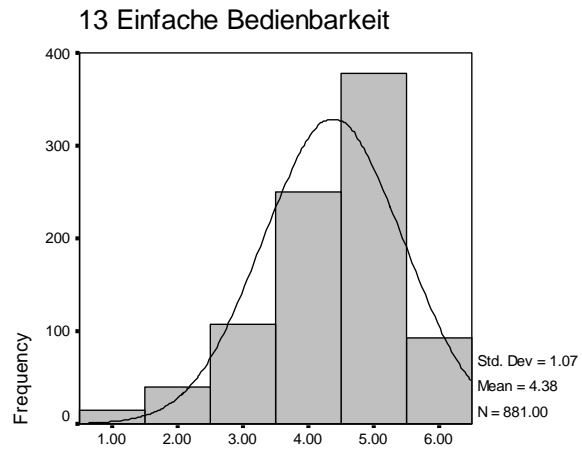


Abbildung C2.4. Trendbereinigtes Normalverteilungsdiagramm von der Mittelwertskala (Item 1 bis 18) der Rohwerte aus dem Unternehmen A ($N = 908$).

Anhang C3*Histogramme der Items 1 bis 18 der ersten Testversion.*





Anhang C4

Die folgenden Tabellen und Abbildungen ergänzen das Kapitel 4.3.4.2 Faktoranalysen.

Tabelle C4.1

Überblick der erklärten Gesamtvarianz der Hauptkomponentenanalyse (Varimax, Items 1 bis 17 der ersten Testversion, N = 881, Missing Values = EM)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.544	38.497	38.497	6.544	38.497	38.497	3.709	21.817	21.817
2	1.375	8.090	46.586	1.375	8.090	46.586	2.854	16.790	38.607
3	1.083	6.371	52.958	1.083	6.371	52.958	2.440	14.351	52.958
4	0.927	5.452	58.410						
5	0.850	4.999	63.409						
6	0.736	4.330	67.739						
7	0.699	4.112	71.851						
8	0.607	3.571	75.422						
9	0.566	3.328	78.750						
10	0.553	3.251	82.001						
11	0.507	2.985	84.986						
12	0.499	2.938	87.924						
13	0.457	2.691	90.614						
14	0.439	2.583	93.197						
15	0.420	2.471	95.669						
16	0.388	2.284	97.953						
17	0.348	2.047	100.000						

Extraction Method: Principal Component Analysis.

Run MATRIX procedure:

PARALLEL ANALYSIS:

Principal Components

Specifications for this Run:

Ncases 881
 Nvars 17
 Ndatsets 1000
 Percent 95

Random Data Eigenvalues

Root	Means	Prcntyle
1.000000	1.245562	1.290537
2.000000	1.197486	1.231535
3.000000	1.160391	1.190364
4.000000	1.128413	1.154058
5.000000	1.098421	1.121690
6.000000	1.071031	1.094805
7.000000	1.044569	1.066549
8.000000	1.019898	1.039880
9.000000	.994581	1.015503
10.000000	.970174	.990152
11.000000	.945651	.966295

```

12.000000      .922034      .942078
13.000000      .896894      .917817
14.000000      .871452      .894190
15.000000      .843811      .869413
16.000000      .813513      .839194
17.000000      .776120      .810094

```

----- END MATRIX -----

Abbildung C4.1. SPSS-Output der Parallelanalyse nach O'Conner (2000) für 1000 Datendateien (Stichproben) mit 881 Fällen und 17 Zufallsvariablen.

Tabelle C4.2

Übersicht der Kommunalität der Items bei unterschiedlicher Faktorenanzahl (Hauptkomponentenanalyse, Varimax, Item 1 bis 17 der ersten Testversion, $N = 881$, Missing values = EM)

Item / Kommunalität	h_1^2	h_2^2	h_3^2
1 Arbeitsrelevante Inhalte	.34	.50	.61
2 Aktuelle Informationen	.41	.54	.57
3 Gestaltung der Informationsarchitektur	.38	.58	.58
4 Verständlichkeit der Informationen	.51	.51	.51
5 Sicherheit im Umgang	.36	.36	.63
6 Erleichterung der Kommunikation	.49	.49	.54
7 Zugangsmöglichkeiten	.27	.40	.45
8 Qualität Suchmaschine	.18	.35	.52
9 Nachfrage und Feedbackmöglichkeiten	.25	.33	.48
10 Effizientere Arbeitsweisen	.57	.57	.59
11 Wissen Dokumente aufschalten	.12	.35	.50
12 Geschwindigkeit	.41	.43	.44
13 Einfache Bedienbarkeit	.47	.53	.53
14 Hilfe bei Problemen	.44	.44	.47
15 Aktuelle Firmenneuigkeiten	.44	.52	.53
16 Angemessene Informationspräsentation	.53	.54	.54
17 Vertrauen in Informationen	.35	.49	.51

h_1^2 = Kommunalität bei Ein-Faktoren-Struktur, h_2^2 = Kommunalität bei Zwei-Faktoren-Struktur, h_3^2 = Kommunalität bei Drei-Faktoren-Struktur.

Anhang C5

Tabelle C5.1

Varianzanalyse (One Way ANOVA) der Itemmittelwerte von der deutsch-, französisch- und italienischsprachigen Stichprobe

		Sum of Squares	df	Mean Square	F	Sig.
v1 Arbeitsrelevante Inhalte	Between Groups	.954	2	.477	.354	.702
	Within Groups	1427.768	1061	1.346		
	Total	1428.722	1063			
v2 Aktuelle Informationen	Between Groups	.703	2	.351	.383	.682
	Within Groups	949.265	1036	.916		
	Total	949.967	1038			
v3 Gestaltung der Informationsarchitektur	Between Groups	26.633	2	13.316	8.108	.000
	Within Groups	1749.128	1065	1.642		
	Total	1775.760	1067			
v4 Verständlichkeit der Informationen	Between Groups	9.824	2	4.912	4.668	.010
	Within Groups	1116.536	1061	1.052		
	Total	1126.360	1063			
v5 Sicherheit im Umgang	Between Groups	11.616	2	5.808	4.532	.011
	Within Groups	1358.406	1060	1.282		
	Total	1370.023	1062			
v6 Erleichterung der Kommunikation	Between Groups	10.094	2	5.047	3.567	.029
	Within Groups	1461.519	1033	1.415		
	Total	1471.614	1035			
v7 Zugangsmöglichkeiten	Between Groups	6.090	2	3.045	2.717	.067
	Within Groups	1192.477	1064	1.121		
	Total	1198.566	1066			
v8 Qualität Suchmaschine	Between Groups	6.108	2	3.054	1.543	.214
	Within Groups	1961.317	991	1.979		
	Total	1967.426	993			
v9 Nachfrage und Feedbackmöglichkeiten	Between Groups	6.572	2	3.286	1.969	.140
	Within Groups	1134.787	680	1.669		
	Total	1141.359	682			
v10 Effizientere Arbeitsweisen	Between Groups	1.140	2	.570	.458	.632
	Within Groups	1278.187	1028	1.243		
	Total	1279.327	1030			
v11 Wissen Dokumente aufschalten	Between Groups	7.998	2	3.999	1.289	.276
	Within Groups	2783.762	897	3.103		
	Total	2791.760	899			
v12 Geschwindigkeit	Between Groups	1.408	2	.704	.528	.590
	Within Groups	1351.689	1014	1.333		
	Total	1353.097	1016			
v13 Einfache Bedienbarkeit	Between Groups	3.706	2	1.853	1.589	.205
	Within Groups	1194.366	1024	1.166		
	Total	1198.072	1026			
v14 Hilfe bei Problemen	Between Groups	29.197	2	14.598	12.149	.000
	Within Groups	934.890	778	1.202		

	Total	964.087	780			
v15 Aktuelle Firmenneuigkeiten	Between Groups	.624	2	.312	.341	.711
	Within Groups	962.686	1051	.916		
	Total	963.310	1053			
v16 Angemessene Informationspräsentation	Between Groups	4.429	2	2.214	1.954	.142
	Within Groups	1144.351	1010	1.133		
	Total	1148.780	1012			
v17 Vertrauen in Informationen	Between Groups	5.381	2	2.690	2.239	.107
	Within Groups	1267.459	1055	1.201		
	Total	1272.839	1057			
v18 Globalzufriedenheit	Between Groups	.796	2	.398	.419	.658
	Within Groups	1007.402	1062	.949		
	Total	1008.197	1064			

Anmerkung: Deutschsprachige Stichprobe N = 881; Französischsprachige Stichprobe N = 143; Italienischsprachige Stichprobe = 42;
Missing values = Pairwise deletion.

Anhang D

Anhang D1

Die Tabellen D1.1 bis D1.4 listen die erhobenen Angaben zu den Probanden aus der Stichprobe der DEZA auf.

Tabelle D1.1

Übersicht der Geschlechterverteilung in der DEZA-Stichprobe

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	weiblich	72	56.7	57.1	57.1
	männlich	54	42.5	42.9	100.0
	Total	126	99.2	100.0	
Missing	keine Antwort	1	.8		
Total		127	100.0		

Tabelle D1.2

Überblick der Alterstruktur der Probanden aus der Stichprobe der DEZA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	bis 29 Jahre	24	18.9	19.0	19.0
	30 bis 39 Jahre	33	26.0	26.2	45.2
	40 bis 49 Jahre	40	31.5	31.7	77.0
	50 Jahre und älter	29	22.8	23.0	100.0
	Total	126	99.2	100.0	
Missing	keine Antwort	1	.8		
Total		127	100.0		

Tabelle D1.3

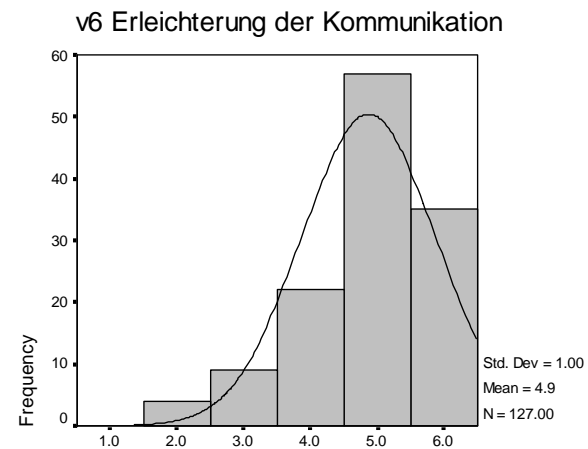
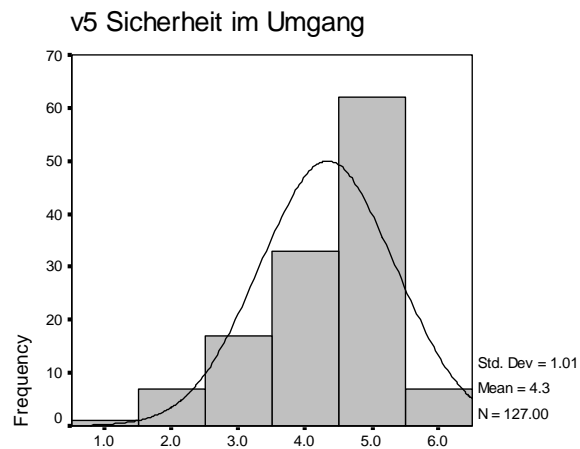
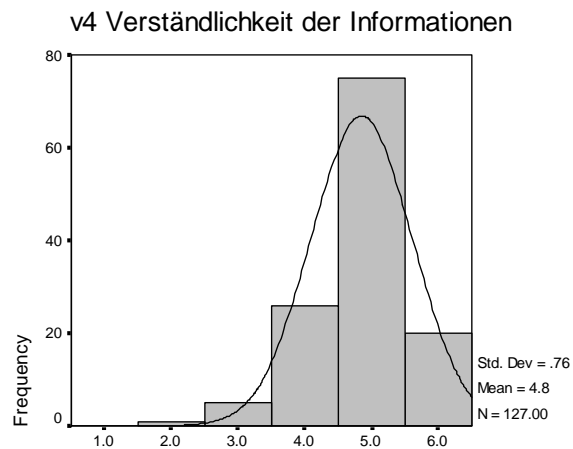
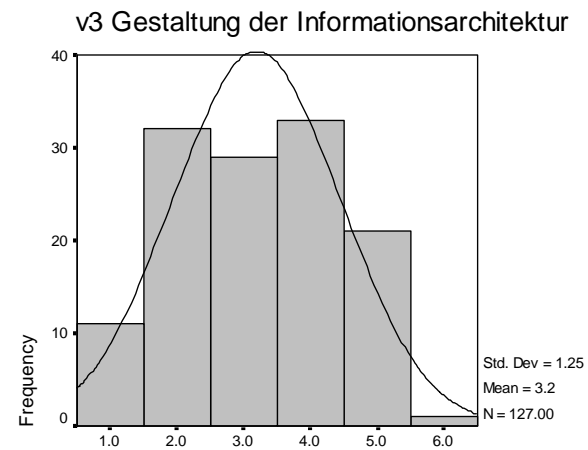
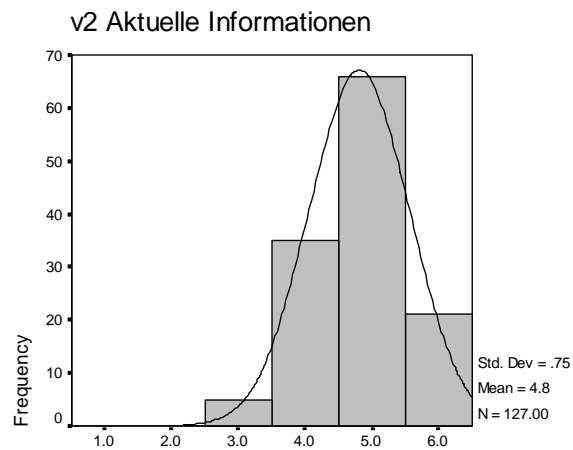
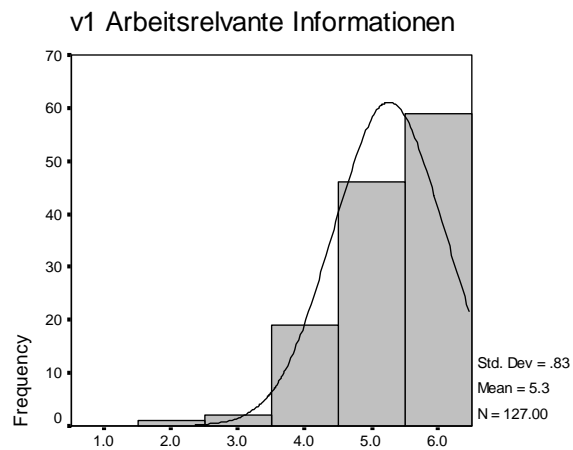
Überblick der Anzahl Dienstjahre der Probanden aus der Stichprobe der DEZA

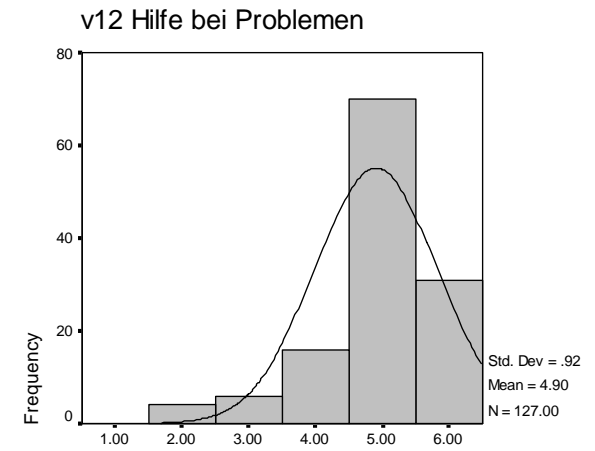
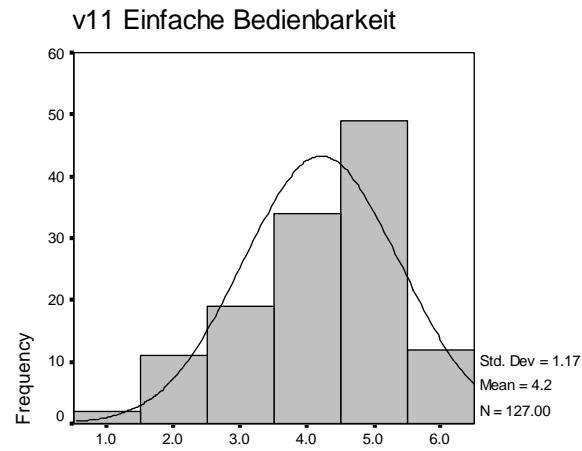
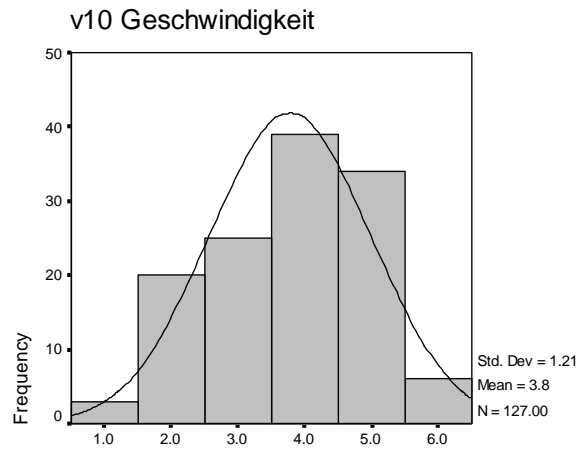
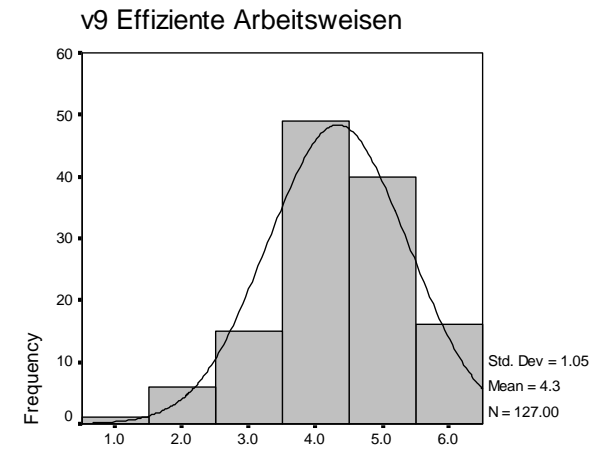
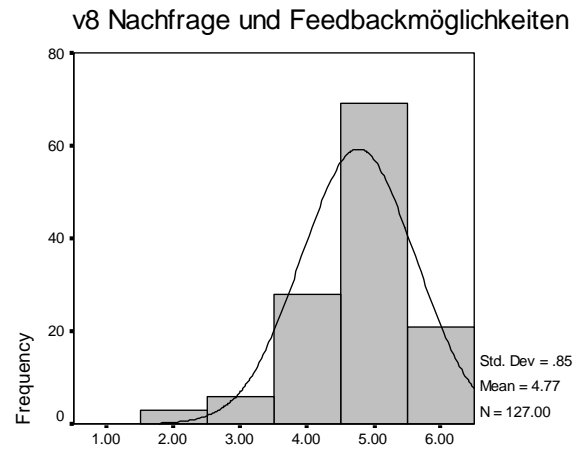
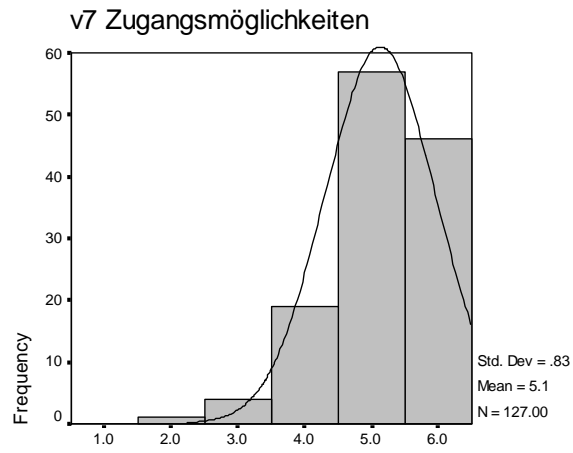
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	weniger als 1 Jahr	17	13.4	13.5	13.5
	1 bis 2 Jahre	10	7.9	7.9	21.4
	3 bis 5 Jahre	45	35.4	35.7	57.1
	6 bis 10 Jahre	25	19.7	19.8	77.0
	11 Jahre und länger	29	22.8	23.0	100.0
	Total	126	99.2	100.0	
Missing	keine Antwort	1	.8		
Total		127	100.0		

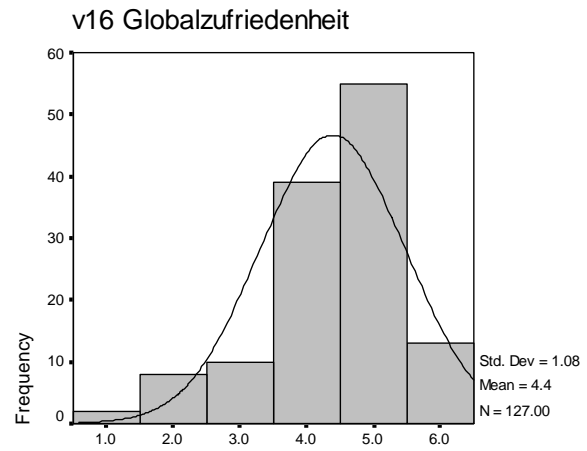
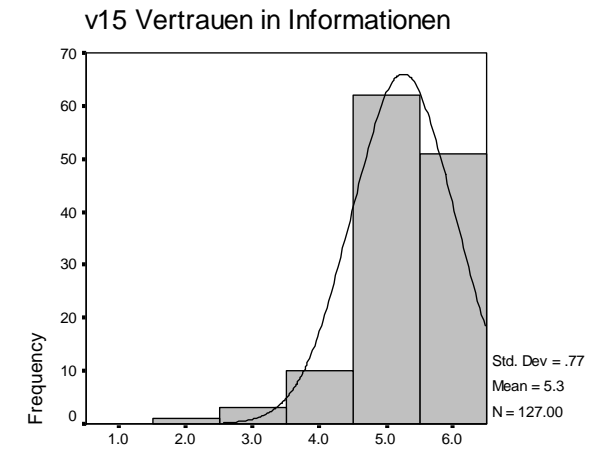
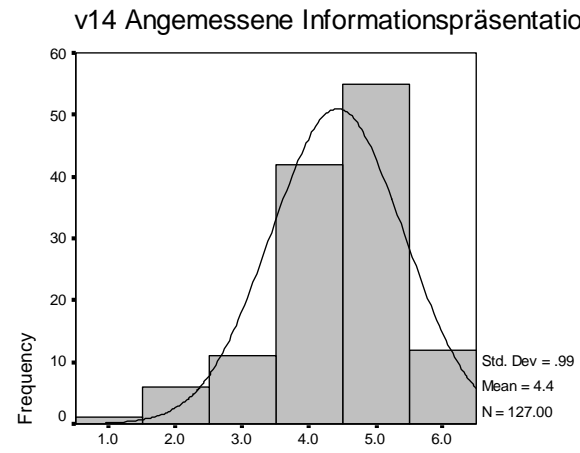
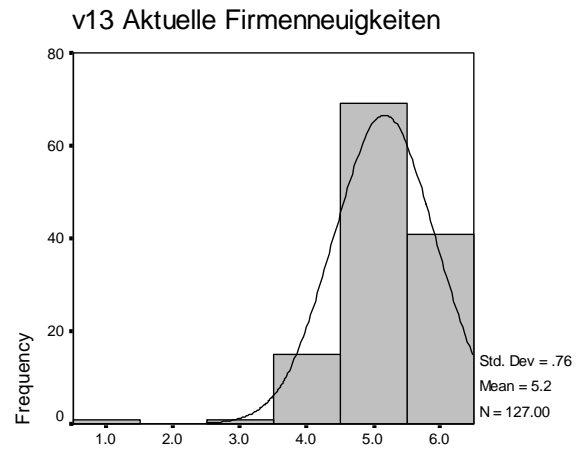
Tabelle D1.4

Überblick der von den Probanden der DEZA-Stichprobe angegebenen durchschnittlichen Benutzung des Intranets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	monatlich	0	0	0	0
	wöchentlich	6	4.7	4.8	4.8
	täglich	41	32.3	32.5	37.3
	mehrmals täglich	79	62.2	62.7	100.0
	Total	126	99.2	100.0	
Missing	keine Antwort	1	.8		
Total		127	100.0		

Anhang D2*Histogramme der Items 1 bis 16 der zweiten Testversion.*





Anhang D3

Die folgenden Tabellen und Abbildungen ergänzen das Kapitel 5.2.4.2 Faktoranalysen.

Tabelle D3.1

Überblick der erklärten Gesamtvarianz der Hauptkomponentenanalyse (Varimax, Items 1 bis 15 der zweiten Testversion, N = 127, Missing values = EM)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.771	31.804	31.804	4.771	31.804	31.804	3.244	21.628	21.628
2	1.648	10.984	42.787	1.648	10.984	42.787	2.148	14.318	35.946
3	1.249	8.326	51.113	1.249	8.326	51.113	1.976	13.171	49.117
4	1.086	7.240	58.353	1.086	7.240	58.353	1.385	9.235	58.353
5	0.891	5.941	64.294						
6	0.827	5.511	69.805						
7	0.764	5.091	74.896						
8	0.654	4.358	79.254						
9	0.593	3.956	83.211						
10	0.526	3.508	86.718						
11	0.504	3.362	90.080						
12	0.474	3.159	93.239						
13	0.388	2.587	95.826						
14	0.345	2.301	98.127						
15	0.281	1.873	100.000						

Extraction Method: Principal Component Analysis.

```

Run MATRIX procedure:

PARALLEL ANALYSIS:

Principal Components

Specifications for this Run:
Ncases      127
Nvars       15
Ndatasets   1000
Percent     95

Random Data Eigenvalues

      Root      Means      Prcntyle
1.000000    1.631942    1.767459
2.000000    1.482737    1.576954
3.000000    1.372212    1.452615
4.000000    1.276430    1.350724
5.000000    1.189175    1.256270
6.000000    1.109803    1.171607
7.000000    1.034836    1.092834
8.000000    .967516      1.022320
9.000000    .902338      .956051
10.000000   .838143      .895352
11.000000   .772023      .823850
12.000000   .707995      .761530
13.000000   .644038      .700699
14.000000   .575259      .633648
15.000000   .495551      .561293

----- END MATRIX -----

```

Abbildung D3.1. SPSS Output der Parallelanalyse nach Horn auf Basis der SPSS-Syntax von O'Conner (2000) für 1000 Datendateien (Stichproben) mit 127 Fällen und 15 Zufallsvariablen.

Tabelle D3.2

Übersicht der Kommunalität der Items 1 bis 15 aus der zweiten Testversion bei unterschiedlicher Faktorenzahl (Hauptkomponentenanalyse, Varimax, $N = 127$, Missing values = EM)

Item / Kommunalität	h_1^2	h_2^2	h_3^2	h_4^2
1 Arbeitsrelevante Informationen	.16	.17	.42	.75
2 Aktuelle Informationen	.07	.52	.55	.59
3 Gestaltung der Informationsarchitektur	.46	.66	.67	.67
4 Verständlichkeit der Informationen	.20	.38	.51	.59
5 Sicherheit im Umgang	.34	.60	.61	.64
6 Erleichterung der Kommunikation	.36	.36	.46	.50
7 Zugangsmöglichkeiten	.23	.23	.37	.57
8 Nachfrage und Feedbackmöglichkeiten	.21	.29	.37	.52
9 Effizientere Arbeitsweisen	.51	.62	.63	.68
10 Geschwindigkeit	.40	.42	.56	.56
11 Einfache Bedienbarkeit	.52	.53	.53	.63
12 Hilfe bei Problemen	.44	.45	.47	.49
13 Aktuelle Firmenneuigkeiten	.15	.37	.37	.37
14 Angemessene Informationspräsentation	.48	.51	.51	.54
15 Vertrauen in Informationen	.25	.31	.65	.65

h_1^2 = Kommunalität bei Ein-Faktoren-Struktur, h_2^2 = Kommunalität bei Zwei-Faktoren-Struktur
 h_3^2 = Kommunalität bei Drei-Faktoren-Struktur, h_4^2 = Kommunalität bei Vier-Faktoren-Struktur

Anhang D4

Französische Version des Intranet Satisfaction Questionnaire (noch nicht validiert).

1	L’Intranet met à ma disposition des contenus importants pour mon travail.
2	Dans l’Intranet, les informations sont actuelles.
3	L’Intranet est conçu de façon claire et sa structure est compréhensible.
4	Les informations que je lis sur l’Intranet sont rédigées clairement et explicitement.
5	L’Intranet facilite la communication interne (par exemple grâce à l’annuaire des collaborateurs, aux communications du Management ou aux forums de discussion, etc.).
6	Lorsque j’utilise le moteur de recherche pour trouver une information, j’obtiens celle-ci en temps utile.
7	L’Intranet me permet de travailler de façon plus efficace (par exemple processus de travail internes, assistance ou obtention d’informations).
8	L’Intranet permet un travail accéléré (par exemple les pages apparaissent rapidement ou les documents peuvent être facilement téléchargés).
9	L’Intranet est facile à utiliser (par exemple personnalisation, utilisation de l’annuaire des collaborateurs, etc.).
10	Je suis satisfait/e du service d’aide et de l’assistance que j’obtiens lorsque je rencontre un problème ou lorsque j’ai une question au sujet de l’Intranet (par exemple HelpDesk ou aide online).
11	L’Intranet me fournit en permanence des nouvelles actuelles relatives à la Société.
12	Dans l’Intranet, les informations nécessaires à mon travail sont présentées de telle manière que je peux facilement les utiliser (par exemple, sans devoir effectuer des modifications compliquées).
13	Dans l’ensemble, je suis satisfait/e de l’Intranet.

Anmerkung. 1= Pas d’accord 6 = Entièrement d’accord. Je ne peux pas juger + champ libre : « Désirez-vous ajouter une précision à votre réponse? »

Italianische Version des Intranet Satisfaction Questionnaire (noch nicht validiert).

1	L’Intranet mette a mia disposizione dei contenuti importanti per il mio lavoro.
2	Nell’Intranet, le informazioni sono attuali.
3	L’Intranet è organizzato chiaramente e la struttura è comprensibile.
4	Quando leggo qualcosa su intranet, le informazioni sono chiare e comprensibili.
5	L’Intranet facilita la comunicazione interna (per esempio tramite l’elenco dei collaboratori, le comunicazioni del Management o i fori di discussione, ecc.).
6	Quando cerco qualcosa con un motore di ricerca, trovo l’informazione desiderata entro un periodo utile.
7	L’Intranet mi permette di lavorare efficientemente (per esempio con i processi di lavoro interni, per il supporto o per la ricerca d’informazioni).
8	L’Intranet permette un lavoro rapido (per esempio rapida presentazione di un sito o scaricamento di documenti).
9	L’Intranet è facile ad utilizzare (per esempio personificazione, uso dell’elenco dei collaboratori, ecc.).
10	Sono contento con l’aiuto, rispettivamente con il sostegno che ricevo quando ho un problema o una domanda riguardante l’Intranet (per esempio Help Desk o guida online).
11	L’Intranet mi comunica continuamente le novità attuali riguardanti la Società.
12	Trovo nell’Intranet le informazioni necessarie per il mio lavoro in tale modo che posso utilizzarle semplicemente (per esempio non sono necessarie complicate modifiche).
13	Nell’insieme, sono soddisfatto dell’Intranet.

Anmerkung. 1= Non sono d’accordo. 6 = Sono pienamente d’accordo. Non posso rispondere + campo aperto: “Desidera com unircarci ancora qualcosa al soggetto di questa dichiarazione?”